Universidade de Brasília

Faculdade de Economia, Administração e Contabilidade

Departamento de Administração

SARAH SABINO DE FREITAS MARCELINO

FORMAÇÃO DE PORTFÓLIO POR MEIO DE MÁQUINAS DE SUPORTE VETORIAL

Brasília – DF

2014

SARAH SABINO DE FREITAS MARCELINO

FORMAÇÃO DE PORTFÓLIO POR MEIO DE MÁQUINAS DE SUPORTE VETORIAL

Monografia apresentada ao Departamento de Administração como requisito parcial à obtenção do título de Bacharel em Administração.

Professor Orientador: Dr. Pedro Henrique Melo Albuquerque

Brasília – DF

2014

SARAH SABINO DE FREITAS MARCELINO

FORMAÇÃO DE PORTFÓLIO POR MEIO DE MÁQUINAS DE SUPORTE VETORIAL

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de Conclusão do Curso de Administração da Universidade de Brasília da aluna

Sarah Sabino de Freitas Marcelino

Doutor, Pedro Henrique Melo Albuquerque
Professor-Orientador

Doutor, Thiago Veiga Marzagão
Professor-Examinador

Mestre, Yuri Sampaio Maluf
Professor-Examinador

Brasília, 20 de Novembro de 2014

A Deus, o único que é digno de toda glória e honra e que me guiou até aqui. À minha família que sempre esteve ao me lado com apoio, incentivo e compreensão.

AGRADECIMENTOS

Ao meu Professor Orientador, Pedro Henrique Melo Albuquerque,por ser meu principal suporte vetorial nessa caminhada.Obrigada Pedro, por todoauxílio, apoio, incentivo, paciência e disponibilidade.Agradeço também aos meus colegas de curso, em especial ao Pedro Alexandre Moura Barros Henrique que me auxiliou e me apoiou na parte final da pesquisa.

RESUMO

A presente pesquisateve como objetivoreplicar a metodologia de Máquinas de Suporte Vetorial proposta por Fan e Palaniswami (2001) no contexto brasileiro de formação de portfólio.O SVM foi então utilizado para verificar se o uso de Máquinas de Suporte Vetorial na formação de portfólios de fato contribui para que o retorno seja superior ao de um *benchmark* do mercado, sendo que o ativo escolhido para tal comparação foi o fundo de índice BOVA11. A amostra foi constituída por 67 ações que compuseram a carteira teórica válida para 2 de Setembro de 2013 a 03 de Janeiro de 2014 e os insumos para o modelo foram dados históricos de preço e indicadores financeiros coletados na base de dados do sistema Economática, no recorte temporal de 2000 a 2013.A função de decisão do SVM classificou os ativos na Classe 1 ou na Classe 2 de acordo com o *ranking* dos *outputs* que foram interpretados como a probabilidade da ação ser classificada como $+1$. Assim, a Classe 1 foi composta dos 25% de ações com maiores probabilidades, e a Classe 2 foi constituída pelas demais ações. Nas classificações de ativos feitas pelo SVM, utilizando os parâmetros ótimos, a máquina acertou a classificação em 73,48% das vezes. No período de teste de aproximadamente 5 anos, o retorno acumulado do *benchmark*foi de 19,34%, enquanto o do SVM foi de 257,36%. Em termos de retorno trimestral médio, o SVM apresentou um retorno médio de 8,26%, enquanto o BOVA11 foi de 1,64%. Os resultados tornaram evidente que o SVM superou *benchmark* em 403,92%, entretanto, o contexto econômicoacentuou em grande medida a discrepância entre os resultados. Por isso, o portfólio formado foi comparado com um segundo *benchmark*de mercado composto por todas as 67 açõesda carteira teórica do Ibovespa utilizada na pesquisa. O retorno trimestral médio deste segundo *benchmark* foi de 7,12% e o retorno acumulado foi de 183,41%. Portanto, novamente o retorno do portfólio escolhido pelo SVM foi superior ao *benchmark*, dessa vez, em 16,08%. Para testar a significância estatística dos resultados e controlar o efeito *Data Snooping*, o método *Bootstrap* foi utilizado.

Palavras-chave: Máquinas de Suporte Vetorial. SVM. Formação de Portfólio.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ABBS – *Accrual Based on Balance Sheet*

ABCF – *Accrual Based on Cash Flow*

AC – Ativo Circulante

AR – *AccountsReceivable*

AT – Ativo Total

BV – *Book Value*

CE – *Cash andEquivalents*

CEx – *Capital Expenditures*

CFFA –*Cash FromFinancingActivities*

CFIA –*Cash FromInvestingActivities*

CFOS – *Cash FromOperatingActivities*

CL – *CurrentLiabilities*

D – *Dividends*

DNEPS – *Diluted Normalized Earnings Per Share*

DP – *Depreciation*

DPR – *DividendPayoutRatio*

ETF – *Exchange TradedFund*

FH – *Financial Health*

GP –*Gross Profit*

I – Imobilizado

IAT –*IncomeAfterTax*

IBT –*IncomeBeforeTax*

LB – Lucro Bruto

LL – Lucro Líquido

MAC – Média das Ações em Circulação

NCIC – *Net Change In Cash*

NI –*Net Income*

NIBEI –*Net Income Before Extraordinary Items*

OI –*OperatingIncome*

PAM – Participação dos Acionistas Minoritários

PC – Passivo Circulante

PL – Patrimônio Líquido

PNC – Passivo Não Circulante

PPA – Participação Patrimonial dos Acionistas

RLP – Realizável a Longo Prazo

ROE –*ReturnonEquity*

ROL – Participação Patrimonial dos acionistas

SAC -– *Snapshot Accrual*

STI – *Short TermInvestiments*

STL – *Short TermLiabilities*

SVM – *Support Vector Machines*

TA – *Total Assets*

TCA – *Total CurrentAssets*

TCL – *Total CurrentLiabilities*

TD – *Total Debt*

TE – *Total Equity*

TI – *Total Inventory*

TL – *Total Liabilities*

TLTD – *Total LongTermDebt*

TR – *Total Revenue*

TS – *Total Shares*

VL – Participação Patrimonial dos Acionistas

SUMÁRIO

1 INTRODUÇÃO

A seleção de ações é uma parte desafiadora e crucial do processo de decisão dos investidores. Considerando o enorme montante de opções de ativos disponíveis no mercado financeiro, segundo Fan e Palaniswami (2001), o desafio da seleção de ações está na identificação dos ativos com potencial de superar o mercado no próximo ano.

Para Tay e Cao (2001, p.309, tradução nossa), “a previsão de séries temporais financeiras é considerada uma das aplicações mais desafiadoras da previsão de séries temporais”. Abu-Mostafa e Atiya (1996) apontam que os especuladores, investidores e empresas, em sua busca para prever o comportamento dos mercados, assumem que as ocorrências futuras se baseiam, pelo menos em parte, em eventos e dados presentes e passados. No entanto, as séries financeiras são permeadas por ruídos, não-estacionariedade e caos determinístico. A primeira característica refere-se à indisponibilidade de informações completas do comportamento passado dos mercados financeiros que poderiam contribuir para a análise da dependência dos preços futuros e passados, sendo assim, ruídos são as informações não incluídas no modelo ou erros de medidas. A não-estacionariedade implica que a distribuição das séries temporais financeiras muda ao longo do tempo e padrões que representam dados passados podem não ser mais aplicáveis. A presença de caos determinístico, por sua vez, significa que no curto prazo, o comportamento das séries temporais financeiras é aleatório, mas no longo prazo apresenta um padrão determinístico.

Este contexto levou muitos economistas a adotarem a Hipótese do Mercado Eficiente, que considera que as mudanças nos preços das ações são independentes do passado e seguem um padrão aleatório (ABU-MOSTAFA; ATIYA, 1996). As mudanças de preços seriam então imprevisíveis e qualquer alteração no preço representaria uma reação imediata a um novo evento ou a uma mudança inesperada de oferta ou demanda. Se houvesse qualquer oportunidade inesperada de lucro, por exemplo, os investidores a explorariam imediatamente de forma que o preço voltaria ao patamar que estava quando essa oportunidade não existia. Ainda segundo essa teoria, quaisquer padrões úteis deveriam refletir no preço corrente, porém, se o mercado é eficiente a ponto de todos os preços das ações refletirem

plenamente todas as informações públicas disponíveis, não se pode esperar que essa forma de análise consiga identificar com antecedência os investimentos com retornos superiores ao mercado. Apesar de existirem vários debates sobre a Hipótese do Mercado Eficiente, é difícil de refutá-la ou não (ABU-MOSTAFA; ATYIA, 1996, p.205).

Na perspectiva de mineração de dados, os retornos futuros das ações são considerados, em alguma medida, previsíveis. De acordo com Fan e Palaniswami (2001), o problema de predição envolve a descoberta de padrões de relacionamentos úteis nos dados e aplicação dessa informação para classificar as ações. Uma abordagem que tem se mostrado promissora para esse problema são as Máquinas de Suporte Vetorial (Support Vector Machine – SVM), propostas por Boser, Guyon e Vapnik, em 1992. Originalmente, as Máquinas de Suporte Vetorial foram desenvolvidas para o reconhecimento de padrões em um conjunto de dados. Segundo Albuquerque (2014, p.11), “por meio desse reconhecimento é possível realizar um processo de inferência indutiva, o qual seria capaz de realizar previsões para um conjunto de dados observados posteriormente à estimação dos parâmetros do modelo”.

Desde sua criação em 1992, as Máquinas de Suporte Vetorial marcaram o início de uma nova era na inteligência artificial, representando uma quebra de pensamento na Teoria de Aprendizado Estatístico (SOMAN, K. P.; LOGANATHAN, R.; AJAY, V., 2011). Com a implementação do Princípio de Minimização do Risco Estrutural, que minimiza o limite superior do erro de generalização ao invés de minimizar apenas o erro empírico, o SVM abre um novo panorama para modelagem de algoritmos de aprendizagem de máquina com maior capacidade de generalização, superando a maioria das dificuldades enfrentadas pelos algoritmos tradicionais, como o *overfitting* e a alta dimensionalidade de dados. Segundo Soman, Loganathan e Ajay (2011), com o SVM, a solução ótima sempre é encontrada.

1.1 Formulação do problema

“A seleção de ações para formação de portfólios envolve a obtenção de

proporções ideais entre os ativos, para construir uma carteira que respeite as preferências dos investidores" (GUPTA; MEHLAWAT; MITTAL, 2012, p.297, tradução nossa). O estudo de Markowitz (1952) é o pioneiro na área de otimização de portfólios. Segundo ele, o retorno esperado e o risco, expresso pela variância desses retornos, são as duas únicas variáveis que interessam à utilidade do investidor. Markowitz (1952) também assume que os investidores são avessos ao risco e somente aceitarão correr mais risco caso o retorno também aumente.

As Máquinas de Suporte Vetorial surgem como método alternativo para seleção de ações, aplicando simultaneamente a minimização do erro de classificação e a maximização da margem geométrica. Portanto, pode-se indagar: No contexto brasileiro, um portfólio formado por meio de Máquinas de Suporte Vetorial possui um retorno superior ao do mercado?

1.2 Objetivo Geral

A presente pesquisateve como objetivo verificar se o uso de Máquinas de Suporte Vetorial contribui para que o retorno do portfólio seja superior a um *benchmark* do mercado, definidocomo o fundo de índice BOVA11.

1.3 Objetivos Específicos

O primeiro passo para alcançar o objetivo geral apresentado será replicar o modelo de Máquinas de Suporte Vetorial proposto por Fan e Palaniswami (2001), no contexto brasileiro para formação de portfólios. Posteriormente, formar um portfólio com ações selecionadas pelo SVM e analisar seu retorno. A terceira etapa será a comparação entre o retorno do portfólio selecionado pelo SVM e o retorno obtido no mesmo período caso o mesmo montante fosse aplicado no fundo de índice BOVA11.

1.4 Justificativa

A aplicação de Máquinas de Suporte Vetorial é um tema bastante recente, visto que o modelo foi proposto inicialmente por Boser, Guyon e Vapnikem 1992 e aplicado em finanças pela primeira vez em 2001 no estudo de Fan e Palaniswami. Apesar de ainda existirem poucos estudos da aplicação do SVM em previsão de séries temporais financeiras, o número de investigações acerca do tema está em constante crescimento. O Gráfico 1 compara a quantidade de publicações e citações encontradas no Google Acadêmico com as palavras chaves *Support Vector Machines* e *Portfolio Markowitz*.

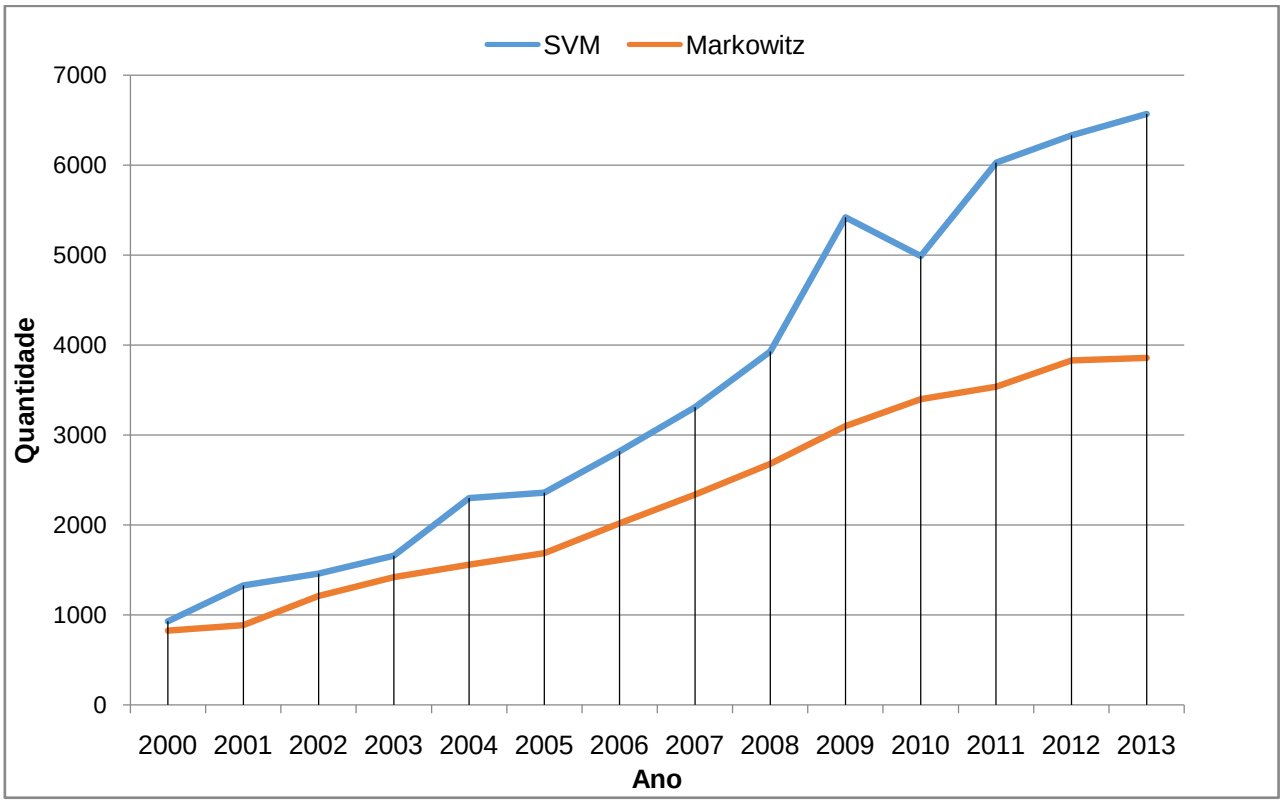


Gráfico 1 - Comparação entre publicações

Fonte: Elaborado pela autora.

A relevância teórica deste projeto de pesquisa relaciona-se à geração de conhecimento científico sobre aplicação de Máquinas de Suporte Vetorial no Mercado, em especial, no Mercado de Ações Brasileiro. Até o momento, publicações em português acerca do tema e aplicações do SVM ao Mercado Brasileiro são inexistentes, portanto, este trabalho visa suprir essa lacuna replicando

a metodologia proposta por Fan e Palaniswami (2001) no Mercado de Ações Brasileiro.

As Máquinas de Suporte Vetorial têm se mostrado eficientes na superação dos ruídos que permeiam as Séries Temporais Financeiras. Fan e Palaniswami (2001), Yu, Lu e Chang (2008) mostram que essa abordagem supera o Mercado consistentemente na classificação de ações, abrindo assim espaço para mais pesquisa sobre esse tema.Huerta, Corbacho e Elkan (2013) apontam que SVMs não lineares conseguem identificar sistematicamente ações com alto ou baixo retorno. No estudo de Kim (2003) as Máquinas de Suporte Vetorial superaram outras técnicas de mineração de dados na previsão do preço de ações do Mercado da Coréia. Lu, Yu e Lin (2008) utilizaram o SVM para prever as melhores ações do Mercado de Taiwan e os resultados mostram que a performance desse modelo supera a de outros métodos do mercado em termos de risco, acurácia e menor erro. Portanto, apesar de algumas limitações, vários trabalhos publicados utilizando as Máquinas de Suporte Vetorial em séries temporais apresentaram bons resultados, justificando assim a utilização dessa abordagem em Séries Temporais Financeiras.

Do ponto de vista aplicado, esta pesquisa é relevante na medida em que contribuirá para a implementação em *software* do modelo SVM para construção do portfólio. Essa ferramenta será de grande utilidade na tomada de decisão de gestores, financistas e investidores.

2 REFERENCIAL TEÓRICO

Para compreensão do contexto no qual as Máquinas de Suporte Vetorial estão inseridas, faz-se necessária a revisão da literatura sobre formação de portfólio e sobre as aplicações mais significativas do SVM em finanças.

Com a publicação do estudo pioneiro de Markowitz (1952), o modelo média-variância revolucionou a forma como as pessoas pensam sobre portfólios de ativos e ganhou ampla aceitação como uma ferramenta prática para otimização de portfólios (LAI; YU; WANG; ZHOU, 2006). Entretanto, a Teoria de Markowitz fornece solução apenas para alocação de ativos previamente selecionados. Como no Mercado Financeiro centenas de diferentes ativos como ações, títulos, opções, *commodities*, contratos futuros, estão disponíveis para negociação e a qualidade destes pode variar bastante, a escolha dos ativos para investimento é um passo crítico, segundo Lai, Yu, Wang e Zhou (2006). Os autores ainda defendem que a qualidade dos ativos escolhidos para investimento é essencial para o bom desempenho do portfólio, mesmo que estejam alocados de forma a minimizar o risco e isso é muitas vezes negligenciado pela Teoria de Markowitz (1952).

O estudo Lai, Yu, Wang e Zhou (2006) teve como objetivo desenvolver um Algoritmo Genético de Otimização de Dois Estágios para formação de portfólios, sendo os dados oriundos dos preços de fechamento diários de 100 ações selecionadas aleatoriamente da Bolsa de Valores de Shangai, *Shanghai Stock Exchange (SSE),* para o período de Janeiro de 2001 a Dezembro de 2004.

No primeiro estágio, o algoritmo genético é usado como uma ferramenta de classificação das ações tendo como insumos indicadores financeiros das ações listadas. O objetivo desse estágio é permitir que os investidores selecionem apenas ações de boa qualidade. No segundo estágio, a alocação dos ativos de boa qualidade é otimizada usando um algoritmo genético baseado na Teoria de Markowitz (1952). Os autores ressaltam que no primeiro estágio algumas ações podem ser consideradas de boa qualidade tendo como base apenas o *ranking* de retornos, entretanto, como ressaltado anteriormente, o gerenciamento de um portfólio não deve focar apenas no retorno, mas também na minimização do risco. Sendo assim, a melhor solução para alocação de ativos na carteira é uma

quantidade de ações que minimize o risco de um determinado nível de retorno esperado e definido pelo investidor.

Os resultados mostraram que o retorno líquido acumulado do portfólio igualmente ponderado foi pior do que o portfólio otimizado pelo algoritmo genético. Isso implica que se um investidor sem experiência escolhe aleatoriamente ações para seu portfólio, o retorno esperado da carteira vai ser aproximadamente igual ao seu valor. Mesmo que o investidor não perca dinheiro com essa seleção de ações, devido ao custo de capital ele já perde algum recurso. Os resultados do estudo de Lai, Yu, Wang e Zhou (2006) também mostram que o desempenho do portfólio diminui à medida que o número de ações aumenta. Segundo os autores, um maior número de ações gera maior flexibilidade para composições de minimização de risco, mas selecionar ações de boa qualidade é um pré-requisito para se obter um bom portfólio. Ações de má qualidade, mesmo se incluídas em composições que minimizem o risco, podem influenciar negativamente o desempenho do portfólio.

Os resultados do estudo também mostraram que se um investidor selecionar apenas ações de boa qualidade, um portfólio mais volumoso não necessariamente supera um portfólio com poucas ações. Dessa forma, segundo Lai, Yu, Wang e Zhou (2006) é sensato que os investidores selecionem um número limite de ações e que todas sejam de boa qualidade.

Visando aprimorar a forma de com que os ativos de um portfólio são selecionados, esta pesquisa utilizará as Máquinas de Suporte Vetorial como ferramenta de classificação de ativos.

2.1 Aplicação de Máquinas de Suporte Vetorial em Finanças

A primeira aplicação direta das Máquinas de Suporte Vetorial em finanças refere-se à aplicação do modelo para classificação de ações e formação de portfólio, abordagem proposta por Fan e Palaniswami em 2001. A utilidade do SVM foi testada com informações contábeis das ações negociadas na *Australian Stock Exchange* para o período de 1992 a 2000. O Quadro 1, mostra todos os indicadores calculados com os relatórios financeiros e dados de preços disponíveis.

Retornosobre Capital	**Investimento**
Profit Before Tax / Total Assets	*Price-EarningsRatio*
Profit Before Tax / Total Capital	*Net Tangible Assets per Share*
Net Income / Total Capital	*Dividend Yield*
Cash Flow / Total Assets	*EarningYield*
Cash Flow / Total Capital	*Shareholders' Equity / Total Market Value*
Rentabilidade	**Crescimento**
Profit BeforeTax / Sales	*Sales Growth*
Profit AfterTax / Sales	*EarningBeforeTaxGrowth*
Net Income / Sales	*EarningAfterTaxGrowth*
Cash Flow / Sales	*Net Recurring Profit Growth*
Profit AfterTax / Equity	*Operating Profit Growth*
Cash Flow / Total Market Value	*Shareholders' FundGrowth*
Profit After Tax / Cash Flow	*Total AssetsGrowth*
Alavancagem	**Liquidez a CurtoPrazo**
Debt / Equity	*CurrentAssets / CurrentLiabilities*
Total Liabilities / Total Capital	*CurrentLiabilities / Total Assets*
Total Liabilities / Shareholders' Equity	*CurrentLiabilities / Equity*
Total Assets / Shareholders' Equity	*Long Term Debt / Total Debt*
Total Assets / Total Market Value	**RetornosobreInvestimentos**
Risco	*Return on Assets*
Profit Before Tax / Current Liabilities	
Profit After Tax / Current Liabilities	
Cash Flow / CurrentLiabilities	

Quadro 1 - Indicadores financeiros utilizados por Fan e Palaniswami (2001)

Fonte: FAN; PALANISWAMI, 2001.

Os resultados foram comparados com um modelo de *benchmark* que foi determinado pelos autores como uma carteira de investimentos uniformemente ponderada composta por todas as ações disponíveis para a classificação.

Neste estudo, para reduzir o nível de ruído e manter a consistência, apenas relatórios anuais foram considerados e relatórios com mais de uma variável faltando

foram descartados. Posteriormente, pela Análise dos Componentes Principais, Fane Palaniswami (2001) agruparam os indicadores financeiros similares em oito categorias: Retorno sobre Capital, Lucratividade, Alavancagem, Investimento, Liquidez a Curto Prazo, Retorno sobre Investimento, Risco. Os dados foram convertidos em vetores de oito elementos, cada elemento representando um único principal componente de cada grupo. Como as empresas possuem ciclos de relatórios diferentes, os retornos das ações foram calculados individualmente usando dados de preços para cada 12 meses a partir da data de publicação.

O retorno esperado das ações foi definido como a variável dependente binária, podendo assumir dois valores: $+1$ que representa ações com retorno excepcional e -1 que representa as ações consideradas normais. Dessa forma, as ações que estavam entre o terceiro e quarto quantil empírico da distribuição de retornos das empresas da Bolsa de Valores Australiana foram classificadas como pertencentes à Classe 1, das melhores ações. Já aquelas que apresentaram retorno entre o primeiro e terceiro quantil empírico, constituíram a Classe 2, classe das piores ações.

Fan e Palaniswami (2001) utilizaram o método de Validação Cruzada com três anos de dados para prever o retorno futuro da ação no ano seguinte. O primeiro e segundo ano de dados foram usados para treinamento, o terceiro ano para validação e com os dados do quarto ano o modelo foi testado.

Quando o SVM foi usado para selecionar 25% das ações de cada ano, o portfólio igualmente ponderado obteve um retorno total de 208% durante um período de 5 anos, superando o desempenho *benchmark* que gerou um retorno de 71%. Portanto, o SVM mostrou-se bastante útil para seleção de ações e este resultado é corroborado também por outros estudos.

Recentemente, Huerta, Corbacho e Elkan (2013) desenvolveram um estudo similar ao de Fan e Palaniswami (2001) e ressaltam que escolheram o SVM para identificar ações com alto ou baixo retorno esperado devido a sua simplicidade e eficácia. Dois diferenciais da abordagem são o fato do SVM ter sido aplicado mensalmente para se ajustar às mudanças do mercado e a seleção dos dados que foram usados para treinar o SVM. Não foram utilizados todos os dados disponíveis, mas um conjunto de dados presentes nos quantis mais altos e baixos da distribuição histórica, também chamados de dados de cauda. Segundo os autores, a porcentagem de ações dentro do quantil escolhido é suficiente para que o SVM

aprenda as correlações entre as características da ação e a classe à qual ela pertence. Foi determinado um quantil de 20%, então 20% das ações de mais alto retorno e 20% das ações de mais baixo retorno foram escolhidas. Segundo essa abordagem, esses 40% dos dados são suficientes para o treinamento do modelo e a omissão das ações que estão no meio da distribuição alavanca o desempenho, pois é possível treinar o classificador mais rapidamente. Os dados coletados estão compreendidos no período de 1981 a 2010, sendo retirados de uma base de dados comum CRSP/Compustat.

Três filtros foram aplicados para formar a base de dados com ativos negociáveis. O primeiro é uma *proxy* para liquidez (LIQ), o segundo, uma *proxy* para o volume de dólar negociado (VDN) e por último, o preço da ação apenas. Para um determinado ativo, o cálculo da liquidez envolveu regredir os retornos de mercado sobre o volume de dólares levando em consideração o sinal do fluxo de pedidos. Já os retornos diários $r(t)$foram regredidos nos preços $p(t)$ e volume$v(t)$ de acordo com a seguinte equação:

(1)

$$r(t) = c + \lambda \,\mathrm{sinal}(t)\, \log v(t)p(t)$$

Onde $\mathrm{sinal}(t)$ assume os valores $+1$ se $r(t) \geq 0$, e -1, caso contrário. O coeficiente de regressão λ foi usado como uma *proxy* inversa da liquidez e estimado usando todas as negociações em uma janela de 91 dias. Segundo os autores, o inverso do fator de liquidez quantifica o impacto do volume de dólar negociado nas alterações de preço daquele dia. O VDN foi calculado multiplicando o volume diário pelo preço ação $v(t)p(t)$. Esse filtro elimina as ações que não possuem capacidade suficiente para serem negociadas em fundos mútuos. Os valores diários do VDN foram suavizados por uma média diária exponencial expressa por:

(2)

$$e(t) = \alpha p(t)v(t) + (1 - \alpha)e(t - 1)$$

com $\alpha = 2/(91 + 1)$. Visando simular as condições reais de negociação, os autores aplicaram os filtros todos os dias em que houve negociação antes da abertura das posições. Os limites de corte dos filtros são 50% inferiores para o VDN e preço, e 50% superiores para λ no filtro LIQ.

Se uma ação que pertencia à lista de negociáveis caiu abaixo da marca de corte durante o período de realização da carteira, a ação foi mantida até que as posições fossem fechadas. Se fosse feito o inverso, se introduziria um viés ao longo da carteira por manter na lista de ações negociáveis apenas aquelas que estavam melhores que a média. Resumindo, os autores aplicam os filtros VDN e LIQ para o modelo não aprender correlações de ações que são difíceis de serem negociadas.

Como cada setor da economia possui características únicas, os autores construíram um modelo para cada um dos seguintes setores: Energia, Materiais, Indústria, Consumo de Luxo, *Staples* do Consumidor, Saúde, Finanças, Tecnologia da Informação, Serviços de Telecomunicações e Utilitários. Os setores Serviços de Telecomunicações e Utilitários não apresentaram um número de ações suficientes para se construir o modelo, portanto, foram descartados.

As características técnicas de cada ativo foram calculadas pelo CRSP e as fundamentais foram obtidas do Compustat. A seleção das características foi feita de acordo com a popularidade destas na literatura. O Quadro 2 elenca os indicadores fundamentais utilizados:

Indicador	**Fórmula ou Variável**
Total Revenue	TR
Gross Profit	GP
OperatingIncome	OI
IncomeBeforeTax	IBT
IncomeAfterTax	IAT
Net Income Before Extraordinary Items	NIBEI
Net Income	NI
Dividends	D
Diluted Normalized Earnings Per Share	DNEPS
Cash andEquivalents	CE
Short TermInvestiments	STI
AccountsReceivable	AR
Total Inventory	TI
Total CurrentAssets	TCA
Total Assets	TA

Short TermLiabilities	STL
Total CurrentLiabilities	CL
Total LongTermDebt	TLTD
Total Debt	TD
Total Liabilities	TL
Total Equity	TE
Total Shares	TS
Depreciation	DP
Cash FromOperatingActivities	CFOA
Capital Expenditures	CEx
Cash FromInvestingActivities	CFIA
Cash FromFinancingActivities	CFFA
Net Change In Cash	NCIC
Snapshot Accrual	$\mathrm{SAC} = \mathrm{TCA} - \mathrm{CE} - \mathrm{CL} + \mathrm{TD}$
Accrual Based on Balance Sheet	$\mathrm{ABBS} = \mathrm{SAC}(quarter) - \mathrm{SAC}(quarter - 4)$
Accrual Based on Cash Flow	ABCF
Financial Health	FH
Working Capital	$\mathrm{TCA} - \mathrm{CL}$
QuickRatio	$(\mathrm{TCA} - \mathrm{TI})/\mathrm{CL}$
DividendPayoutRatio	$\mathrm{DPR} = \mathrm{D}/\mathrm{NI}$
Book Value	BV
Book Value - Total Debt	$\mathrm{BV} - \mathrm{TD}$
Receivables to Sales	AR/TR
Debt to Assets	TD/TA
Debt to Equity	TD/TE
Cash to Assets	CE/TA
Liabilities to Income	T/NI
ReturnonEquity	$\mathrm{ROE} = \mathrm{NI}/\mathrm{TE}$
Sales per Share	TR/TS

Quadro 2 - – Indicadores fundamentais utilizados por Huerta, Corbacho e Elkan (2013)

Fonte: HUERTA; CORBACHO; ELKAN, 2013

Os portfólios foram formados com a classificação dos *outputs* do SVM, sendo as ações classificadas nas posições mais altas no *ranking* utilizadas para vendas de longo prazo na carteira, e as ações em posições mais baixas, usadas para vendas de curto prazo. Foram formadas carteiras de 10 posições de longo prazo igualmente ponderadas e 10 posições de curto prazo também igualmente ponderadas. O estudo chegou a um retorno anual de 15% com volatilidade próxima a 8% para o portfólio formado.

O estudo de Emir, Dinçer e Timor (2012) teve como objetivo construir um modelo financeiro ótimo que permitisse a classificação das melhores ações do Mercado Turco. Para este propósito, anualmente, as ações que apresentaram os 10 retornos mais altos foram classificadas como “1” e as demais, classificadas como “0”. Segundo os autores, a aplicação de redução de dimensionalidade dos dados antes do processamento destes para classificação, melhora o resultado final.

Os dados foram coletados para cada ação que compunha o Índice *Istanbul Stock Exchange (ISE)*no período de 2002 a 2010 e este estudo foi inovador ao utilizar tanto parâmetros técnicos como fundamentalistas para a análise. Os dados técnicos foram 13 indicadores do Índice *Istanbul Stock Exchange (ISE)* listados no Quadro 3.

Atributo	**Fórmula ou Variável**
Growth in Assets	GA
Growth in Net Profit	GNP
EquityGrowth	EG
CurrentAssets / Assets	CA/A
FixedAssets / Assets	FA/TA
Equity / Assets	E/TA
Equity / TangibleAssets	E/TGA
Return on Assets	ROA = NI/TA
Net Profit / CurrentAssets	NP/CA
Return on Equity	ROE = NI/TE
Earnings per Share	NI − DPS/AOS
Price-EarningsRatio	MVPS/EPS
Market to Book Value	MV/BV

Quadro 3 - Indicadores técnicos utilizados por Emir, Dinçer e Timor (2012)

Fonte: EMIR; DINÇER; TIMOR, 2012.

Já a análise fundamentalista foi feita com 14 indicadoresconsiderados essenciais para representar as empresas do ISE como um todo, apesar de pertencerem a diferentes setores. Estes estão elencados no Quadro 4.

Atributo	Fórmula ou Variável
Mass Index (MASS)	$\sum_{1}^{25} \frac{9-\text{day EMA of (High} - \text{Low)}}{9-\text{day EMA of a } 9-\text{day EMA of (High} - \text{Low)}}$
AverageTrue Range (ATR)	$\text{ATR}_t = \frac{\text{ATR}_{t-1} \times (n-1) + \text{TR}_t}{n}$ O primeiro ATR é calculado usando a seguinte média aritmética: $\text{ATR}_t = \frac{1}{n}\sum_{i=1}^{n} TR_i$
Momentum (Mo)	$C_t - C_{t-4}$
Chaikin Money Flow Indicator (CMF)	$CMF = \frac{\sum_{t=20}^{t} CLV_t \times volume_t}{\sum_{t=20}^{t}(vol_t)}$ onde, $CLV = \frac{(close_1 - low_1) - (high_1 - close_1)}{(high_1 - low_1)}$
Commodity Channel Index (CCI)	$\frac{(M_t - SM_t)}{(0.015\ D_t)}$ onde: $M_t = (\frac{H_t + L_t + C_t}{3})$ $SM_t = \frac{\sum_{i=1}^{n} M_{t-i+1}}{n}$
Moving Average Convergence-Divergence Trading Method (MACD)	$MCD = 2(DIF - DEA)$ $DIF = EMA(12) - EMA(26)$ $DEA_t = \frac{2}{10}dif + \frac{8}{10}DEA_{t-1}$ $EMA_t(n) = \frac{2}{N+1}C_t + \frac{N-1}{N} + 1MA_{t-1}(n)$
ExponentialMovingAverage (EMA)	$EMA_{todau} = EMA_{yesterday} + \alpha(price_{todau} - EMA_{yesterday})$
RelativeStrength Index (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i}/n) \,/\, (\sum_{i=0}^{n-1} Dw_{t-i}/n)}$
Money Flow Index (MFI)	$100 - \frac{100}{(1 + \text{Money Flow Ratio})}$
Stochastics %K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$
Triple Exponential Smoothing of the Log of Closing Price (TRIX)	$TripleEMA_0 = (1-f)^3(p_0 + 3fp_1 + 6f^2p_2 + 10f^3p_3 + \cdots)$

William's %R	$\frac{H_n - C_t}{H_n - L_n} \times 100$

Quadro 4 - Indicadores fundamentalistas utilizados por Emir, Dinçer e Timor (2012)

Fonte: Elaborado pela autora.

Para fins de comparação, um modelo de Rede Neural foi aplicado nas mesmas circunstâncias e os resultados mostraram que as Máquinas de Suporte Vetorial apresentaram desempenho superior na acurácia da previsão. Portanto, os resultados empíricos do estudo de Emir, Dinçer e Timor (2012) também corroboram para o sucesso do SVM como modelo para previsão em séries temporais financeiras.

No mesmo âmbito de abordagens para construção de portfólios, Gupta, Mehlawat e Mittal (2012) desenvolveram uma abordagem híbrida para facilitar as tomadas de decisão dos investidores. Primeiramente, utilizaram as Máquinas de Suporte Vetorial para classificar as ações em três classes pré-definidas de acordo com o desempenho delas em três indicadores financeiros: liquidez, retorno e risco.

Como retorno do portfólio considerou-se o retorno a curto prazo, equivalente ao desempenho médio dos ativos no período de 12 meses, e o retorno a longo prazo, também equivalente ao desempenho médio dos ativos, mas para um período de 36 meses. O risco da carteira foi definido como o desvio semi-absoluto de rentabilidade do portfólio abaixo do retorno esperado. Já a liquidez, foi considerada como a probabilidade de conversão de um investimento em dinheiro, sem qualquer perda significativa de valor e medida através da taxa de *turnover*.

Ativos da Classe 1 foram classificados como ativos líquidos, já que o indicador de liquidez foi o mais alto nesta classe. Ativos da Classe 2 foram classificados como de alto rendimento, uma vez que apresentaram altos retornos. Ativos da Classe 3 foram classificados como ativos de menor risco, visto que em comparação com as demais classes, esses ativos apresentaram o menor desvio padrão, mas retorno e liquidez médios.

A base de dados foi composta por 150 ativos listados no *National Stock Exchange (NSE*), o principal mercado de ativos financeiros da Índia. O conjunto de treinamento foi composto por 60% do total dos dados e o conjunto de teste por 40%. O segundo passo do estudo foi a aplicação de um algoritmo genético, mais

especificamente, *Real CodedGeneticAlgorithm* (RCGA), em cada uma das três classes para formação de portfólios ótimos. O portfólio formado a partir das ações da Classe 1 apresentou maior liquidez, mas um nível de risco médio. O portfólio formado a partir da Classe 2 apresentou maior nível de retorno e maior nível de risco. Já o portfólio da Classe 3 apresentou o menor nível de risco comparado aos demais portfólios, e como esperado, um nível de retorno médio. Sendo assim, os autores concluem que investidores à procura de maior liquidez deveriam investir em ativos da Classe 1. Já os investidores à procura de maiores retornos deveriam optar pela Classe 2 e àqueles à procura de investimentos mais seguros deveriam investir em ativos da Classe 3. Estes resultados indicam que a abordagem desenvolvida é capaz de classificar os ativos com boa acurácia e ainda mais, é capaz gerar portfólios otimizados para cada classe de ativos de acordo com as preferências dos consumidores.

As Máquinas de Suporte Vetorial também podem ser aplicadas na previsão da direção do mercado. Kim (2003) analisou a aplicabilidade das Máquinas de Suporte Vetorial na previsão da direção das alterações diárias nos preços das ações em comparação com dois modelos: BPN (*Back-Propagation Neural Network)* e CBR (*Case-basedreasoning*).Os dados utilizadosforam as observações diárias dos preços das ações que compõem o Índice de Mercado da Coreia (KOSPI) e 12 indicadores técnicos para período de Janeiro de 1989 a Dezembro de 1998. Os indicadores selecionados como *inputs* estão descritos no Quadro 5.

Nome do Atributo	**Fórmula**
Stochastics %K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$
%D (3-period moving average of %K)	$\frac{\sum_{i=0}^{\%n-1} \%K_{t-i}}{n}$
Slow %D	$\frac{\sum_{i=0}^{\%n-1} \%D_{t-i}}{n}$
Momentum	$C_t - C_{t-4}$
Price Rate of Change (ROC)	$\frac{C_t}{C_{t-n}} \times 100$
William's %R	$\frac{H_n - C_t}{H_n - L_n} \times 100$

A/D Oscilador	$\frac{H_t - C_{t-1}}{H_t - L_t}$
Disparity5	$\frac{C_t}{MA_5} \times 100$
Disparity10	$\frac{C_t}{MA_{10}} \times 100$
PriceOscilator (OSCP)	$\frac{M_5 - M_{10}}{MA_5}$
Commodity Channel Index (CCI)	$\frac{(M_t - SM_t)}{(0.015\ D_t)}$ onde: $M_t = (\frac{H_t + L_t + C_t}{3})$ $SM_t = \frac{\sum_{i=1}^{n} M_{t-i+1}}{n}$ $D_t = \frac{\sum_{i=1}^{n} \lvert M_{t-i+1} - SM_t \rvert}{n}$
RelativeStrength Index (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i}/n) \,/\, (\sum_{i=0}^{n-1} Dw_{t-i}/n)}$

Quadro 5 - Atributos selecionados por Kim (2003)

Fonte: KIM, 2003.

O autor classificou as alterações diárias dos preços em duas classes: “0” ou “1”. A primeira classe foi composta por ações cujo preço do dia posterior foi menor do que do dia anterior. Já a segunda classe foi composta por ações cujo índice no dia posterior foi mais alto se comparado ao dia anterior. 80% dos dados foram usados para treinamento e estimação dos parâmetros e os 20% restantes foram utilizados para validação do modelo. Para desenvolver os experimentos, o software LIBSVM foi utilizado.

O desempenho na predição das séries temporais P, foi avaliada usando a seguinte equação:

(3)

$$\mathrm{P} = \frac{1}{m}\sum_{i=1}^{m} R_i \ (i = 1,2, \dots, m)$$

Onde R_i é o resultado da predição para o *i*-ésimo dia de negociação e é definido por:

$$R_i = \begin{cases} 1\ se\ PO_i = AO_i \\ 0\ caso\ contrário, \end{cases}$$

PO_i é o valor previsto do *output* para o *i*-ésimo dia de negociação, AO_ié o *output* atual para o *i*-ésimo dia de negociação e *m* é a quantidade de exemplos de teste.

Os resultados empíricos mostram que no conjunto de validação, o SVM obteve um desempenho na previsão de 57,83% contra 54,73% e 51,97% dos modelos BPN e CBR, respectivamente. Fica evidente então que as Máquinas de Suporte Vetorial superaramos dois modelos no nível de acurácia da previsão e isso pode ser atribuído ao fato de que o SVM implementa o Princípio da Minimização do Risco Estrutural, permitindo uma melhor generalização. Este estudo concluiu que o SVM é uma alternativa promissora em previsão de séries temporais financeiras.

Zhang e Zhao (2009) por sua vez, aplicaram o SVM no mercado de câmbio para prever mudanças nas taxas de câmbio euro/dólar. Neste estudo, os *inputs* para o modelo foram indicadores técnicos, sendo os dados oriundos do sistema Bloomberg no intervalo de 10 de julho de 2007 a 9 de julho de 2009. Os indicadores utilizados assim como suas fórmulas estão representados noQuadro 6.

Indicador	Fórmula
Moving Average Line	$MA_t(n) = \frac{1}{N}C_t + \frac{N-1}{N}MA_{t-1}(n)$
Moving Average Convergence and Divergence Line	$MCD = 2(DIF - DEA)$ $DIF = EMA(12) - EMA(26)$ $DEA_t = \frac{2}{10}dif + \frac{8}{10}DEA_{t-1}$ $EMA_t(n) = \frac{2}{N+1}C_t + \frac{N-1}{N} + 1MA_{t-1}(n)$
Random Index	$K_t = \frac{2}{3}K_{t-1} + \frac{1}{3}RSV_t$ $D_t = \frac{2}{3}D_{t-1} + \frac{1}{3}K_t$ $J = 3D - 2K$ $RSV_t = \frac{C_t - L_n}{H_n - L_t} \times 100\%$
Relative Strength Index	$RSI(n) = \frac{A}{A+B} \times 100\%$
BIAS	$BIAS(n) = \frac{C_t - MA(n)}{MA(n)} \times 100\%$

Quadro 6 - Técnicos selecionados por Zhang e Zhao (2009)

Fonte: ZHANG; ZHAO, 2009.

Os autores ressaltam que as mudanças na taxa de câmbio dependem dos ajustes políticos do governo, portanto, os preços das ações estão intimamente relacionados aos resultados de suas análises e, genericamente, pode-se dizer que a mudança de uma taxa de câmbio está mais próxima de um processo estocástico. Para os operadores do Mercado determinarem se o preço de uma taxa de câmbio subirá ou cairá, eles precisam de muitas experiências e um vasto conhecimento sobre indicadores e sobre seu comportamento recente para somente depois tomarem uma decisão. As Máquinas de Suporte Vetorial por sua vez, precisam estudar dados históricos das taxas de câmbio e estabelecer um modelo de classificação, sendo este um processo bem menos oneroso.

Segundo os autores, análises de indicadores técnicos de câmbio podem ser descritos como um problema de aprendizagem geral. Primeiramente é preciso reconhecer se a análise técnica é válida, isto é, se os indicadores técnicos e a tendência da taxa de câmbio têm alguma ligação intrínseca. Se houver esse tipo de relacionamento, a chave do problema é achar uma função que minimize o Risco Esperado e seja aplicável em uma grande amostra. Nesse contexto, os *inputs* são os indicadores técnicos e os *outputs,* que indicam a mudança no preço futuro, derivam do relacionamento entre os indicadores e a tendência da taxa de câmbio. Entretanto, a escolha dos indicadores não é uma tarefa fácil.

Como na maioria dos estudos que utilizam Máquinas de Suporte Vetorial, Zhang e Zhao (2009) classificam o *output* do modelo em duas classes: Classe 1, composta pelas observações em que houve aumento no preço, isto é, $y_i = +1$, e Classe 2, formada pelas observações em que houve queda no preço, ou seja, $y_i = -1$. Os dias em que não houve variação no preço foram ignorados. Os resultados empíricos mostraram que a precisão da previsibilidade do SVM é maior que 60%. Sendo assim, Zhang e Zhao (2009) chegaram à conclusão de que com o SVM, é possível fazer previsões independente da complexidade do Mercado Financeiro.

Observa-se que há uma diversidade de indicadores que podem ser usados como insumos para as Máquinas de Suporte Vetorial. O Apêndice A deste trabalho traz a compilação de todos os indicadores utilizados nos estudos apresentados.

3 MÉTODOS E TÉCNICAS DE PESQUISA

Esta pesquisa seguiu a metodologia de Máquinas de Suporte Vetorial proposta por Fan e Palaniswami (2001) aplicada ao Mercado Financeiro Brasileiro.

3.1 Tipo e descrição geral da pesquisa

A presente pesquisa é considerada Correlacional Quantitativa, visto que teve como objetivo analisar como o retorno de um portfólio formado por SVM se comporta, conhecendo o comportamento de outros métodos de seleção de ações. Este é um estudo de campo com dados secundários.

A variável dependente é o retorno futuro da ação, sendo esta uma variável discreta binária $y = \pm 1$, onde$+1$ representa ações com retornos futuros excepcionais e -1 representa as ações consideradas normais.As variáveis independentes são os preços das ações e informações do mercado, coletados na base de dados do Economática. Assim como no estudo de Fan e Palaniswami (2001), indicadores financeiros foram utilizados como insumos para o SVM, e estes foram obtidos por meio do Economática. O Quadro 7 elenca os indicadores utilizados assim como a sua fórmula e classificação, segundo o Economática.

Classificação	Indicador	Fórmula
Dados por Ação	LucroporAção	$\frac{\mathrm{LL} - \mathrm{DAP}}{\mathrm{MAC}}$
	Valor Patrimonial porAção	$\frac{\mathrm{PL}}{\mathrm{MAC}}$
Estrutura de Capital	Exigível Total / Ativo Total (%)	$\frac{\mathrm{AT} - \mathrm{PL} - \mathrm{PAM}}{\mathrm{AT}} \times 100$
	Exigível Total / PatrimônioLíquido (%)	$\frac{\mathrm{AT} - \mathrm{PL} - \mathrm{PAM}}{\mathrm{PL} + \mathrm{PAM}} \times 100$
	AtivoFixo / PatrimônioLíquido (%)	$\frac{\mathrm{I}}{\mathrm{PL} + \mathrm{PAM}} \times 100$

Liquidez	LiquidezGeral	$\frac{\text{AC} + \text{RLP}}{\text{PC} + \text{PNC}}$
	LiquidezCorrente	$\frac{\text{AC}}{\text{PC}}$
Rentabilidade	Giro do Ativo	$\frac{\text{VL}}{\text{AT}}$
	Giro do PatrimônioLíquido	$\frac{\text{VL}}{\text{PPA}}$
	MargemBruta (%)	$\frac{\text{LB}}{\text{ROL}} \times 100$
	MargemLíquida (%)	$\frac{\text{LL} + \text{PAM}}{\text{ROL}} \times 100$
	Rentabilidade do Ativo (%)	$\frac{\text{LL} + \text{PAM}}{\text{AT}} \times 100$
	Rentabilidade Patrimonial Final (%)	$\frac{\text{LL} + \text{PAM}}{\text{PL} + \text{PAM}} \times 100$
	Rentabilidade Patrimonial Média (%)	$\frac{\text{LL} + \text{PAM}}{\text{PL (no início do período)} + \text{PAM (no início do período)} + \text{PL} + \text{PAM}} \times 100$
	Rentabilidade Patrimonial Inicial (%)	$\frac{\text{LL} + \text{PAM}}{\text{PL (no início do período)} + \text{PAM (no início do período)}} \times 100$

Quadro 7 - Indicadores financeiros utilizados na pesquisa

Fonte: Elaborado pela autora.

Segundo Downes e Goodman (2002), os indicadores podem ser interpretados da seguinte forma:

O indicador Lucro por Ação indica a parte do lucro de uma empresa alocada em cada ação do capital ordinário, sendo um indicador da rentabilidade da ação. Já o indicador Valor Patrimonial por Ação evidencia o valor do patrimônio líquido sobre o número total de ações em circulação. Se comparado com o preço corrente da ação, pode ser usado como um parâmetro para indicar se a ação está subavaliada. Entretanto, essa métrica não pode ser usada sozinha, pois representa uma visão bem limitada da situação da empresa.

A taxa Exigível Total / Ativos Total inclui obrigações de longo e curto prazo, assim como ativos tangíveis e intangíveis, definindo a proporção de ativos de uma

empresa que está sendo financiada com dívida ao invés de capital próprio. Quanto maior a taxa, maior o risco financeiro. Já a taxa Exigível Total / Patrimônio Líquido, é basicamente a divisão das obrigações totais pelo patrimônio líquido dos acionistas, indicando qual a proporção de capital próprio e dívida que a empresa está usando para financiar seus ativos. Quando o valor da taxa for alto, pode se inferir que a empresa tem financiado em grande medida o seu crescimento com a dívida e isso pode resultar em ganhos voláteis devido às despesas de juros adicionais. O indicador Ativo Fixo / Patrimônio Líquido por sua vez, mede a contribuição do patrimônio líquido dos acionistas e a contribuição das fontes de dívida no ativo imobilizado da empresa. Se o valor final for maior que 1, significa que o patrimônio liquido é menor que o ativo imobilizado e que a empresa tem usado dívidas para financiar parte do ativo imobilizado. Por outro lado, se for inferior a 1, significa que o patrimônio líquido é maior que os ativos imobilizados e que o patrimônio líquido está financiando não apenas os ativos fixos, mas também uma parte do capital de giro.

O indicador de Liquidez Geral aponta o quanto a empresa possui em dinheiro, bens e direitos realizáveis a curto e longo prazo. Entretanto, para se avaliar a capacidade de pagamento da empresa de fato, é preciso observar a estrutura de prazos e o ciclo operacional. Já a Liquidez Correnterelaciona quantos reais a empresa possui disponível de forma imediata para conversão em dinheiro com as dívidas de curto prazo. Se o resultado do quociente for maior que 1, a empresa possui uma folga de recursos para uma possível liquidação das obrigações.Se for igual a 1, os valores dos direitos e obrigações a curto prazo são equivalentes. Porém, se for menor que 1, este é um indicativo de que não haveria disponibilidade suficiente para quitar as obrigações a curto prazo, caso fosse preciso.

O Giro do Ativo mede a quantidade de receitas geradas por cada unidade monetária de ativo, sendo assim um indicador de eficiência da alocação dos ativos da empresa. Em geral, quanto maior for essa taxa, melhor, entretanto, comparações com essa taxa devem ser feitas apenas entre companhias de um mesmo setor. O Giro do Patrimônio Líquido por sua vez, mede a habilidade da empresa de gerar vendas dado os investimentos na participação patrimonial dos acionistas ordinários e preferenciais. Se for igual a 10, por exemplo, essa taxa indica que para cada unidade monetária investida na participação patrimonial, a empresa gera 10 dólares de renda.

O indicador de Margem Bruta apresenta quanto a empresa obtém de retorno das vendas, retirando os custos das mercadorias vendidas e serviços prestados. Então, quanto maior a margem, maior a rentabilidade. Já a Margem Líquida mostra qual o lucro líquido gerado para cada unidade de venda realizada na empresa, sendo um bom indicador da margem operacional.

A taxa Rentabilidade Patrimonial (Pat final) indica o montante de lucro líquido como uma porcentagem do patrimônio líquido no fim do período. Ou seja, é uma medida de rentabilidade da empresa à medida que mostra o quanto de lucro líquido foi gerado com o dinheiro que os acionistas investiram. Este patrimônio líquido não incluiu a participação de acionistas preferenciais. As taxas Rentabilidade Patrimonial (Pat médio) e Rentabilidade Patrimonial (Pat inicial) também podem ser interpretadas da mesma forma, diferindo apenas no fato de que o lucro líquido é indicado como uma porcentagem do patrimônio líquido no fim, meio e no início do período, respectivamente.

Neste estudo, o SVM foi o classificador, ou seja, o modelo que classificou o desempenho das ações por meio do aprendizado com os dados históricos e a porcentagem de acertos foi utilizada como medida de acurácia. Entretanto, para o problema de seleção de ações, apenas a acurácia de previsão não é um indicador suficiente da capacidade de classificação do modelo. Por este motivo, nesta pesquisa, assim como em Fan e Palaniswami (2001), para comprovar o desempenho do SVM, o retorno do portfólio igualmente ponderado escolhido pela máquina foi comparado com o retorno gerado por um modelo*benchmark*, definido como o fundo de índice BOVA11. Este ETF (*Exchange TradedFund*) é um fundo de investimento que tem comoobjetivo obter uma taxa de rentabilidade semelhante ao desempenho do Ibovespa e suas cotas são negociadas na BM&FBOVESPA por meio do código BOVA11.Diferentemente do estudo de Fan e Palaniswami (2001), o SVM foi utilizado para selecionar portfólios a cada trimestre e não a cada ano.

Segundo Bruni, Fuentes e Famá (1997), a maximização da eficiência dos portfólios pode ser obtida através da razão retorno/risco, então, a relação retorno/risco de ambas opções de investimento também foram calculadas a fim de se criar mais um parâmetro de comparação.

Posteriormente, o método *Bootstrap* foi aplicado para controlar os efeitos de *Data Snooping* e avaliar a significância estatística dos resultados encontrados.

3.2 Caracterização da organização, setor ou área

Este projeto de pesquisa estudou a aplicabilidade das Máquinas de Suporte Vetorial para previsão do retorno das ações negociadas na BM&FBOVESPA. No Mercado de Capitais do Brasil existiam duas grandes bolsas: a BOVESPA (Bolsa de Valores de São Paulo e a BM&F (Bolsa de Mercados e Futuros). Em 2008, foi criada a BM&FBOVESPA.SA (Bolsa de Valores, Mercadorias e Futuros), fruto da fusão dessas duas bolsas.

A BM&FBOVESPA.SA é líder na América Latina e está entre as maiores bolsas do mundo em valor de mercado. Ela é a principal instituição de intermediação no Brasil para operações no mercado de capitais e provê sistemas de negociação de ações, derivativos de ações, derivativos financeiros, títulos de renda fixa, títulos públicos federais, moedas à vista e *commodities* agropecuárias. Como única bolsa de valores, mercadorias e futuros em operação no Brasil, a BM&FBOVESPA tem o papel de fomentar o Mercado Brasileiro por meio de inovações e desenvolvimento de produtos, além de programas de educação para a população. Sediada na cidade de São Paulo, a BM&FBOVESPA possui escritório de representação nos Estados Unidos (Nova York), no Reino Unido (Londres) e na China (Xangai), para oferecer suporte aos participantes daqueles mercados nas atividades com os clientes estrangeiros e no relacionamento com os órgãos reguladores, além de divulgar seus produtos e práticas de governança a potenciais investidores.

Para ser listada na BM&FBOVESPA, a empresa precisa abrir o capital e assim ter suas ações negociadas publicamente. O primeiro passo é protocolar um pedido de registro da companhia aberta na Comissão de Valores Mobiliários (CVM), órgão regulador e fiscalizador do Mercado de Capitais Brasileiro. De posse dessa autorização, a empresa pode abrir seu capital com ou sem oferta de ações no Mercado. A Oferta Pública Inicial é a primeira colocação pública de títulos no mercado e como estes ativos estão sendo negociados pela primeira vez, se diz que essa operação acontece no mercado primário. Os recursos obtidos são destinados ao caixa da empresa, dado que o vendedor é a própria companhia. A partir da negociação dessas ações pela segunda vez, ou seja, na segunda, terceira, *n*-ésima negociação do produto, ela estará sendo feita no mercado secundário. No mercado

secundário, quem vende é o acionista, portanto a empresa não recebe nenhuma parcela dos recursos originados na venda. Apesar disso, a empresa tem total interesse que suas ações sejam negociadas e o preço delas suba. Primeiramente porque emissões primárias de ações e admissão de novos sócios pelas operações no mercado secundário configuram uma fonte de financiamento que não tem limitação. Enquanto a empresa tiver projetos viáveis e rentáveis, existirão investidores interessados em financiá-los. Além disso, a cotação de suas ações no mercado acionário é um indicativo de valor, pois reflete a percepção de muitos investidores sobre as perspectivas futuras da companhia. Portanto, a empresa ganha visibilidade ao ser regularmente mencionada na mídia e acompanhada pela comunidade financeira.

Este estudo se ateve mais especificamente às ações que compõem o índice Ibovespa. "O Ibovespa é resultado de uma carteira teórica de ativos, elaborada de acordo com os critérios estabelecidos em sua metodologia" (METODOLOGIA DO ÍNDICE BOVESPA, 2014, p.2). Ele foi criado em 02/01/1968 e mede a rentabilidade média das cotações das ações de maior negociabilidade e representatividade negociadas na Bovespa. O Ibovespa não reflete apenas a variação do preço das ações como também o impacto da distribuição dos proventos, sendo, dessa forma, considerado um indicador do retorno total das ações que o compõem.

3.3 População e amostra

A população dessa pesquisa consistiu em todas as ações disponíveis para negociação na BM&FBOVESPA e ações que compuseram o Ibovespa no período de 2000 a 2013 constituíram a amostra. Visto que a carteira teórica do Ibovespa não é fixa, sendo ajustada a cada quadrimestre, a amostra deste estudo se ateve à carteira teórica válida para 2 de Setembro de 2013 a 03 de Janeiro de 2014, por essa ser a mais recente do recorte temporal.

3.4 Caracterização dos instrumentos de pesquisa

Para análise dos dados este estudo utilizou o *software*livre R. Este instrumento foi criado inicialmente por Robert Gentleman e Ross Ihaka do Departamento de Estatística da Universidade de Auckland, Nova Zelândia. R é uma linguagem e um ambiente para computação que fornece uma ampla variedade de técnicas estatísticas e gráficas como, modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, *clustering.*O R é altamente expansível com o uso de pacotes, que são bibliotecas para funções específicas. Neste estudo foi utilizado o pacote *kernlab*, útil em métodos de aprendizagem em máquina baseados em *Kernel* para classificação, regressão, *clustering* e redução de dimensionalidade.

3.5 Procedimentos de coleta e de análise de dados

A primeira etapa da coleta de dados foi a busca dos indicadores financeiros apresentados no Quadro 7 no período de 2000 e 2013 para as ações que compuseram o Ibovespa na carteira teórica do último quadrimestre de 2013. A Tabela 1, elenca todas as 71 ações, bem como o percentual de sua participação na carteira.

Tabela 1- Carteira Teórica Ibovespa para 2 de Set. de 2013 a 3 de Jan. de 2014

(continua)

Código	Ação	Participação (%)
ABEV3	AMBEV S/A	5,540
AEDU3	ANHANGUERA	0,815
ALLL3	ALL AMER LAT	0,394
BBAS3	BRASIL	2,428
BBDC3	BRADESCO	1,671
BBDC4	BRADESCO	7,382
BBSE3	BBSEGURIDADE	2,316
BISA3	BROOKFIELD	0,043
BRAP4	BRADESPAR	0,459

(continuação)

Código	Ação	Participação (%)
BRFS3	BRF SA	3,460
BRKM5	BRASKEM	0,418
BRML3	BR MALLS PAR	0,881
BRPR3	BR PROPERT	0,418
BVMF3	BMFBOVESPA	2,315
CCRO3	CCR SA	1,698
CESP6	CESP	0,523
CIEL3	CIELO	3,167
CMIG4	CEMIG	1,404
CPFE3	CPFL ENERGIA	0,632
CPLE6	COPEL	0,358
CRUZ3	SOUZA CRUZ	0,938
CSAN3	COSAN	0,459
CSNA3	SID NACIONAL	0,699
CTIP3	CETIP	0,836
CYRE3	CYRELA REALT	0,262
DTEX3	DURATEX	0,232
ECOR3	ECORODOVIAS	0,322
ELET3	ELETROBRAS	0,164
ELET6	ELETROBRAS	0,249
ELPL4	ELETROPAULO	0,099
EMBR3	EMBRAER	1,570
ENBR3	ENERGIAS BR	0,263
ESTC3	ESTACIO PART	0,898
EVEN3	EVEN	0,157
FIBR3	FIBRIA	0,514
GFSA3	GAFISA	0,105
GGBR4	GERDAU	1,248
GOAU4	GERDAU MET	0,468
GOLL4	GOL	0,134
HGTX3	CIA HERING	0,316
HYPE3	HYPERMARCAS	0,667
ITSA4	ITAUSA	3,002
ITUB4	ITAUUNIBANCO	9,692
JBSS3	JBS	1,302
KLBN11	KLABIN S/A	0,536

(conclusão)

Código	Ação	Participação (%)
KROT3	KROTON	1,328
LAME4	LOJAS AMERIC	0,636
LIGT3	LIGHT S/A	0,231
LREN3	LOJAS RENNER	0,907
MMXM3	MMX MINER	0,018
MRFG3	MARFRIG	0,226
MRVE3	MRV	0,247
NATU3	NATURA	0,733
OIBR4	OI	0,668
PCAR4	P.ACUCAR-CBD	1,806
PDGR3	PDG REALT	0,226
PETR3	PETROBRAS	5,120
PETR4	PETROBRAS	8,116
QUAL3	QUALICORP	0,500
RENT3	LOCALIZA	0,580
RSID3	ROSSI RESID	0,065
SANB11	SANTANDER BR	1,554
SBSP3	SABESP	0,839
SUZB5	SUZANO PAPEL	0,406
TBLE3	TRACTEBEL	0,735
TIMP3	TIM PART S/A	1,135
UGPA3	ULTRAPAR	1,845
USIM5	USIMINAS	0,406
VALE3	VALE	3,825
VALE5	VALE	5,124
VIVT4	TELEF BRASIL	1,270
VIVT4	TELEF BRASIL	1,270

Fonte:BM&FBOVESPA. Carteira Teórica Ibovespa válida para 15/6/2014. Disponível em: <http://www.bmfbovespa.com.br/indices/ResumoCarteiraTeorica.aspx?Indice=Ibovespa&idioma=pt-br>. Acesso em: 15 de Junho de 2014.

Fan e Palaniswami (2001), agruparam os 37 indicadores financeiros utilizados em seu estudo pela técnica de Análise dos Componentes principais, a fim de agrupar os indicadores similares em um único conjunto, o que resumiu e diminuiu o volume de dados. Entretanto, como nesta pesquisa foram utilizados apenas 15 indicadores financeiros, ou seja, menos da metade do total utilizado por Fan e

Palaniswami (2001),o agrupamento destes não agregaria muito valor e por isso tal técnica não foi utilizada.

Os dados dos indicadores financeiros coletados são trimestrais, sendo que, para a maioria das empresas da carteira, as datas de divulgação dos resultados estavam organizadas da seguinte forma: 31/03 para o primeiro trimestre, 30/06 para o segundo trimestre, 30/09 para o terceiro trimestre e 31/12 para o quarto trimestre. A empresa COSAN (CSNA3) foi a única que não continha os dados dispostos nesse padrão de datas ao longo de todo o recorte temporal, pois os resultados dessa empresa referentes ao quarto trimestre não foram divulgados no mês de dezembro do mesmo ano, mas sim, em 31/01 do ano seguinte. Assim, de 2005 a 2009 as suas datas de divulgação foram: 31/01 para os resultados do quarto trimestre do ano anterior, 30/04, para o primeiro trimestre do ano corrente, 31/07, para o segundo trimestre e 31/10 para o terceiro trimestre. A partir de 2009, a empresa passou a divulgar os resultados trimestrais nas mesmas datas das demais empresas da carteira. Assim, para análise dos dados, fez-se necessário o ajuste temporal dos trimestres de 2005 a 2009 da ação CSAN3.

A base de dados do Economática possui algumas limitações. Entre elas está o fato de que nem todos os indicadores financeiros são calculados para todas as ações, decorrente na maioria das vezes da não obrigatoriedade da divulgação de certos demonstrativos contábeis a partir dos quais os indicadores são calculados.Portanto, os 15 indicadores utilizados nesta pesquisa foram selecionados após uma exaustiva análise de quais indicadores estavam disponíveis para a grande maioria das empresas da carteira no recorte temporal delimitado.

Além da dificuldade de encontrar todos os indicadores financeiros para todas as empresas da carteira, a base de dados do Economática não contém o histórico dos indicadores financeiros de todos os anos do recorte temporal para todas elas. Dois casos críticos foram as ações BBSE3 e QUAL3 que apresentaram dados históricos de apenas dois e três anos, respectivamente, e por isso, foram descartadas.

A próxima etapa foi a coleta no Economática do histórico de cotações das ações em questão. Mais uma vez, para algumas ações o histórico completo das cotações não estava disponível e, por essa razão, as ações FIBR3 e KLBN11 também tiveram que ser descartadas. É importante ressaltar que os dados de cotações foram coletados já convertidos peloEconomáticaao poder aquisitivo da

data da informação mais recente da série do índice de inflação utilizado. No Apêndice B desta pesquisa encontra-se um manual com instruções para coletas de dados neste sistema. Depois dos descartes feitos, a carteira final desta pesquisa foi composta de 67 ações, listadas na Tabela 2.

Tabela 2 - Carteira de ações utilizada na pesquisa

(continua)

Código	Ação	Participação (%)
ABEV3	AMBEV S/A	5,540
AEDU3	ANHANGUERA	0,815
ALLL3	ALL AMER LAT	0,394
BBAS3	BRASIL	2,428
BBDC3	BRADESCO	1,671
BBDC4	BRADESCO	7,382
BISA3	BROOKFIELD	0,043
BRAP4	BRADESPAR	0,459
BRFS3	BRF SA	3,460
BRKM5	BRASKEM	0,418
BRML3	BR MALLS PAR	0,881
BRPR3	BR PROPERT	0,418
BVMF3	BMFBOVESPA	2,315
CCRO3	CCR SA	1,698
CESP6	CESP	0,523
CIEL3	CIELO	3,167
CMIG4	CEMIG	1,404
CPFE3	CPFL ENERGIA	0,632
CPLE6	COPEL	0,358
CRUZ3	SOUZA CRUZ	0,938
CSAN3	COSAN	0,459
CSNA3	SID NACIONAL	0,699
CTIP3	CETIP	0,836
CYRE3	CYRELA REALT	0,262
DTEX3	DURATEX	0,232
ECOR3	ECORODOVIAS	0,322
ELET3	ELETROBRAS	0,164
ELET6	ELETROBRAS	0,249
ELPL4	ELETROPAULO	0,099

(continuação)

Código	Ação	Participação (%)
EMBR3	EMBRAER	1,570
ENBR3	ENERGIAS BR	0,263
ESTC3	ESTACIO PART	0,898
EVEN3	EVEN	0,157
GFSA3	GAFISA	0,105
GGBR4	GERDAU	1,248
GOAU4	GERDAU MET	0,468
GOLL4	GOL	0,134
HGTX3	CIA HERING	0,316
HYPE3	HYPERMARCAS	0,667
ITSA4	ITAUSA	3,002
ITUB4	ITAUUNIBANCO	9,692
JBSS3	JBS	1,302
KROT3	KROTON	1,328
LAME4	LOJAS AMERIC	0,636
LIGT3	LIGHT S/A	0,231
LREN3	LOJAS RENNER	0,907
MMXM3	MMX MINER	0,018
MRFG3	MARFRIG	0,226
MRVE3	MRV	0,247
NATU3	NATURA	0,733
OIBR4	OI	0,668
PCAR4	P.ACUCAR-CBD	1,806
PDGR3	PDG REALT	0,226
PETR3	PETROBRAS	5,120
PETR4	PETROBRAS	8,116
RENT3	LOCALIZA	0,580
RSID3	ROSSI RESID	0,065
SANB11	SANTANDER BR	1,554
SBSP3	SABESP	0,839
SUZB5	SUZANO PAPEL	0,406
TBLE3	TRACTEBEL	0,735
TIMP3	TIM PART S/A	1,135
UGPA3	ULTRAPAR	1,845
USIM5	USIMINAS	0,406
VALE3	VALE	3,825

(conclusão)

Código	Ação	Participação (%)
VALE5	VALE	5,124
VIVT4	TELEF BRASIL	1,270

Fonte: Elaborado pela autora.

Dispondo da base de dados das cotações diárias de cada ação, o retorno diário líquidopara cada uma delas foi calculado com a fórmula (4), utilizando os respectivos preços de fechamento ajustados a splits e dividendos.

(4)

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} - 1 = \frac{P_t}{P_{t-1}} - 1$$

Sendo P_t o preço de fechamento de um dia e P_{t-1},o preço de fechamento do dia anterior a este. Para que os retornos dos ativos tivessem a mesma periodicidade dos indicadores financeiros, calculou-se o retorno trimestral bruto, que equivale ao produtório dos retornos diários líquidos compreendidos naquele trimestre, somados de 1, conforme a seguinte fórmula:

(5)

$$\mathrm{R_{t\ bruto}} = \left(\frac{\mathrm{P_t}}{\mathrm{P_{t-1}}} + 1\right) \times \left(\frac{\mathrm{P_{t-1}}}{\mathrm{P_{t-2}}} + 1\right) \times \left(\frac{\mathrm{P_{t-2}}}{\mathrm{P_{t-3}}} + 1\right) \times \left(\frac{\mathrm{P_{t-3}}}{\mathrm{P_{t-4}}} + 1\right) \times \cdots \times \left(\frac{\mathrm{P_{t-k+1}}}{\mathrm{P_{t-k}}} + 1\right)$$

Ou seja,

(6)

$$\mathrm{R_{t\ bruto}} = \prod \left(\frac{\mathrm{P_t}}{\mathrm{P_{t-k}}} + 1\right)$$

Posteriormente, foi calculado o retorno trimestral líquido, que é o próprio retorno trimestral bruto, subtraído de um, como mostra a fórmula (7):

(7)

$$R_{t\ líquido} = \left(\frac{P_t}{P_{t-1}} + 1\right) - 1$$

O próximo passo foi a junção dos dados históricos de indicadores financeiros e dos retornos calculados em uma única base de dados. Esta por sua vez, foi dividia em três bases com recortes temporais diferentes para a Validação Cruzada. A primeira base constitui o conjunto de treinamento, com 42 ações no período do primeiro trimestre de 2000 ao terceiro trimestre de 2006, totalizando 26 trimestres. Já a segunda base compôs o conjunto de validação com as mesmas 42 ações, mas no período do quarto trimestre de 2006 ao primeiro trimestre de 2009, totalizando 10 trimestres. Assim, no período do primeiro trimestre de 2000 até o primeiro de 2009, 70% foi composto pelo período de treinamento e 30%, pelo período de validação.

A terceira base de dados constituiu o conjunto de teste, formado por 25 ações, diferentes das 42 citadas acima, para o período do segundo trimestre de 2009 ao quarto trimestre de 2013, totalizando 19 trimestres. Em todas as três bases de dados, cada linha representa um trimestre de cada ação e cada coluna, uma variável, isto é, indicadores financeiros e retorno. O Quadro 8 apresenta as ações que compuseram cada conjunto.

Conjunto de Treinamento	Conjunto de Validação	Conjunto de Teste
ABEV3	ABEV3	AEDU3
ALLL3	ALLL3	BISA3
BBAS3	BBAS3	BRML3
BBDC3	BBDC3	BRPR3
BBDC4	BBDC4	BVMF3
BRAP4	BRAP4	CCRO3
BRFS3	BRFS3	CIEL3
BRKM5	BRKM5	CSAN3
CESP6	CESP6	CTIP3
CMIG4	CMIG4	DTEX3
CPFE3	CPFE3	ECOR3
CPLE6	CPLE6	ENBR3
CRUZ3	CRUZ3	ESTC3
CSNA3	CSNA3	EVEN3
CYRE3	CYRE3	GOLL4

ELET3	ELET3	HYPE3
ELET6	ELET6	JBSS3
ELPL4	ELPL4	KROT3
EMBR3	EMBR3	MMXM3
GFSA3	GFSA3	MRFG3
GGBR4	GGBR4	MRVE3
GOAU4	GOAU4	NATU3
HGTX3	HGTX3	PDGR3
ITSA4	ITSA4	RENT3
ITUB4	ITUB4	SANB11
LAME4	LAME4	
LIGT3	LIGT3	
LREN3	LREN3	
OIBR4	OIBR4	
PCAR4	PCAR4	
PETR3	PETR3	
PETR4	PETR4	
RSID3	RSID3	
SBSP3	SBSP3	
SUZB5	SUZB5	
TBLE3	TBLE3	
TIMP3	TIMP3	
UGPA3	UGPA3	
USIM5	USIM5	
VALE3	VALE3	
VALE5	VALE5	
VIVT4	VIVT4	

Quadro 8 - Ações do conjunto de treinamento e validação

Fonte: Elaborado pela autora.

Levando em consideração que o cálculo do retorno do ativo sempre recorre a um período anterior e que a base estava organizada de forma que os dados trimestrais estavam agrupados por ação, a base de treinamento teve que ser defasada em um período. Assim, se garantiu que os dados financeiros de uma ação não seriam utilizados para explicar o retorno de outra ação no período seguinte.

Após estes ajustes, os ativos foram classificados em duas classes. A Classe 1 ($y_i = +1$) foi composta pelos 25% das ações com maiores retornos e a Classe 2 ($y_i = -1$) foi composta pelas demais ações que representam 75%.

A próxima etapa foi a construção da Máquina de Suporte Vetorial por meio da biblioteca *kernlab*do *software* R. Para definição dos parâmetros C e σ, criou-se uma sequência para o parâmetro C variando de 1 a 100 e outra para o parâmetro σ, variando de 0,0001 a 2. Um *grid* foi construído com estas duas sequências para que a combinação ótima de parâmetros fosse encontrada, isto é, para que o par de parâmetros que gerasseo menor erro na etapa de validação fosse encontrado.Como medida de acurácia do desempenho do SVM, utilizou-se a porcentagem de previsões corretas de classificação da ação em relação ao total de classificações feitas.

A função *Kernel* utilizada neste trabalho foi o *Kernel*Gaussiano. Este é o mais popular nas aplicações do SVM em finanças, principalmente porque, como será demonstrado na fórmula (54), ele mapeia um ponto em um espaço de dimensões infinitas, permitindo uma busca mais generalizada e rápida da solução ótima. Por essa razão, foi utilizado nesta pesquisa, assim como nos estudos de Tay e Cao (2001), Huerta, Corbacho e Elkan (2013), Dinçer e Timor (2012) e Kim (2003). Sua formulação tradicional é dada por:

(8)

$$k(x,y) = \exp\left(\frac{\| x - y \|^2}{2\sigma^2}\right)$$

Entretanto, no *software* R, ele está implementado da seguinte forma:

(9)

$$k(x,y) = -\sigma \| x - y \|^2$$

Considerando que σrepresenta a medida de não linearidade do SVM.

Sabe-se que um problema comum em testes de modelos em finanças é o viés do *Data Snooping*. O efeito de *Data Snooping* ocorre quando uma base de dados é utilizada mais de uma vez , com o propósito de inferência ou seleção de um modelo, existindo assim a possibilidade de que algum resultado satisfatório ocorra por acaso , ao invés de ocorrer por mérito inerente ao procedimento . Então, para controlar os efeitos de *Data Snooping* e verificar a significância estatística destes

resultados foi aplicado o método *Bootstrap*com 10 000 amostragens. O *Bootstrap* caracteriza-se por amostragens sucessivas e com reposição a partir da amostra original, permitindo o estabelecimento de intervalos de confiança dos resultados da estratégia.

3.5.1 Metodologia Máquinas de Suporte Vetorial

O bom desempenho de um modelo classificador de padrões é alcançado quando a capacidade da função de classificação é compatível com o tamanho do conjunto de treinamento. Classificadores com um grande número de parâmetros ajustáveis geralmente apresentam grande capacidade de aprender os padrões do conjunto de treinamento sem erro, mas isso vem acompanhado por uma baixa capacidade de generalização. Sendo assim, deve existir um equilíbrio entre a capacidade de generalização do classificador e sua complexidade.

Em 1992,Boser, Guyon e Vapnik descreveram um algoritmo que automaticamente regula a capacidade de classificação da função por meio da maximização da margem entre os dados do conjunto de treinamento e o limite da classe.

As Máquinas de Suporte Vetorial são originárias deste estudo de Boser, Guyon e Vapnik (1992) e o seu maior diferencial é justamente a construção de um hiperplano que separa os dados em duas classes ou mais, para atingir a separação máxima entre elas. A Figura 1, mostra os diversos hiperplanos possíveis para a separação dos dados.

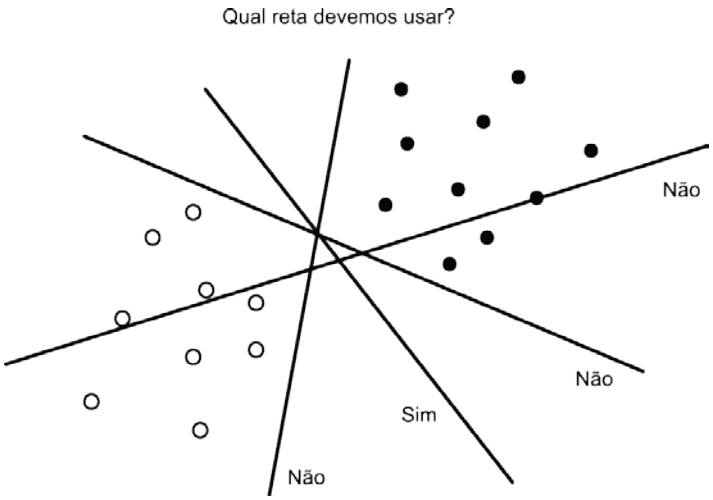


Figura 1 - Escolhendo o melhor hiperplano para classificação

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

De acordo Fan e Palaniswami (2001), a aplicação do Princípio de Minimização do Erro Empírico aplicado em métodos bastante difundidos como Redes Neurais, não garante um menor erro real. O SVM resolve essa questão com a implementação do Princípio da Minimização do Risco Estrutural, o qual procura minimizar o limite superior do erro de generalização, em vez de minimizar apenas o erro do processo de estimação. Isso significa que na classificação de novas observações de classes desconhecidas, a chance de haver um erro na predição, baseado na aprendizagem do classificador, será mínima.

Sendo assim, a aprendizagem do SVM pode ser entendida como a descoberta do hiperplano central que maximiza a margem de forma que as observações da Classe 1 ($y_i = +1$) fiquem o mais separadas possível das observações da Classe 2 ($y_i = -1$). A Figura 2 ilustra o conceito de máxima margem para um conjunto de dados que podem ser separados de maneira linear e direta.

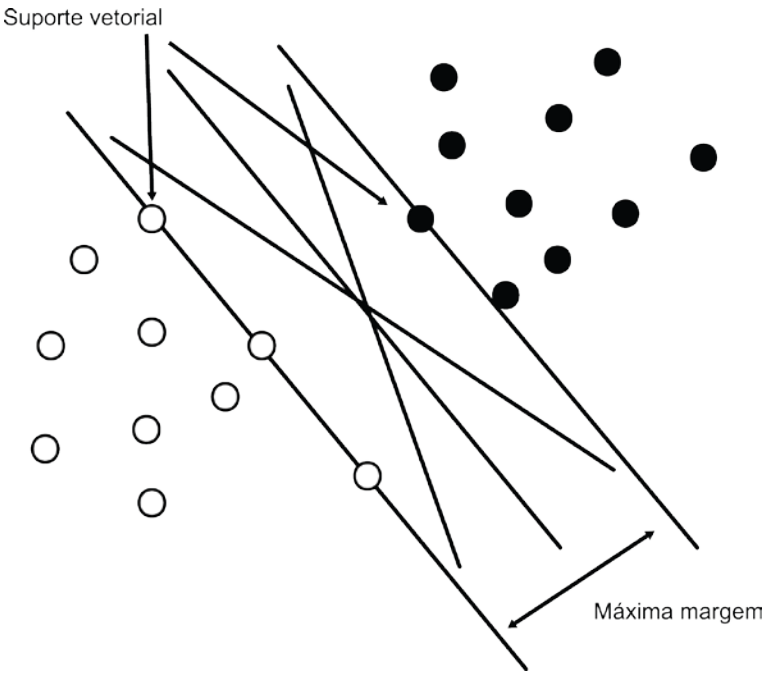


Figura 2 - Classificador de máxima margem

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

Os pontos localizados em cima das retas paralelas ao classificador são chamados de Suportes Vetoriais. Estes pontos possuem um papel crucial na teoria, justificando o nome do método, sendo que o termo “Máquinas” refere-se ao algoritmo.

3.5.2 Formulação do SVM Linear

O modelo clássico de Máquinas de Suporte Vetorial é o modelo de classificação linear dicotômica, que tem como objetivo encontrar uma função de decisão com a seguinte forma:

(10)

$$f(x) = sinal(w^T x - \gamma)$$

Onde *x* é um vetor de dimensão *p* x *1*representando o vetor de uma observação arbitrária com *p* variáveis, *w* é o vetor de parâmetros de dimensão *p* × 1 e γé um parâmetro escalar, denominado termo de viés.

A formulação do problema de separação linear tem como insumo para a estimação, uma matriz $\mathbf{A}$ de dimensão *n* x *p*, onde cada linha representa uma observação de uma população e cada coluna, uma característica, isto é, uma variável dessa população. Além disso, para fins de estimação deve-se considerar também um vetor y que representa o grupo no qual cada observação se encontra, sendo ele de dimensão *n* x 1 e contendo somente os valores $+1$ ou -1.

Considere o conjunto de dados abaixo:

(11)

$$\begin{bmatrix} i & x_1 & x_2 & y_i \\ 1 & 1 & 1 & -1 \\ 2 & 2 & 1 & -1 \\ 3 & 1 & 2 & -1 \\ 4 & 2 & 2 & -1 \\ 5 & 4 & 4 & +1 \\ 6 & 4 & 5 & +1 \\ 7 & 5 & 4 & +1 \\ 8 & 5 & 5 & +1 \end{bmatrix}$$

x_1 e x_2 representam duas variáveis do conjunto de dados, i representa as

observações e y_i a classe à qual cada observação pertence. A plotagem desses dados é dada pelaFigura 3.

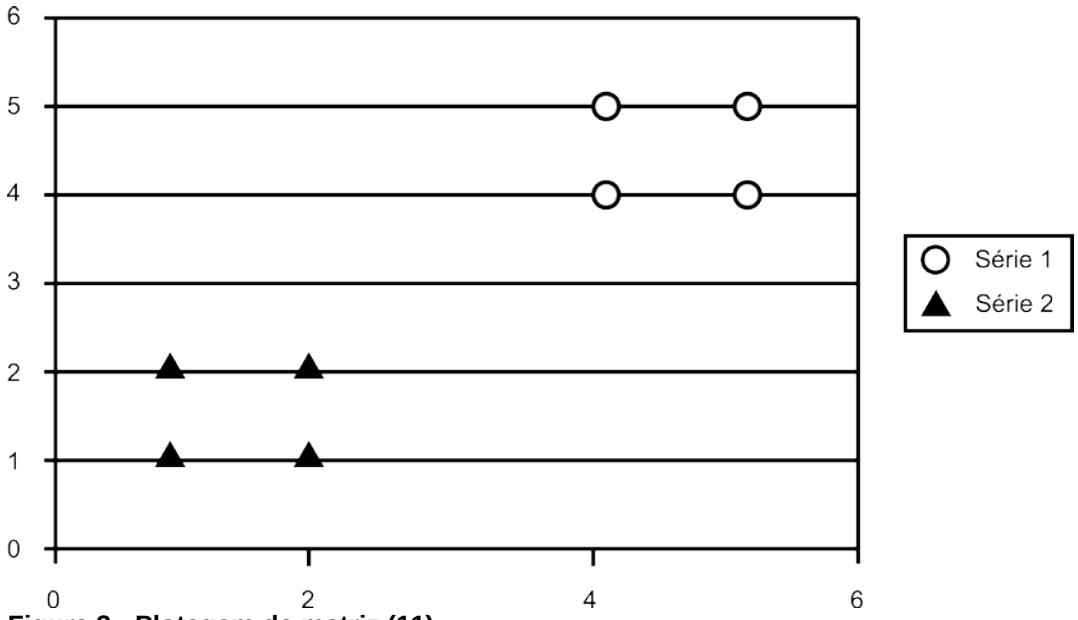


Figura 3 - Plotagem da matriz (11)

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

Aplicando o SVM a esse conjunto de dados, o objetivo é achar o hiperplano de máxima margem, da forma $w_1x_1 + w_2x_2 - \gamma = 0$, e dois outros planos que limitarão cada uma das classes, assumindo a forma $w_1x_1 + w_2x_2 - \gamma \geq +1$ e $w_1x_1 + w_2x_2 - \gamma \leq -1$. Em outras palavras, o SVM visa encontrar dois planos tal que os pontos com $d = -1$ satisfaçam a restrição $w_1x_1 + w_2x_2 - \gamma \leq -1$ e os pontos com $d = +1$ satisfaçam $w_1x_1 + w_2x_2 - \gamma \geq +1$.

As restrições podem ser escritas da seguinte forma:

(12)

$$1w_1 + 1w_2 - \gamma \leq -1$$
$$2w_1 + 1w_2 - \gamma \leq -1$$
$$1w_1 + 2w_2 - \gamma \leq -1$$
$$2w_1 + 2w_2 - \gamma \leq -1$$
$$4w_1 + 4w_2 - \gamma \geq +1$$
$$4w_1 + 5w_2 - \gamma \geq +1$$
$$5w_1 + 4w_2 - \gamma \geq +1$$
$$5w_1 + 5w_2 - \gamma \geq +1$$

Isso equivale à:

(13)

$$(-1)\ (1w_1 + 1w_2 - \gamma) \geq +1$$
$$(-1)\ (2w_1 + 1w_2 - \gamma) \geq +1$$
$$(-1)\ (1w_1 + 2w_2 - \gamma) \geq +1$$
$$(-1)\ (2w_1 + 2w_2 - \gamma) \geq +1$$
$$(+1)\ (4w_1 + 4w_2 - \gamma) \geq +1$$
$$(+1)\ (4w_1 + 5w_2 - \gamma) \geq +1$$
$$(+1)\ (5w_1 + 4w_2 - \gamma) \geq +1$$
$$(+1)\ (5w_1 + 5w_2 - \gamma) \geq +1$$

A forma matricial resultante é:

(14)

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 2 & 2 \\ 4 & 4 \\ 4 & 5 \\ 5 & 4 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \gamma \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Ou:

(15)

$$\mathbf{D}(\mathbf{A}w - \gamma 1) \geq 1$$

Onde,

(16)

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 2 & 2 \\ 4 & 4 \\ 4 & 5 \\ 5 & 4 \\ 5 & 5 \end{bmatrix} = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ x_4^T \\ x_5^T \\ x_6^T \\ x_7^T \\ x_8^T \end{bmatrix};\ \mathbf{D} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Sabe-se que a distância do hiperplano $w_1 x_1 + w_2 x_2 - \gamma = +1$ até a origem é

$\frac{|-\gamma-1|}{\sqrt{w_1^2+w_2^2}}$ e que a distância até o hiperplano $w_1x_1 + w_2x_2 - \gamma = -1$ é $\frac{|-\gamma+1|}{\sqrt{w_1^2+w_2^2}}$.

Portanto, a distância entre esses dois planos é $\frac{2}{\sqrt{w_1^2+w_2^2}}$. O objetivo então é achar o w e o γ que maximizem a distância e ao mesmo tempo satisfaçam a matriz de restrições.

Maximizar a margem $\frac{2}{\sqrt{w_1^2+w_2^2}}$ equivale a minimizar o seu recíproco$\frac{w_1^2+w_2^2}{2} = \frac{1}{2}w^Tw$, então o problema de programação quadrática pode ser escrito de duas formas:

(17)

Maximize: $\varsigma = \frac{2}{||w||}$

Sujeito à

$\mathbf{D}(\mathbf{A}w - \gamma 1) \geq 1$

Para $w \in \mathbb{R}^p$, $\gamma \in \mathbb{R}$.

Ou,

(18)

Minimize : $\varsigma^* = \frac{1}{2}\, w^Tw$

Sujeito à

$\mathbf{D}(\mathbf{A}w - \gamma 1) \geq 1$

Para$w \in \mathbb{R}^p$, $\gamma \in \mathbb{R}$.

Sintetizando, o treinamento do SVM consiste em achar $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ e γ dada uma matriz de dados $\mathbf{A}$ e o vetor de classes y. Como apresentado anteriormente, a função de decisão do SVM linear é dada por:

(19)

$f(x) = sinal(w^Tx - \gamma)$

Isso significa que para uma nova observação, o sinal de $w^Tx - \gamma$ a

classificará na Classe 1 ou na Classe 2.

O problema primal do SVM também pode ser escrito na sua forma dual, dada pelo Dual de Wolfe (1961):

(20)

$$Max_{\lambda \geq 0}\left[Min_{w,\gamma} \mathrm{L}(\mathrm{w},\gamma,\lambda)\right]$$

Para determinado λ devemos encontrar we γ que minimizem $\mathrm{L}(\mathrm{w},\gamma,\lambda)$ e depois substituir o resultado na função Lagrangeana para maximizá-la em função de λ. Note que $\mathrm{L}(\mathrm{w},\gamma,\lambda)$ é um escalar e a função Lagrangeana é dada por:

(21)

$$\mathrm{L}(\mathrm{w},\gamma,\lambda) = \frac{1}{2}w^T w - \lambda^T\{\mathbf{D}[(\mathbf{A}w - \gamma 1) - 1]\}$$

Resolvendo as condições de primeira ordem, se tem:

(22)

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathrm{L}(\mathrm{w},\gamma,\lambda) = 0 \rightarrow \frac{\mathrm{d}}{\mathrm{d}\gamma}\left(\frac{1}{2}w^T w - \lambda^T\mathbf{D}\mathbf{A}w + \lambda^T\mathbf{D}\gamma 1 - \lambda^T\mathbf{D}1\right) = 0 \rightarrow\ \lambda^T\mathbf{D}1 = 0$$

(23)

$$\frac{\mathrm{d}}{\mathrm{d}w}\mathrm{L}(\mathrm{w},\gamma,\lambda) = 0 \rightarrow \frac{\mathrm{d}}{\mathrm{d}w}\left(\frac{1}{2}w^T w - \lambda^T\mathbf{D}\mathbf{A}w + \lambda^T\mathbf{D}\gamma 1 - \lambda^T\mathbf{D}1\right) = 0 \rightarrow$$

$$\rightarrow \frac{1}{2}2w - \lambda^T\mathbf{D}\mathbf{A}w = 0 \rightarrow w^T - \lambda^T\mathbf{D}\mathbf{A} = 0$$

Isso significa que $w^T = \lambda^T\mathbf{D}\mathbf{A}$ **e** $w = \mathbf{A^T}\mathbf{D}\lambda$**.** Substituindo os resultados na função Lagrangeana, o problema Dual é:

(24)

$$\mathrm{L}(\lambda) = -\frac{1}{2}\lambda^T\mathbf{D}\mathbf{A}\mathbf{A^T}\mathbf{D}\lambda + \lambda^T 1$$

Sujeito à

$1^T\mathbf{D}\lambda = 0$

$\lambda \geq 0$

Para definição algébrica do problema, considere:

(25)

$$\lambda^T \mathbf{DAA^T D}\lambda = (\lambda_1 \dots \lambda_n)\begin{pmatrix} y_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & y_n \end{pmatrix}\begin{pmatrix} x_1^T x_1 & \cdots & x_1^T x_n \\ \vdots & \ddots & \vdots \\ x_n^T x_1 & \cdots & x_n^T x_n \end{pmatrix}\begin{pmatrix} y_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & y_n \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

$$= (\lambda_1 y_1 \dots \lambda_n y_n)\begin{pmatrix} x_1^T x_1 & \cdots & x_1^T x_n \\ \vdots & \ddots & \vdots \\ x_n^T x_1 & \cdots & x_n^T x_n \end{pmatrix}\begin{pmatrix} \lambda_1 y_1 \\ \vdots \\ \lambda_n y \end{pmatrix}$$

$$= (\lambda_1 y_1 x_1^T x_1 + \lambda_2 y x_2^T x_1 + \lambda_3 y_3 x_3^T x_1 + \cdots + \lambda_n y_n x_n^T x_1 + \cdots + \lambda_1 y_1 x_1^T x_n + \cdots + \lambda_n y_n x_n^T x_n)\begin{pmatrix} \lambda_1 y_1 \\ \vdots \\ \lambda_n y_n \end{pmatrix}$$

$$= (\sum_{i=1}^{n} \lambda_i y_i x_i^T x_1, \dots, \sum_{i=1}^{n} \lambda_i y_i x_i^T x_n)\begin{pmatrix} \lambda_1 y_1 \\ \vdots \\ \lambda_n y_n \end{pmatrix}$$

$$= (\sum_{i=1}^{n} \lambda_i y_i x_i^T x_1)\lambda_1 y_1 + \cdots + (\sum_{i=1}^{n} \lambda_i y_i x_i^T x_n)\lambda_n y_n$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_i y_i x_i^T x_j \lambda_j y_j$$

Então,

(26)

$$\mathrm{L}(\mathrm{w}, \gamma, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \lambda_i [y_i (w^T x_i - \gamma) - 1]$$

$$\mathrm{L}(\mathrm{w}, \gamma, \lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j \left(x_i^T x_j\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y\left(x_i^T x_j^T\right) + \sum_{i=j}^{n} \lambda_i$$

$$\mathrm{L}(\mathrm{w}, \gamma, \lambda) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j \left(x_i^T x_j^T\right) + \sum_{i=j}^{n} \lambda_i$$

Assim, o Lagrangeano em função somente de λ é:

(27)

$$\mathrm{Max}\ \mathrm{L}(\lambda) = \sum_{i=j}^{n} \lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j \left(x_i^T x_j^T\right)$$

Sujeito à

$$\sum_{i=j}^{n} \lambda_i \, y_i = 0$$

$$\lambda \geq 0, i = 1,2, \dots, n.$$

3.5.3 L1-SVM com Margem Suave: *Kernel* Linear

Agora, considere um conjunto de dados que não são totalmente separáveis por um hiperplano. A solução neste caso é achar o hiperplano que “melhor” separa esses dados, ou seja, aquele de máxima margem e que permite que apenas um pequeno número de pontos caia na classe errada. O desvio desses pontos em relação à sua classe original é denominado “erro”. Desejamos, portanto, encontrar o hiperplano com o mínimo de pontos que contribuem para o erro. A condição de máxima margem e a de minimização do número de pontos que contribuem para o erro são contraditórias, pois uma margem maior gerará mais pontos com erros. Por isso, o parâmetro C é introduzido e ele representa o custo do erro, isto é, o peso de se fazer uma classificação errada.

A Figura 4 ilustra o seguinte conjunto de dados:

(28)

$$\begin{bmatrix} y & x_1 & x_2 \\ +1 & 1.0 & 0.8 \\ +1 & 3.0 & 2.5 \\ +1 & 2.5 & 1.0 \\ -1 & 1.0 & 1.8 \\ -1 & 3.0 & 4.5 \\ -1 & 2.5 & 2.8 \end{bmatrix}$$

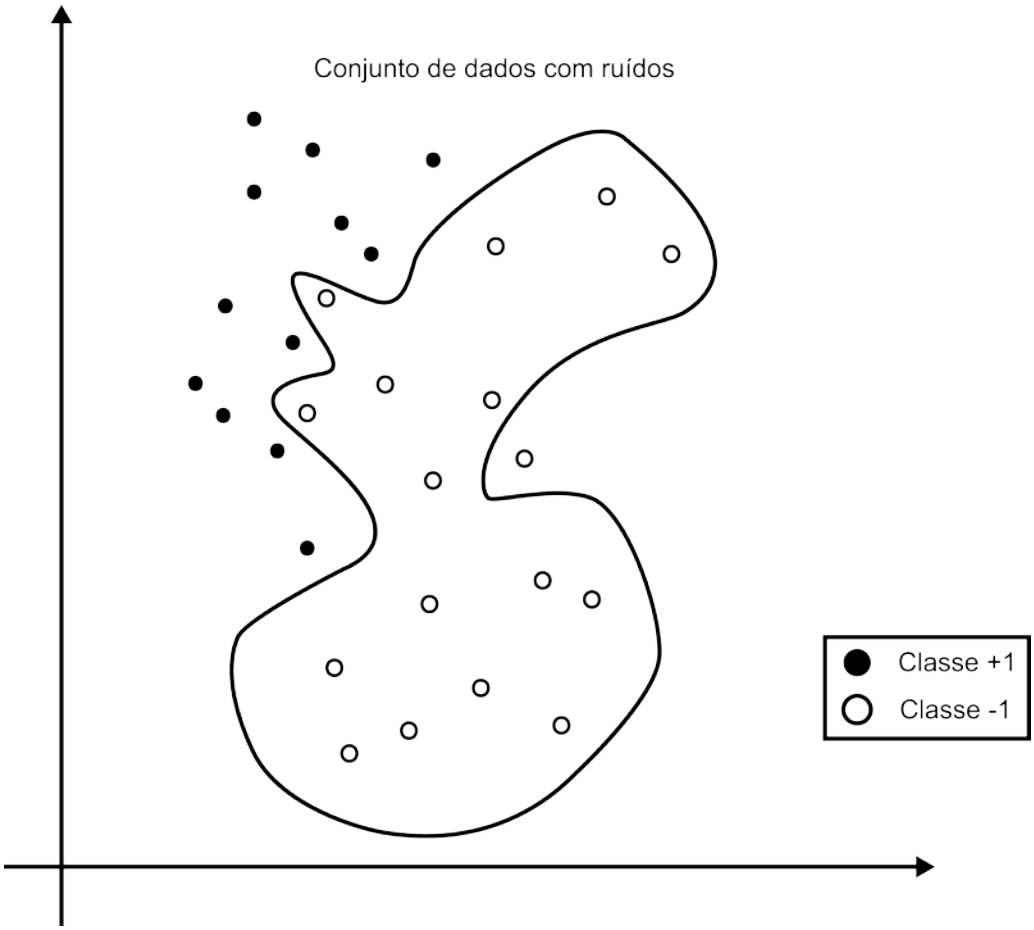


Figura 4 - Conjunto de dados que requer o SVM linear com margem suave

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

Neste caso, segundo Soman, Loganathan e Ajay (2011), não recomenda-se o uso de um SVM não linear porque o risco de *overfitting*com um modelo mais robusto é grande. Considere também as seguintes restrições:

(29)

$$1w_1 + 0.8w_2 - \gamma \geq +1$$
$$3w_1 + 2.5w_2 - \gamma \geq +1$$
$$2.5w_1 + 1.0w_2 - \gamma \geq +1$$
$$1w_1 + 1.8w_2 - \gamma \leq -1$$
$$3w_1 + 4.5w_2 - \gamma \leq -1$$
$$2.5w_1 + 2.8w_2 - \gamma \leq -1$$

Como permite-se o erro de treinamento ξ_ie os pontos não são linearmente separáveis, as restrições passam a ser:

(30)

$$1w_1 + 0.8w_2 - \gamma + \xi_1 \geq +1$$
$$3w_1 + 2.5w_2 - \gamma + \xi_2 \geq +1$$
$$2.5w_1 + 1w_2 - \gamma + \xi_3 \geq +1$$
$$1w_1 + 1.8w_2 - \gamma + \xi_4 \leq -1$$
$$3w_1 + 4.5w_2 - \gamma + \xi_5 \leq -1$$
$$2.5w_1 + 2.8w_2 - \gamma + \xi_6 \leq -1$$

Isso equivale à:

(31)

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \left\{ \left(\begin{bmatrix} 1.0 & 0.8 \\ 3.0 & 2.5 \\ 2.5 & 1.0 \\ 1.0 & 1.0 \\ 3.0 & 4.5 \\ 2.5 & 2.8 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} - \gamma \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right) \right\} + \geq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Logo, a forma matricial primal do L1-SVM é:

(32)

Minimize : $\varsigma^* = \frac{1}{2}\, w^T w + \mathrm{C}1^T \xi$

Sujeito à

$\mathbf{D}(\mathbf{A}w - \gamma 1) + \xi \geq 1$

$\xi \geq 0$

Para achar a forma primal, basta seguir os mesmos passos descritos para o SVM Linear, atentando para o fato de que agora se terá mais um multiplicador de Lagrange.

(28)

$$\mathrm{Max}_{\substack{\lambda \geq 0 \\ \mu \geq 0}} \left[\mathrm{Min}_{\mathrm{w},\gamma,\xi} \mathrm{L}(\mathrm{w}, \gamma, \lambda, \mu, \xi) \right]$$

Seguindo a mesma lógica, para determinado λ e μ devemos encontrar w, γ e ξ que minimizem $\mathrm{L}(\mathrm{w}, \gamma, \lambda, \mu, \xi)$ e depois substituir o resultado na função Lagrangeanapara maximizá-la em função de λ e μ. Lembrando que $\mathrm{L}(\mathrm{w}, \gamma, \lambda, \mu, \xi)$ é um escalar e a função Lagrangeana é dada por:

(34)

$$\mathrm{L}(\mathrm{w},\gamma,\lambda,\mu,\xi) = \frac{1}{2}w^T w + \mathrm{C}1^T\xi - \lambda^T[\mathbf{D}(\mathbf{A}w - \gamma 1) + \xi - 1] - \mu^T\xi$$

Resolvendo as condições de primeira ordem, se tem:

(35)

$$\frac{\mathrm{d}}{\mathrm{d}w}\mathrm{L}(\mathrm{w},\gamma,\lambda,\mu) = 0 \rightarrow \frac{\mathrm{d}}{\mathrm{d}w}\left(\frac{1}{2}w^T w + \mathrm{C}1^T\xi - \lambda^T\boldsymbol{DA}w + \lambda^T\mathbf{D}\gamma 1 - \lambda^T\xi + \lambda^T - \mu^T\xi\right)$$
$$= 0 \rightarrow w^T - \lambda^T\mathbf{DA} = 0 \rightarrow\ w^T = \lambda^T\mathbf{DA}$$

(36)

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathrm{L}(\mathrm{w},\gamma,\lambda,\mu) = 0 \quad \rightarrow \frac{\mathrm{d}}{\mathrm{d}\gamma}\left(\frac{1}{2}w^T w + \mathrm{C}1^T\xi - \lambda^T\mathbf{DA}w + \lambda^T\mathbf{D}\gamma 1 - \lambda^T\xi + \lambda^T - \mu^T\xi\right)$$
$$= 0 \rightarrow\ \lambda^T\mathbf{D}1 = 0$$

(37)

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\mathrm{L}(\mathrm{w},\gamma,\lambda,\mu) = 0 \rightarrow \frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{1}{2}w^T w + \mathrm{C}1^T\xi - \lambda^T\mathbf{DA}w + \lambda^T\mathbf{D}\gamma 1 - \lambda^T\xi + \lambda^T - \mu^T\xi\right)$$
$$= 0 \rightarrow\ \mathrm{C}1^T - \lambda^T - \mu^T = 0$$

Vale ressaltar que como $\lambda \geq 0$, $\mu \geq 0$ e $1^T - \lambda^T - \mu^T$, o maior valor de μ é C e λ_i é máximo quando $\mu_i = 0$. Portanto, os multiplicadores de Langrange possuem valores truncados. Essa é maior diferença entre o SVM de Margem Suave e o SVM Linear Clássico.

Substituindo os resultados na função Lagrangeana, o problema Dual é:

(38)

$$L(\lambda) = -\frac{1}{2}\lambda^T\mathbf{DAA^T D}\lambda + \lambda^T 1$$

Sujeito a

$1^T\mathbf{D}1 = 0$

$0 \leq \lambda \leq \mathrm{C}1$

Existem outras variações do SVM Linear, como o SVM L2-Norm SVM , onde ao invés de se minimizar apenas o erro, opta-se por minimizar a soma dos quadrados dos erros. A formulação do L2-Norm SVM é:

(39)

$$\text{Minimize} : \varsigma^* = \frac{1}{2}\, w^T w + \frac{\mathrm{C}}{2}\sum_{i=1}^{m} \xi_i^2$$

Sujeito a

$$\mathbf{D}(w^T x - \gamma) + \xi_i - 1 \geq 0$$

$$\xi_i \geq 0$$

3.5.4 SVM Não Linear

Um problema de separação linear é apenas um caso particular e por isso torna-se necessária uma formulação capaz de lidar com problemas mais complexos. De acordo com Soman, Loganathan e Ajay (2011), o problema de aprendizagem para SVMs é definido como uma relação de dependência desconhecida e não linear entre dados de alta dimensão, representados por um vetor ou matriz **A,** e uma variável alvo ou variável *target*, representada por um escalar $y$ou vetor $\mathcal{y}$**,**no caso de SVMs de múltiplas classes. Essa relação é descrita por um mapeamento ou função $y = f(\mathrm{x})$, onde a matriz ou vetor x são os *inputs* e a variável y, é o *output.* A única informação disponível é o conjunto de treinamento $T_D = \{(x_i, y_i)\},\ i = 1,2,\dots,m$, onde m representa o número de pares do conjunto de treinamento e é, portanto, igual à dimensão do conjunto de treinamento T_D.

Este problema é similar à Inferência Estatística Clássica que segundo Soman, Loganathan e Ajay (2011), baseia-se nos seguintes pressupostos:

i. Os dados podem ser modelados por um conjunto de parâmetros de funções lineares, sendo este um fundamento de um paradigma paramétrico do aprendizado de dados experimentais.

ii. Na maioria dos problemas, o componente estocástico é assumido para seguir uma Distribuição Normal de probabilidade.

iii. Como consequência do segundo pressuposto, para a estimação de parâmetros, o paradigma de indução estatística é o método com maior verossimilhança, e este é reduzido a minimização da função de custo da soma dos erros ao quadrado na maioria das aplicações.

Os três pressupostos nos quais o paradigma estatístico clássico se baseia tornaram-se inadequados para vários problemas da atualidade devido às seguintes razões:

i. Problemas da atualidadepossuem alta dimensionalidade e se o mapeamento básico não for bem suavizado, a “Maldição da Dimensionalidade” é desencadeada. Em outras palavras, para que o modelo tenha um bom desempenho, o número de elementos de treinamento requeridos será uma função exponencial da dimensão do espaço de característica.

ii. A distribuição dos dados reais pode ser bastante diferente de uma Distribuição Normal e o modelo deve considerar essa diferença a fim de construir um algoritmo de aprendizagem eficaz.

iii. A partir dos dois primeiros pontos, conclui-se que função de custo da soma de erros ao quadrado, deve ser substituída por um novo paradigma de indução que é uniformemente melhor, a fim de modelar distribuições não-Gaussianas.

Esse contexto levou Vapnik a formular métodos baseados em funções *Kernel* para aprendizado em dados de alta dimensão.

Em situações como a da Figura 5, se há uma tentativa de separação linear dos dados, o grau de tolerância para erros de classificação deve ser bem alto. Neste caso, é preferível construir um mapeamento dos dados em algum espaço de dimensão maior, em um “espaço de característica” no qual eles serão linearmente separáveis.

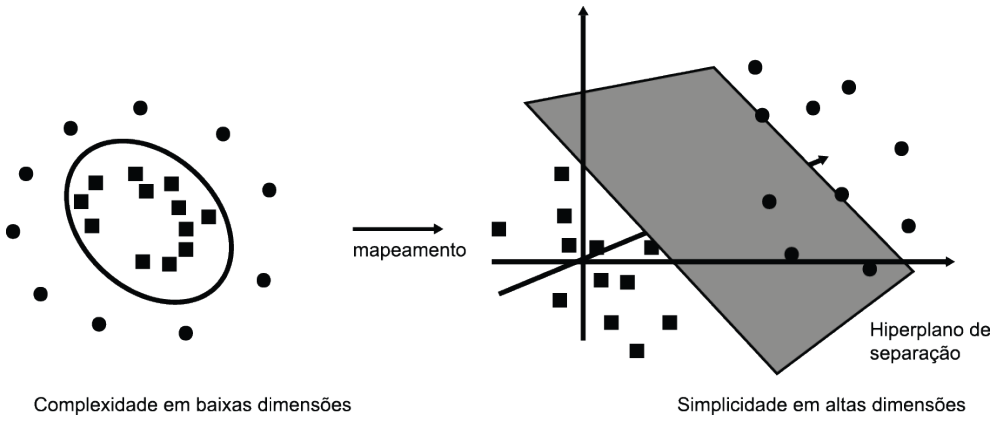


Figura 5 - Classificador não linear
Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

Para distinguir a dimensão original dos dados e o espaço de característica, Soman, Loganathan e Ajay (2011) denominam o primeiro de “espaço de *input*”. A Figura 6ilustra o processo de mapeamento não linear no espaço de característica.

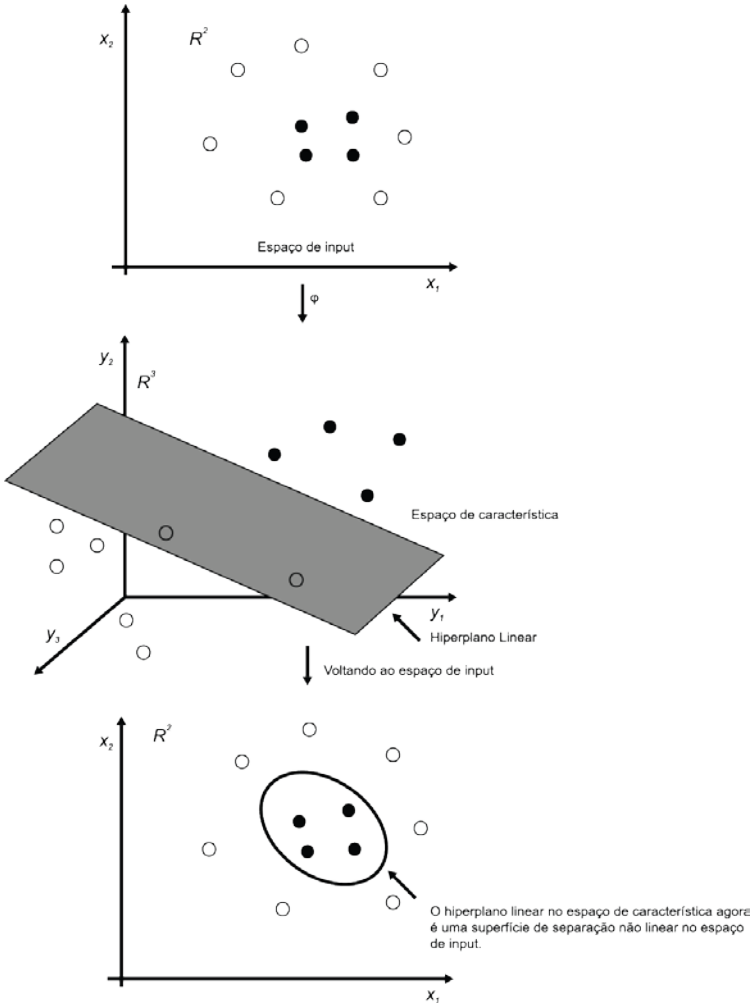


Figura 6 - Processo de mapeamento

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

Fica evidente então que a estratégia para se trabalhar com a não linearidade de dados é criar novas dimensões por meio do processo de mapeamento e este é descrito da seguinte forma:

(40)

$$x \to \phi(x)$$
$$\mathbb{R}^p \to \mathbb{R}^q \quad \text{tal que } q \gg p.$$

Na situação ilustrada na Figura 6, por exemplo,partiu-se de um espaço bidimensional (x_1, x_2) para um espaço tridimensional $(t_1, t_2, t_3) =$

$(x_1^2, x_2^2, \sqrt{2}x_1x_2)$. Porém, existem diversos mapeamentos que podem levar a diferentes espaços de características e o desafio que surge é justamente identificar qual o melhor para determinado problema de classificação de forma que este minimize o erro de generalização.

3.5.5 Métodos *Kernel*

O mapeamento em questão é definido pela função *Kernel* $\mathrm{K}(x_i, x_j)$ que representa uma medida similaridade ou proximidade entre os pontos, sendo descrita sinteticamente como:

(41)

$$\mathrm{K}(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

ou

(42)

$$\mathrm{K}(x_i, x_j) = g\left(-\sigma * dist(x_i, x_j)\right)$$

Onde $dist(x_i, x_j)$ é a distancia entre x_i e x_j, σ é um parâmetro escalar de "suavização" da função e $g(z)$ é uma função que decresce a medida que *z* aumenta. O valor de σ define a intensidade da inclinação da função *Kernel* de acordo com a distância entre os pontos. A Gráfico 2 ilustra essa relação entre a inclinação da função e a variação de σ para uma curva Gaussiana.

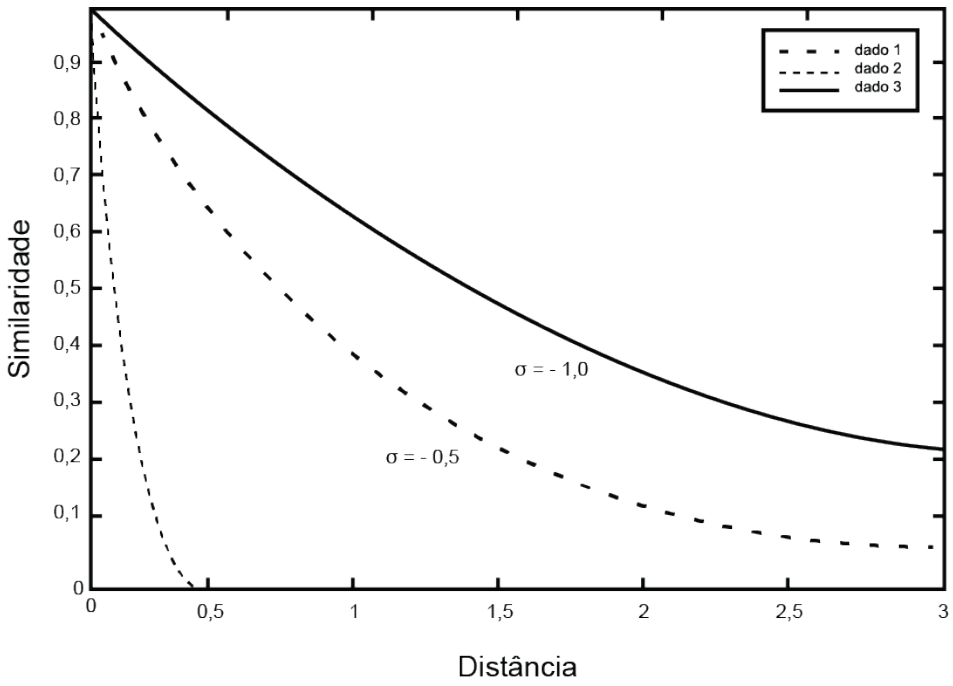


Gráfico 2 - Valor *Kernel* como uma função de distância

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

O *Kernel* leva cada ponto x em um mapa $\phi(x)$. Dessa forma, ao invés de se trabalhar com a matriz $\mathbf{A}$, trabalha-se com $\mathbf{F}$, construída a partir de $\phi(x)$. Como $\mathbf{F}$é de dimensão muito grande,o cálculo da matriz $\mathbf{FF^T}$exige uma quantidade enorme de operações, o que torna o processo bastante oneroso. Essa explosão de dimensionalidade pode ser evitada com o uso do *KernelTrick*. A função $\mathrm{K}(x_i, x_j)$, por ser uma função do espaço de *input,* não requer o mapeamento, mas sim o resultado escalar do produto $\phi(x_i)^T\phi(x_j)$no espaço de característica. Os produtos escalares são encontrados diretamente com o cálculo dos *Kernels*$\mathrm{K}(x_i, x_j)$ para determinado conjunto de vetores de dados de treinamento em um espaço de *input*. Então, basta que a matriz $\mathbf{FF^T}$ seja conhecida para construir um SVM que opere em um espaço de dimensão extremamente alta, até mesmo infinita. Por meio dessa artimanha o algoritmo do treinamento do SVM não linear se torna o mesmo que o do SVM linear, tornando as operações e a abordagem do problema da Maldição da Dimensionalidade bem mais simples.

Considere a seguinte matriz de dados e que em sua dimensão original eles não são linearmente separáveis:

(43)

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Sabe-se que:

(44)

$$\mathbf{F} = \begin{pmatrix} \phi^T(x_i) \\ \vdots \\ \phi^T(x_n) \end{pmatrix}$$

(45)

$$\mathbf{F^T} = \left(\phi(x_i), \dots, \phi(x_n)\right)$$

Logo,

(46)

$$\mathbf{FF^T} = \begin{pmatrix} \phi^T(x_1)\phi(x_1) & \cdots & \phi^T(x_1)\Phi(x_n) \\ \vdots & \ddots & \vdots \\ \phi^T(x_n)\phi(x_1) & \cdots & \phi^T(x_n)\Phi(x_n) \end{pmatrix}$$

A partir do *Kernel* Quadrático, dado por $\phi(x) = \begin{pmatrix} x^2 \\ \sqrt{2}x \\ 1 \end{pmatrix}$, é possível levar cada ponto em $\mathbb{R}^2$ para $\mathbb{R}^3$ e construir a matriz $\mathbf{F}$.Como $\mathbf{K} = \mathbf{FF^T} = \phi^T(x_i)\phi\left(x_j\right)$, a matriz *Kernel* pode ser escrita como:

(47)

$$\mathbf{K} = \begin{bmatrix} \phi(x_1)^T\phi(x_1) & \phi(x_1)^T\phi(x_2) & \phi(x_1)^T\phi(x_3) \\ \phi(x_2)^T\phi(x_1) & \phi(x_2)^T\phi(x_2) & \phi(x_2)^T\phi(x_3) \\ \phi(x_3)^T\phi(x_1) & \phi(x_3)^T\phi(x_2) & \phi(x_3)^T\phi(x_3) \end{bmatrix}$$

(48)

$$\mathbf{K} = \begin{bmatrix} ({x_1}^T x_1 + 1)^2 & ({x_1}^T x_2 + 1)^2 & ({x_1}^T x_3 + 1)^2 \\ ({x_2}^T x_1 + 1)^2 & ({x_2}^T x_2 + 1)^2 & ({x_2}^T x_3 + 1)^2 \\ ({x_3}^T x_1 + 1)^2 & ({x_3}^T x_2 + 1)^2 & ({x_3}^T x_3 + 1)^2 \end{bmatrix}$$

Dado que o *Kernel* em questão é $K\left(x_i, x_j\right) = (x_i^T x_j + 1)^2$ e $x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $x_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, $x_3 = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$, tem-se que:

(49)

$$\mathbf{K} = \begin{pmatrix} 36 & 144 & 324 \\ 144 & 676 & 1600 \\ 324 & 1600 & 3844 \end{pmatrix}$$

Outro exemplo simples de uma situação na qual há a necessidade da construção de modelos de classificação não lineares é dado pela Figura 7, onde o limite de separação é quadrático. É evidente que nenhum hiperplano linear de separação com erro zero pode ser encontrado nessa situação. A melhor função linear de separação geraria cinco classificações erradas, sendo três na classe negativa e duas na classe positiva. Entretanto, se a função não linear de separação for utilizada, as classes podem ser separadas sem nenhum erro.

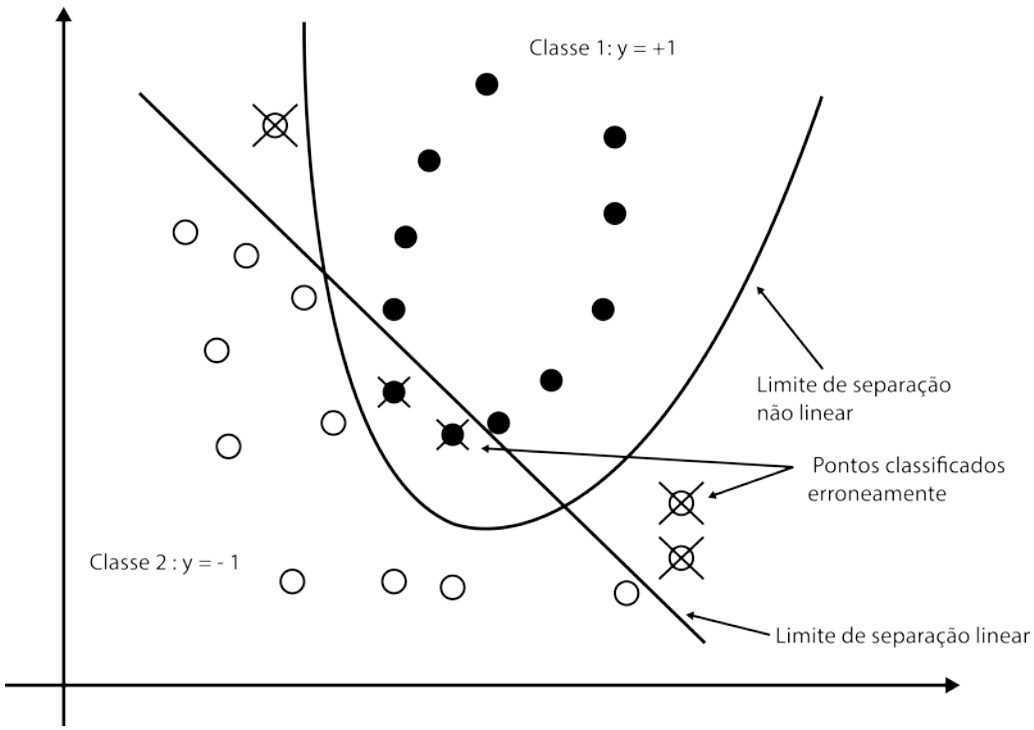


Figura 7- Separador linear *versus* separador não linear

Fonte: Traduzido de SOMAN, LOGANATHAN; AJAY, 2011.

Agora suponha $x \in \Re^2$, isto é, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Se o mapeamento escolhido for $\phi = \begin{bmatrix} x_1^2 \\ \sqrt{x_1 x_2} \\ x_2^2 \end{bmatrix}$, ou seja, $\Re^2 \longrightarrow \Re^3$, então o produto interno será:

(50)

$$\phi(x_i)^T\phi(x_j) = \begin{bmatrix} x_{i1}^2 & \sqrt{2}x_{i1}^2x_{i2}^2 & x_{i2}^2 \end{bmatrix}\begin{bmatrix} x_{j1}^2 \\ \sqrt{2}x_{j1}^2x_{j2}^2 \\ x_{j2}^2 \end{bmatrix}$$

$$= \begin{bmatrix} x_{i1}^2x_{j1}^2 & 2x_{i1}^2x_{i2}^2x_{j1}^2x_{j2}^2 & x_{i2}^2x_{j2}^2 \end{bmatrix} = ({x_i}^T x_j)^2 = K(x_i, x_j)$$

É possível notar novamente que para calcular o produto escalar no espaço de característica $\phi(x_i)^T\phi(x_j)$ não há necessidade de se aplicar de fato o mapeamento $\phi = \begin{bmatrix} x_1^2 \\ \sqrt{x_1x_2} \\ x_2^2 \end{bmatrix}$, já que o produto pode ser encontrado diretamente calculando $({x_i}^T x_j)^2$.

Para exemplificar ainda mais, assuma agora o seguinte mapeamento:

(51)

$$\phi = \begin{bmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$

Ou seja, nesse caso há um mapeamento $\Re^2 \longrightarrow \Re^5$ e um termo de viés como o valor da sexta dimensão, sendo este uma constante. Então, o produto interno no espaço de característica $\mathbf{F}$ é dado como:

(52)

$$\phi(x_i)^T\phi(x_j) = \left[1 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 2x_{i1}x_{i2}x_{j1}x_{j2} + x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2\right]$$

$$= 1 + 2{x_i}^T x_j + ({x_i}^T x_j)^2 = ({x_i}^T x_j + 1)^2 = K(x_i, x_j)$$

Ou

(53)

$$K(x_i, x_j) = ({x_i}^T x_j + 1)^2 = \phi(x_i)^T\phi(x_j)$$

Sintetizando, um *Kernel* polinomial de grau p é calculado da seguinte forma:

(54)

$$\mathbf{K} = \begin{bmatrix} (x_1{}^T x_1 + 1)^p & (x_1{}^T x_2 + 1)^p & . & (x_1{}^T x_m + 1)^p \\ . & . & (x_i{}^T x_j + 1)^p & . \\ . & . & . & . \\ (x_m{}^T x_1 + 1)^p & (x_m{}^T x_2 + 1)^p & . & (x_m{}^T x_m + 1)^p \end{bmatrix}$$

Nota-se que o produto interno é calculado na sua dimensão original.

Soman, Loganathan e Ajay (2011) apresentam o seguinte exemplo com o *Kernel*Gaussiano para provar a operação do SVM em dimensões infinitas. Considerando sua função que é dada por:

(55)

$$K(x, y) = \exp(-\sigma \parallel x - y \parallel^2)$$

Onde γ é um parâmetro positivo para controlar o raio, pode-se expandi-la, o que resulta em:

(56)

$$\exp(-\sigma \parallel x - y \parallel^2) = \exp(\parallel x \parallel^2) \exp(-\sigma \parallel y \parallel^2) \exp(2\sigma x^T y)$$

Como,

(57)

$$\exp(2\sigma x^T y) = 1 + 2\sigma x^T y + \frac{(2\sigma)^2}{2!}(x^T y)^2 + \frac{(2\sigma)^3}{3!} + (x^T y)^3 + \cdots$$

infere-se que $\exp(2\sigma x^T y)$ é um somatório infinito de polinômios, ou seja, o *Kernel*Gaussiano é um *Kernel* que mapeia um ponto em um espaço de dimensões infinitas. Vale ressaltar também que $\exp(-\sigma \parallel x \parallel^2)$ e $\exp(-\sigma \parallel y \parallel^2)$ também são *Kernels* assim como o produto deles, $K(x, y) = \exp(-\sigma \parallel x - y \parallel^2)$.

Os métodos *Kernel* possuem várias vantagens que os tornam atrativos. Eles são considerados universalmente aplicáveis, pois à medida que a amostra de treinamento m se torna arbitrariamente grande, $m \to \infty$, a estimativa do *Kernel* se aproxima da função alvo ideal, com apenas leves restrições. Outra vantagem dos métodos *Kernel* é que não se faz necessário nenhum treinamento para construir o modelo, o próprio conjunto de treinamento é o modelo. Os procedimentos são também conceitualmente simples e fáceis de serem explicados.

Entretanto, os métodos *Kernel* também possuem algumas desvantagens que têm impedido várias aplicações, principalmente no campo da mineração de dados. Como não há um modelo concreto para construção dos *Kernels*, não é possível discernir como a função depende das respectivas variáveis preditoras$x$. Então, pode-se dizer que a predição dos métodos *Kernel* é uma “caixa preta”. Além disso, para fazer uma previsão, os *Kernels* precisam examinar toda a base de dados, o que requer uma memória computacional suficiente para armazenar todo o conjunto de dados, e a computação necessária para fazer cada previsão é proporcional ao tamanho da amostra de formação m. Então, no caso de grandes conjuntos de dados, esse método se torna mais lento do que métodos concorrentes. Talvez a maior limitação dos métodos *Kernel* seja estatística já que para qualquer m finito, a acurácia da previsão depende criticamente da escolha da função $dist(x, x')$ e ainda há o risco do desencadeamento da Maldição da Dimensionalidade.

3.5.6 Tipos de *Kernel*

Para que um *Kernel* seja válido, ele deve obedecer às Condições de Mercer. Soman, Loganathan e Ajay (2011) definem o Teorema de Mercercomo uma representação de uma função simétrica, positiva definida em um quadrado cuja soma converge para o produto de funções. Em outras palavras, existe um mapeamento $\phi(x)$ e uma expansão $K(x_i, x_j) = \phi^T(x_i)\phi(x_j)$ se e somente se:

Teorema 1. *Seja χ o domínio de uma função, considere uma função real $K(.,.)$ bivariada, simétrica e contínua definida em $\chi.\chi$. Considere $\oint$ um espaço de característica. Então, existe uma transformação $\phi: \chi \to \oint$ tal que $K(x,y) = \Phi^T(x)\Phi(y)$ se e somente se, K satisfaz a condição de Mercer.*

Portanto, o *Kernel* é admissível quando é uma função bivariada, simétrica e positiva definida em um quadrado integrável, ou seja, o *Kernel* é admissível quando para toda função $g(x) \in L^2(\mathbb{R}^2)\left[i, e \int g^2(x)dx < \infty\right]$. Então, se $\int K(x_i, x_j)g(x_i)g(x_j)d(x_i)d(x_j) \geq 0$, o *Kernel* é dito admissível.

Entretanto, segundo Souza (2010), alguns *Kernels* com funções que não são estritamente positivas definidas também têm apresentado bons desempenhos na prática. Um exemplo é o *SigmoidKernel*, que apesar do seu vasto uso, não é positivo definido para certos valores dos seus parâmetros.

Várias funções podem ser aplicadas como função *Kernel* no SVM e cada uma delas constrói uma superfície de separação diferente no espaço de *input*. O Quadro 9 apresenta os diversos tipos de *Kernel.*

Kernel	**Fórmula**
Linear Kernel	$k(x,y) = x^T y + c$
Polynomial Kernel	$k(x,y) = (\alpha x^T y + c)^d$
Gaussian Kernel	$k(x,y) = \exp\left(-\frac{\| x-y \|^2}{2\sigma^2}\right)$
Exponential Kernel	$k(x,y) = \exp\left(-\frac{\| x-y \|}{2\sigma^2}\right)$
Laplacian Kernel	$k(x,y) = \exp\left(-\frac{\| x-y \|}{\sigma}\right)$
ANOVA Kernel	$k(x,y) = \sum_{k=1}^{n} \exp(-\sigma(x^k - y^k)^2)^d$
Hyperbolic Tangent (Sigmoid) Kernel	$k(x,y) = \tanh(\alpha x^T y + c)$
Rational Quadratic Kernel	$k(x,y) = 1 - \frac{\| x-y \|^2}{\| x-y \|^2 + c}$
Multiquadric Kernel	$k(x,y) = \sqrt{\| x-y \|^2 + c^2}$
Inverse Multiquadric Kernel	$k(x,y) = \frac{1}{\sqrt{\| x-y \|^2 + c^2}}$
Circular Kernel	$k(x,y) = \frac{2}{\pi}\arccos\left(-\frac{\| x-y \|}{\sigma}\right) - \frac{2}{\pi}\frac{\| x-y \|}{\sigma}\sqrt{1-\left(\frac{\| x-y \|}{\sigma}\right)^2}$ if $\| x-y \| < -\sigma$, zero otherwise
Spherical Kernel	$k(x,y) = 1 - \frac{3}{2}\frac{\| x-y \|}{\sigma} + \frac{1}{2}\left(\frac{\| x-y \|}{\sigma}\right)^2$ if $\| x-y \| < -\sigma$, zero otherwise
Wave Kernel	$k(x,y) = \frac{\theta}{\| x-y \|} + \sin\frac{\| x-y \|}{\theta}$

Power Kernel	$k(x,y) = -\| x-y \|^d$
Log Kernel	$k(x,y) = -\log(\| x-y \|^d + 1)$
Spline Kernel	$k(x,y) = \prod_{i=1}^{d} 1 + x_i y_i + x_i y_i \min(x_i, y_i) - \frac{x_i + y_i}{2} \min(x_i, y_i)^2 + \frac{\min(x_i, y_i)^3}{3}$ Com $x, y \in \mathbb{R}^d$
B-Spline (Radial Basis Function) Kernel	$B_n = \frac{1}{n!} \sum_{k=0}^{n+1} \binom{n+1}{k} (-1)^k (x + \frac{n+1}{2} - k)_+^n$ $x_+^d = \begin{cases} x^d, \text{if } x > 0 \\ 0, \text{otherwise} \end{cases}$
Bessel Kernel	$k(x,y) = -Bessel_{(nu+1)}^n(\sigma \lvert x - x' \rvert^2)$
Cauchy Kernel	$k(x,y) = \frac{1}{1 + \frac{\|x-y\|^2}{\sigma^2}}$
Chi-Square Kernel	$k(x,y) = 1 - \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{\frac{1}{2}(x_i + y_i)}$
Histogram Intersection Kernel	$k(x,y) = \sum_{i=1}^{n} \min(x_i, y_i)$
Generalized Histogram Intersection Kernel	$k(x,y) = \sum_{i=1}^{m} \min(\lvert x_i \rvert^\alpha, \lvert y_i \rvert^\beta)$
Generalized T-Student Kernel	$k(x,y) = \frac{1}{1 + \| x-y \|^d}$
Wavelet Kernel	$k(x,y) = \prod_{i=1}^{N} h\left(\frac{x_i - c}{a}\right) h\left(\frac{y_i - c}{a}\right)$

Quadro 9 - Funções *Kernel*

Fonte: Elaboração da autora com base em SOUZA (2010).

Segundo Souza (2010), a escolha do *Kernel* depende em grande parte do problema em questão. A motivação para essa escolha pode ser bastante intuitiva e

depende diretamente do que se está tentando modelar, ou seja, depende do tipo de informação que se espera extrair dos dados.

3.5.7 Formulação do SVM Não Linear com Margem Suave

Com as considerações apresentadas, fica evidente que a matriz *Kernel* substitui a matriz **A** na formulação do SVM e assim, o problema de separação não linear com margem suave, pode ser escrito como:

(58)

$$\text{Minimize} : \varsigma^* = \frac{1}{2}\, w^T w + \mathrm{C}1^T \xi$$

Sujeito à

$$\mathbf{D}[\mathbf{\Phi}(\mathbf{x})w] + \xi \geq 1$$

$$\xi \geq 0$$

Onde,

(59)

$$\mathbf{\Phi} = \begin{pmatrix} \phi(\mathbf{x}_1)^T \\ \phi(\mathbf{x}_2)^T \\ \vdots \\ \phi(\mathbf{x}_n)^T \end{pmatrix} \text{ e } w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

Ou seja, $\mathbf{\Phi}$ é uma matriz de dimensão $n \times q$ e nesse caso o vetor w possui dimensão $q \times 1$. Assim como no caso linear, para a resolução do problema (58) é mais interessante trabalhar com o Dual de Wolfe (1961). Então, primeiramente é preciso, encontrar a função Lagrangeanaque pode ser escrita como:

(60)

$$\mathrm{L}(\mathrm{w}, \gamma, \lambda, \mu, \xi) = \frac{1}{2} w^T w + \mathrm{C}1^T \xi - \lambda^T [\mathbf{D}[\mathbf{\Phi}(\mathbf{x})w] + \xi - 1] - \mu^T \xi$$

Resolvendo as condições de primeira ordem, se tem:

(61)

$$\frac{\mathrm{d}}{\mathrm{d}w}\mathrm{L}(\mathrm{w},\gamma,\lambda,\mu)=0\rightarrow\frac{\mathrm{d}}{\mathrm{d}w}\left(\frac{1}{2}w^{T}w+\mathrm{C1}^{T}\xi-\lambda^{T}\mathbf{D\Phi}(\mathbf{x})w-\lambda^{T}\xi+\lambda^{T}-\mu^{T}\xi\right)=0$$
$$\rightarrow w^{T}-\lambda^{T}\mathbf{D\Phi}(\mathbf{x})=0\rightarrow\ w^{T}=\lambda^{T}\mathbf{D\Phi}(\mathbf{x})$$

(62)

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\mathrm{L}(\mathrm{w},\gamma,\lambda,\mu)=0\rightarrow\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\frac{1}{2}w^{T}w+\mathrm{C1}^{T}\xi-\lambda^{T}\mathbf{D\Phi}(\mathbf{x})w-\lambda^{T}\xi+\lambda^{T}-\mu^{T}\xi\right)=0$$
$$\rightarrow\ \mathrm{C1}^{T}-\lambda^{T}-\mu^{T}=0$$

Substituindo,

(63)

$$\mathrm{L}(\lambda)=\frac{1}{2}\lambda^{T}\mathbf{D\Phi}(\mathbf{x})\mathbf{\Phi}(\mathbf{x})^{\mathbf{T}}\mathbf{D}\lambda-\lambda^{T}\mathbf{D\Phi}(\mathbf{x})\mathbf{\Phi}(\mathbf{x})^{\mathbf{T}}\mathbf{D}\lambda+\mathrm{C1}^{T}\xi-\lambda^{T}\xi-\mu^{T}\xi+\lambda^{T}1$$

(64)

$$\mathrm{L}(\lambda)=-\frac{1}{2}\lambda^{T}\mathbf{D\Phi}(\mathbf{x})\mathbf{\Phi}(\mathbf{x})^{\mathbf{T}}\mathbf{D}\lambda+\xi(\mathrm{C1}^{T}-\lambda^{T}-\mu^{T})+\lambda^{T}1$$

Sendo assim, o Dual de Wolfe (1961) do SVM Não Linear com Margem Suave é:

(65)

$$\mathrm{L}(\lambda)=-\frac{1}{2}\lambda^{T}\mathbf{D\Phi}(\mathbf{x})\mathbf{\Phi}(\mathbf{x})^{\boldsymbol{T}}\mathbf{D}\lambda+\lambda^{T}1$$

Sujeito à

$0\leq\lambda\leq\mathrm{C1}$

$\lambda\geq 0$

3.5.8 Parâmetros do SVM

Segundo Kim (2003), uma das vantagens do SVM é o fato dele depender de um pequeno número de parâmetros, diferente da maioria modelos de previsão. Entretanto, a escolha destes, embora sejam poucos, é essencial para o bom desempenho do modelo. Não há um valor definido para cada uma das constantes e a escolha delas é mais um desafio.

O valor do parâmetro C é importante porque indica o peso que se dá para uma classificação errada, ou seja, quanto custa o erro. Tay e Cao(2001) inferiram que no conjunto de treinamento, um valor muito pequeno para C causaria *underfitting* e um valor muito alto, *overfitting*. Por isso, defendem que o valor apropriado para o C é entre 1 e 100. No estudo de Kim (2003), os resultados corroboram essa hipótese, pois o desempenho do SVM no conjunto de validação aumenta quando$C$ aumenta de 1 a 78, mas cai quando Cé 100.

Na literatura existente a maioria das aplicações utiliza a validação cruzada no processo de seleção dos parâmetros. Gupta, Mehlawat e Mittal (2012) por exemplo, determinaram os valores de C e σpor meio de uma validação cruzada com 10 etapas e chegaram aos valores ótimos de 2^{25} e 2^{-3}, respectivamente. Fan e Palaniswami (2001) por sua vez, utilizaram um ano de dados para selecionar os parâmetros C e σ que oferecessem maior acurácia, por meio da validação cruzada. A Classe $+1$, composta pelos ativos que apresentaram retorno excepcional correspondeu a um terço da Classe -1, composta pelos ativos com retornos "normais". Assim, os dados de treinamento do estudo foram sempre não balanceados e por essa razão foram atribuídos valores diferentes de C para as diferentes classes.

Já Huerta, Corbacho e Elkan (2013) para escolher os parâmetros C e σ que maximizam o desempenho do modelo na etapa de validação, optaram por criar pares com valores de C e σ, usados para construir um portfólio no tempo t e estes pares foram chamados de a. Foi definido um conjunto de 16 pares com $C = 0.5, 1, 2, 4$ e $\sigma = 0.5, 1, 2, 4$. A qualidade da escolha é definida como $Q(t, a)$, sendo atualizada com uma média exponencial móvel, descrita por:

(66)

$$Q(t, a) = (1 - \alpha)Q(t - 1, a) + \alpha R(t - 1, a)$$

Onde α é uma taxa de aprendizagem escolhida com a média de três anos e $R(t - 1, a)$ é a remuneração obtida com a escolha a no período imediatamente anterior. Para fins de aplicação, os autores consideraram o retorno puro como $R(t - 1, a)$ e os valores dos meta parâmetros usados para formar o portfólio no tempo t foram obtidos por meio de:

(67)

$$a_t = \arg\max_a Q(t, a)$$

Para o parâmetro *Kernel* σ, Tay e Cao (2001) encontraram o intervalo de 1 a 100 como o mais adequado e sugerem o raciocínio contrário ao proposto para o C. Um valor muito pequeno de σ geraria *overfitting* nos dados de treinamento e um σgrande causaria *underfitting*.Os resultados de Kim (2003) mostram que o desempenho do SVM nos dados de treinamento diminuiu com o valor de σ e no conjunto de validação, o desempenho se manteve estável com a variação de parâmetro de 25 a 100.

4 RESULTADOS E DISCUSSÃO

Nesta pesquisa a Máquina de Suporte Vetorial foi construída tendo como insumos os resultados trimestrais dos 15 indicadores financeiros escolhidos. Infere-se que estes 15 indicadores explicam o retorno trimestral líquido do período seguinte de cada ativo, e este por sua vez, determina a classificação da ação em uma das duas classes ($y_i = +1 \text{ ou} y_i = -1$).

O SVM foi treinado com os dados de treinamento das 42 ações listadas no Quadro 8, no período do primeiro trimestre de 2000 ao terceiro trimestre de 2006. Já a previsão foi feita para estas mesmas ações no recorte temporal de validação, compreendido entre o quarto trimestre de 2006 até o primeiro trimestre de 2009. O modelo probabilístico da função *ksvm* foi utilizado para que o SVM interpretasse os *outputs* como a probabilidade dos ativos seremclassificados como $+1$. Dessa forma, a função de decisão do SVM não classificou os ativos nas Classes 1 e 2 apenas pelo sinal do *output*, mas sim pela sua probabilidade de assumir o valor $+1$. A Classe 1 foi composta então pelos 25% de ações com maiores probabilidades e a Classe 2, pelo restante dos ativos.

No *grid* criado para a definição dos melhores parâmetros, o espaço paramétrico de σ foi definido empiricamente como $\sigma = [0{,}0001\,, 2]$. Apesar de Kim (2003) sugerir o intervalo $\sigma = [25{,}100]$ e Tay e Cao (2012) sugerirem $\sigma = [1{,}100]$, os dados destes estudos foram padronizados em intervalos de [-1,1] e [-0,9 , 0,9], respectivamente, o que difere do estado dos dados desta pesquisa.

Quanto ao parâmetro C, no estudo de Fan e Palaniswami (2001), como as Classes 1 e 2 eram não balanceadas, já que a primeira representava apenas25% do total dos ativos enquanto asegunda representava 75%, foram utilizados dois valores de C diferentes, um para cada classe. Nesta pesquisa, as classes também foram constituídas de forma não balanceada, entretanto, visando uma simplificação, apenas um valor de C foi utilizado. O espaço paramétrico dessa constante foi então definido empiricamente como [1,100], o que coincidiu com o intervalo sugerido por Tay e Cao (2001) e corroborado por Kim (2003).

Nesta pesquisa chegou-se aos valores ótimos de $\mathrm{C} = 1$ e $\sigma = 0{,}0001$que resultaram em uma acurácia de 0,7347932. Ou seja, nas classificações de ativos

feitas pelo SVM, utilizando estes parâmetros ótimos, a máquina acertou a classificação em 73,48% das vezes. Este é um resultado bastante satisfatório, entretanto para comprovar a aplicabilidade do SVM na formação de portfólios, o retorno gerado pela máquina foi comparado com o retorno gerado pelo ETF BOVA11 no mesmo período.

O portfólio formado por meio do SVM apresentou um retorno trimestral médio de 8,26% e o BOVA11 teve um retorno médio de 1,64%, portanto, o SVM superou o *benchmark* em 403,92%. Considerando os 19 trimestres testados, o retorno acumulado do SVM foi de 257,36% e o do BOVA11, 19,33%.

No que tange ao risco dos investimentos, o BOVA11 apresentou um risco trimestral médio de 12,21% e o portfólio, 18,29%. Como foram encontrados 22 pares de parâmetros ótimos, ou seja, 22 pares retornaram a acurácia máxima e de forma arbitráriao primeiro par foi escolhido, pode-se inferir que a volatilidade do portfólio é influenciada não apenas pelos retornos obtidos, mas também pelos parâmetros utilizados no modelo. Estes parâmetros ainda podem ser melhorados em grande medida e mesmo que se trabalhe com a mesma acurácia, existe a possibilidade de alteração nos valores de risco e retorno caso outro par ótimo seja utilizado.

Ao longo do recorte temporal delimitado para esta pesquisa, o ativo BOVA11 teve de fato resultados muito baixos devido ao contexto econômico do período, o que aumentou a discrepância entre os retornos. Então, calculou-se também o retorno de outro *benchmark* de mercado composto por todas as 67 açõesda carteira teórica do Ibovespa utilizada nesta pesquisa. Ou seja, os retornos trimestrais médios foram calculados com todas as ações da carteira e não apenas com aquelas que foram classificadas como boas e por isso foram selecionadas pelo SVM. O retorno trimestral médio deste segundo *benchmark* foi de 7,12% e o retorno acumulado foi de 183,41%, com risco trimestral médio de 19,89%. Portanto, novamente o retorno do portfólio escolhido pelo SVM foi superior ao *benchmark*, dessa vez, em 16,08% e com um risco 8,78% menor.

No que tange a relação retorno/risco, obteve-se 0,4516 para o portfólio formado pelo SVM, 0,1343 para o ETF BOVA11 e 0,3579 para o *benchmark* de Mercado. Isso significa que para cada 1% de risco no portfólio selecionado pelo SVM, há 0,45% de retorno. Já para BOVA 11, cada 1% de risco vem acompanhado por 0,13% de retorno. O benchmark de Mercado , por sua vez, apresentou 0,35%

de retorno para cada 1% de risco. Como essa razão relaciona a maximização do retorno e a minimização do risco, quanto maior o valor do quociente, melhor é a opção de investimento. Dessa forma, mais uma vez, os resultados obtidos com o uso do SVM foram superiores. A Tabela 3 agrupatodos os resultados encontrados na pesquisa.

Tabela 3- Resultados da Pesquisa

Opção de Investimento	Retorno Acumulado	Retorno Trimestral Médio	Risco Trimestral Médio	Retorno/Risco
Portfólio SVM	257,36%	8,26%	18,29%	0,4516
BOVA11	19,33%	1,64%	12,21%	0,1343
Benchmark de Mercado	183,41%	7,12%	19,89%	0,3579

Fonte: Elaborado pela autora.

Segundo o *Booststrap*aplicado, 75%dos retornos ficaram acima de 53,76%, 50% ficaram acima de 88,40% e 25% acima de 117,65%. O retorno mínimo encontrado foi de –39,38% e o máximo foi de 182,74%. Para a probabilidade de um retorno ser positivo, foi encontrado o valor de 97,88%. O histograma abaixo explicita a média dos retornos de 84,68% e a distribuição destes ao longo de todo período de teste. As barras vermelhas indicam a posição dos quartis.

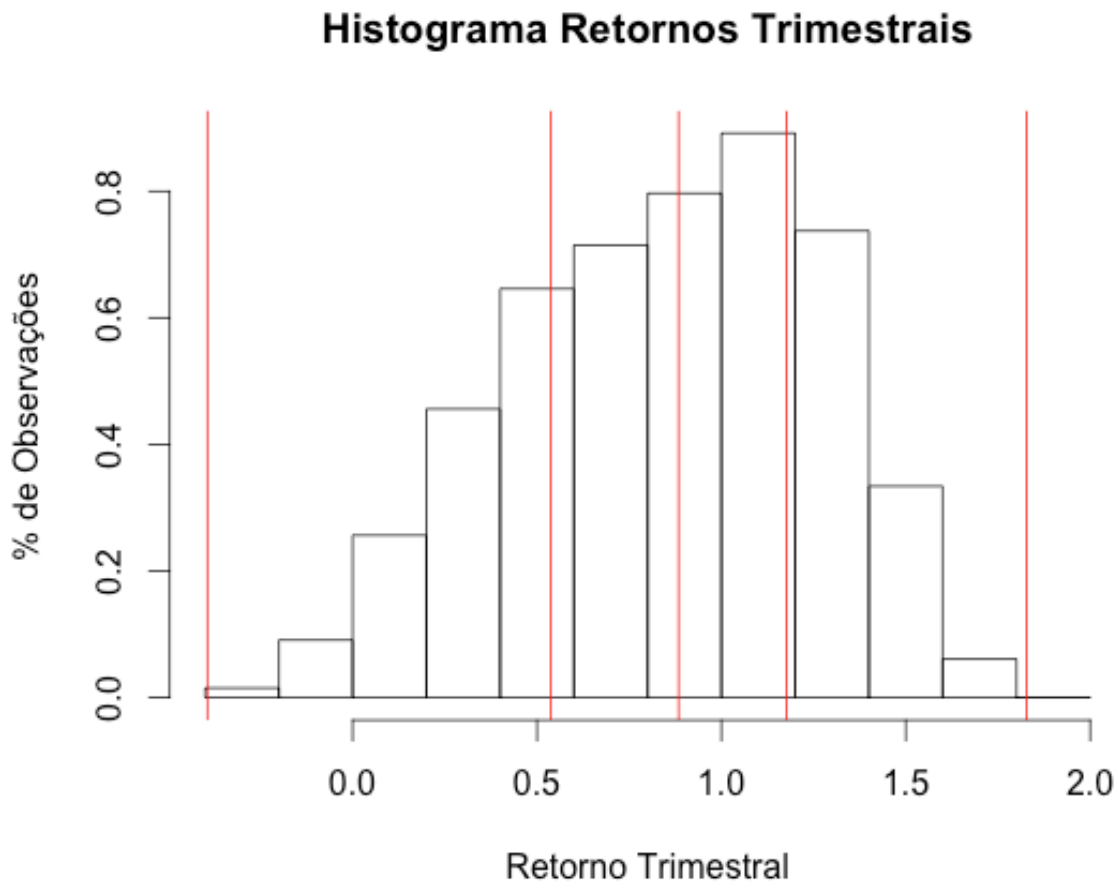


Gráfico 3 –Histograma da distribuição dos retornos trimestrais pelo *Bootstrap*

Fonte: Elaborado pela autora.

5 CONCLUSÕES E RECOMENDAÇÕES

Esta pesquisa visou replicar o modelo de Máquinas de Suporte Vetorial proposto por Fan e Palaniswami (2001) para formação de portfólios, aplicado ao contexto brasileiro. Para tal, as Máquinas de Suporte Vetorial foram utilizadas paraverificar seo uso do SVM na formação de portfólios de fato contribui para que o retorno seja superior ao de um *benchmark* do mercado, sendo que o ativo escolhido para tal comparação foi o fundo de índiceBOVA11.

A população desta pesquisa consistiu em todas as ações disponíveis para negociação na BM&FBOVESPA e a amostra foi constituída pelas 71 ações que compuseram a carteira teórica válida para 2 de Setembro de 2013 a 03 de Janeiro de 2014. Os históricos de preço e informações contábeis foram coletados na base de dados do sistema Economáticano recorte temporal de 2000 a 2013.Devido a algumas limitações dessa base, a amostra foi reduzida para 67 ações e foram utilizados apenas 15 indicadores financeiros.

Assim como no estudo de Fan e Palaniswami (2001), a Validação Cruzada foi utilizada para evitar *overtraining*e para isso os dados foram divididos em três conjuntos com recortes temporais diferentes. O primeiro conjunto foi o de treinamento, constituído por 42 ações no período do primeiro trimestre de 2000 ao terceiro trimestre de 2006, totalizando 26 trimestres. Já o segundo conjunto, foi composto pelas mesmas 42 ações, mas no período do quarto trimestre de 2009 ao primeiro trimestre de 2009, totalizando 10 trimestres. Assim, no recorte temporal do primeiro trimestre de 2000 até o primeiro de 2009, 70% foi usado paratreinamento e 30%para validação. O terceiro conjunto foi o de teste e este foi formado por 25 ações, diferentes das 42 utilizadas nos outros conjuntos. O período de teste compreendeu o período do segundo trimestre de 2009 ao quarto trimestre de 2013, totalizando 19 trimestres.

Primeiramente, o SVM foi utilizado para classificar as ações em duas Classes: Classe 1$(y = +1)$, das melhores ações, e Classe 2$(y = -1)$, das piores ações. Os *outputs* do SVM foram convertidos em probabilidades dos ativos serem classificados como $+1$ e por meio do *ranking* dessas probabilidades, as ações que

ficaram dentro do conjunto dos 25% de maiores probabilidades foram classificadas como pertencentes à Classe 1 e as demais, como pertencentes à Classe 2.

No processo de replicação da metodologia de Fan e Palaniswami (2001), algumas adaptações tiveram que ser feitas para a aplicação no contexto brasileiro. Primeiramente, como as datas de divulgação dos resultados dos indicadores financeiros coletados estavam organizadas de forma trimestral, a classificação de ativos pelo SVM foi feita trimestralmente e não anualmente, como aconteceu na análise de Fan e Palaniswami (2001).Em segundo lugar, a técnica de Análise dos Componentes Principais utilizada por estes autores para agrupar os indicadores não se mostrou agregadora ao contexto desta pesquisa já que foram utilizados apenas 15 indicadores e, por isso, não foi utilizada. Esta pesquisa também diferiu do estudo de Fan e Palaniswami (2001) no fato de ter sido utilizado apenas um parâmetro$C$ para ambas as classes, mesmo com odesbalanceamento entre elas.

Apesar das limitações, os resultados desta pesquisa, corroboram para a hipótese de aplicabilidade das Máquinas de Suporte Vetorial na abordagem de formação de portfólio.A melhor performance obtida para o modelo foi de 73,48% de acurácia e ela foi encontrada para 22 pares distintos de parâmetros, sendo o primeiro deles$C = 1$ e $\gamma = 0{,}0001$. A máquina foi construída com estes parâmetros para a formação de portfólios trimestrais durante aproximadamente 5 anos em um conjunto de teste.O BOVA11 gerou retorno trimestral médio 1,64% e o SVM apresentou a média de 8,26%, superando o *benchmark* em 403,92%. No fim dos 19 trimestres testados, o retorno acumulado do SVM foi de 257,36% e o do BOVA11, 19,33%.No que tange ao risco dos investimentos, o BOVA11 apresentou um risco trimestral médio de 12,21% e o portfólio, 18,29%.

Como de fato os resultados do BOVA11 foram muito baixos devido ao contexto econômico ao longo do recorte temporal delimitado para esta pesquisa, outro *benchmark* foi calculado. Ele foi constituído de todas 67 as ações da amostra dessa pesquisa. Dessa forma, os retornos trimestrais médios foram calculados com todas as ações da carteira e não apenas com aquelas que compuseram o portfólio selecionado pelo SVM. O retorno trimestral médio deste segundo *benchmark* foi de 7,12% e o retorno acumulado foi de 183,41%, com risco médio trimestral de 19,89%. Novamente o retorno do portfólio escolhido pelo SVM foi superior e com um risco 8,78% menor.

O portfólio formado pelo SVM também apresentou a melhor relação retorno/risco de 0,4516, contra 0,1343 do BOVA11 e 0,3579 do *benchmark* de Mercado.

As 10 000 amostragens do *Bootstrap*apontaram que25% dos retornos ficaram acima de 53,76%, 50% ficaram acima de 88,40% e 25% acima de 117,65%. Além disso, o retorno mínimo encontrado foi de –39,38%, o máximo foi de 182,74% e para a probabilidade de um retorno ser positivo, foi encontrado o valor de 97,88%.

Os resultados de previsão obtidos e o teste de significância estatística aplicado corroboram a hipótese de superioridade do método inovador das Máquinas de Suporte Vetorial na formação de portfólios, caracterizado pela construção de um hiperplano que separe os dados em duas classes ou mais, para atingir a separação máxima entre elas e pela implementação do Princípio da Minimização do Risco Estrutural, o qual procura minimizar o limite superior do erro de generalização, em vez de minimizar apenas o erro do processo de estimação.

Muitas lacunas na abordagem de formação de portfólios por meio das Máquinas de Suporte Vetorial podem ser exploradas. Para estudo futuros, sugere-se o aprimoramento da forma de definição dos parâmetros ótimos, construção da máquina com diferentes tipos de *Kernel*ou combinações deles e definição dos *inputs*, ou seja, variáveis mais adequadas para o modelo.Aprofundamentos comoestes poderão elevar em grande medida a acurácia de classificação e o retorno gerado pelo portfólio, contribuindoassim, para o aperfeiçoamento do método.

REFERÊNCIAS

ABU-MOSTAFA, Y. S.; ATIYA, A. F. Introduction to financial forecasting.**AppliedIntelligence**, Springer, v. 6, n. 3, p. 205–213, 1996.

ALBUQUERQUE, Pedro H. M. Previsão de séries temporais financeiras por meio de máquinas de suporte vetorial e ondaletas. 2014. Tese, pós-doutorado. Universidade de São Paulo, Instituto de Matemática e Estatística, São Paulo.

BM&FBOVESPA. **Carteira Teórica Ibovespa válida para 15/6/2014**. Disponível em: <http://www.bmfbovespa.com.br/indices/ResumoCarteiraTeorica.aspx?Indice=Ibovespa&idioma=pt-br>. Acesso em: 15 de Junho de 2014.

BM&FBOVESPA. **Metodologia do índice Bovespa**. Disponível em: <http://www.bmfbovespa.com.br/Indices/download/IBOV-Metodologia-pt-br.pdf>. Acesso em: 2 de Junho de 2014.

BOSER, Bernhard E.; GUYON, Isabelle M.; VAPNIK, Vladimir N.A Training Algorithm for Optimal Margin Classifiers.In: ANNUAL WORKSHOP ON COMPUTACIONAL LEARNING, 5, 1992, Pittsburgh. **ACM Press**. Pittsburgh: Haussler D, jul 1992. p.144-152 .

BRUNI, Adriano Leal; FUENTES, Júnio; FAMÁ, Rubens. A moderna teoria de portfólios e a contribuição dos mercados latinos na otimização da relação risco *versus* retorno de carteira internacionais: Evidências Empíricas Recentes (1996-1997).III Semead, FEA/USP, 1998.

DOWNES, John; GOODMAN, Jordan E.Dicionário de termos financeiros de investimento.São Paulo: Nobel, 1993. 645 p.

EMIR, Senol; DINÇER, Hasan; TIMOR, Mehpare. A stock selection model based on fundamental and technical analysis variables by using artificial neural networks and support vector Machines. **Review of Economics & Finance**, Istanbul, p. 106-122, mar. 2012.

FAN, Alan; PALANISWAMI, Marimuthu. Stock selection using support vector machines. **Neural Networks, 2001. Proceedings. IJCNN 2001. (IEEE World Congress on Computational Intelligence).IEEE International Joint Conference on.** [S.l.], 2001. v. 3, p. 1793–1798.

GUPTA, Pankaj; MEHLAWAT, Kumar M.; MITTAL, Garima.Asset portfolio optimization using support vector machines and real-coded genetic algorithm.**Journal of Global Optimization**, Springer, v. 53, n. 2, p. 297–315, 2012.

HUERTA, Ramon; CORBACHO, Fernanda; ELKAN, Charles. Nonlinear support vector machines can systematically identify stocks with high and low future returns. **Algorithmic Finance**, IOS Press, v. 2, n. 1, p. 45–58, 2013.

KIM, Kyoung-jae. Financial time series forecasting using support vector machines. **Neurocomputing**, Elsevier, v. 55, n. 1, p. 307–319, 2003.

LAI, Kin K. *et al*. A Double-Stage Genetic Optimization Algorithm for Portfolio Selection. In: INTERNACIONAL CONFERENCE, 13, ICONIP 2006, Hong Kong, China. Springer-Verlag Berlin Heidelberg, oct 2006. p. 929-937.

LU, Ruei-Shan; YU, Shang-Wu; LIN, Yi-Hsien.The prediction of applying smooth support vector regression and back propagation network in mutual fund performance.**Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence).IEEE International Joint Conference on**. [S.l.], 2008. p. 3192–3196.

MARKOWITZ, Harry. Portfolio selection. **Journal of Finance**, vol.7, n. 1, p. 77-91, mar. 1952.

MERCER, John.Functions of positive and negative type, and their connection with the theory of integral equations.**Philosophical Transactions of the Royal Society of London.Series A, Containing Papers of a Mathematical or Physical Character,** The Royal Society, v. 209, p. 415–446, 1909. ISSN 02643952. Disponívelem: <http://www.jstor.org/stable/91043>.

SOMAN, K. P.; LOGANATHAN, R.; AJAY, V. Machine Learning with SVM and Other Kernel Methods. 1 ed. New Delhi: PHI Learning Private Limited, 2011. 477 p.

SOUZA, César, R. Kernel Functions for Machine Learning Applications. .Net, Rio de Janeiro, mar. 2010. Disponível em: < http://crsouza.blogspot.com.br/2010/03/kernel-functions-for-machine-learning.html >. Acessoem: 2 jul. 2014.

TAY, Francis. E.H.; CAO, Lijuan. Application of support vector machines in financial time series forecasting. **Omega**, Elsevier, v. 29, n. 4, p. 309–317, 2001.

WOLFE, P. A duality theorem for non-linear programming.**Quarterly of Applied Mathematics**, n. 19, p. 239–244, 1961.

YU, Shang-Wu; LU, Ruei-Shan.; CHANG, Chia-Hao. A study on application of smooth support vector classification to stock selection in taiwan's stock market. **The International Conference on Computational Intelligence in Economics and Finance, 7, 2008.Kainan University, Taoyuan, Taiwan, 2008.**

ZHANG, Zuoquan; ZHAO, Qin. The application of svms method on exchange rates fluctuation. **Discrete Dynamics in Nature and Society**, Hindawi Publishing Corporation, v. 2009, 2009.

APÊNDICES

Apêndice A –Indicadores Utilizados nos Estudos Apresentados

Indicador	Fórmulaouvariável	Artigos que usaram o indicador
%D (3-period moving average of %K)	$\frac{\sum_{i=0}^{\%n-1} \%K_{t-i}}{n}$	Kim (2003)
A/D Oscilador	$\frac{H_t - C_{t-1}}{H_t - L_t}$	Kim (2003)
Accounts Receivable	AR	Huerta, Corbacho e Elkan (2013)
Accrual Based on Balance Sheet	$\mathrm{ABBS} = \mathrm{SAC}(quarter) - \mathrm{SAC}(quarter - 4)$	Huerta, Corbacho e Elkan (2013)
Accrual Based on Cash Flow	ABCF	Huerta, Corbacho e Elkan (2013)
Average True Range (ATR)	$\mathrm{ATR}_t = \frac{\mathrm{ATR}_{t-1} \times (n-1) + \mathrm{TR}_t}{n}$ O primeiro ATR é calculado usando a seguinte média aritmética: $\mathrm{ATR}_t = \frac{1}{n}\sum_{i=1}^{n} TR_i$	Emir, Dinçer e Timor (2012)
BIAS	$BIAS(n) = \frac{C_t - MA(n)}{MA(n)} \times 100\%$	Zhang e Zhao (2009)
Book Value	BV	Huerta, Corbacho e Elkan (2013)
Book Value - Total Debt	$\mathrm{BV} - \mathrm{TD}$	Huerta, Corbacho e Elkan (2013)
Capital Expenditures	CEx	Huerta, Corbacho e Elkan (2013)
Cash and Equivalents	CE	Huerta, Corbacho e Elkan (2013)
Cash Flow / Current Liabilities	$\mathrm{CF/CL}$	Fan e Palaniswami (2001)
Cash Flow / Sales	$\mathrm{CF/S}$	Fan e Palasniwami (2001)

Cash Flow / Total Assets	$\mathrm{CF/TA}$	Fan e Palaniswami (2001)
Cash Flow / Total Capital	$\mathrm{CF/TC}$	Fan e Palaniswami (2001)
Cash Flow / Total Market Value	$\mathrm{CF/MV}$	Fan e Palaniswami (2001)
Cash From Financing Activities	CFFA	Huerta, Corbacho e Elkan (2013)
Cash From Investing Activities	CFIA	Huerta, Corbacho e Elkan (2013)
Cash From Operating Activities	CFOA	Huerta, Corbacho e Elkan (2013)
Cash to Assets	$\mathrm{CE/TA}$	Huerta, Corbacho e Elkan (2013)
Chaikin Money Flow Indicator (CMF)	$CMF = \frac{\sum_{t=20}^{t} CLV_t \times volume_t}{\sum_{t=20}^{t}(vol_t)}$ onde, $CLV = \frac{(close_1 - low_1) - (high_1 - close_1)}{(high_1 - low_1)}$	Emir, Dinçer e Timor (2012)
Commodity Channel Index (CCI)	$\frac{(M_t - SM_t)}{(0.015\, D_t)}$ onde: $M_t = (\frac{H_t + L_t + C_t}{3})$ $SM_t = \frac{\sum_{i=1}^{n} M_{t-i+1}}{n}$	Emir, Dinçer e Timor (2012); Kim (2003)
Current Assets / Assets	$\mathrm{CA/TA}$	Emir, Dinçer e Timor (2012)
Current Assets / Current Liabilities	$\mathrm{CA/CL}$	Fan e Palaniswami (2001)
Current Liabilities / Equity	$\mathrm{CL/E}$	Fan e Palaniswami (2001)
Current Liabilities / Total Assets	$\mathrm{CL/TA}$	Fan e Palaniswami (2001)
Debt / Equity	$\mathrm{TD/E}$	Fan e Palaniswami (2001); Huerta, Corbacho e Elkan (2013)

Debt to Assets	$\mathrm{TD/TA}$	Huerta, Corbacho e Elkan (2013)
Depreciation	DP	Huerta, Corbacho e Elkan (2013)
Diluted Normalized Earnings per Share	DNES	Huerta, Corbacho e Elkan (2013)
Disparity10	$\frac{C_t}{MA_{10}} \times 100$	Kim (2003)
Disparity5	$\frac{C_t}{MA_5} \times 100$	Kim (2003)
Dividend Yield	DY	Fan e Palaniswami (2001)
Dividend Payout Ratio	$\mathrm{DPR} = \mathrm{D/NI}$	Huerta, Corbacho e Elkan (2013)
Dividends	D	Huerta, Corbacho e Elkan (2013)
Earning After Tax Growth	EATG	Fan e Palaniswami (2001)
Earning Before Tax Growth	EBTG	Fan e Palaniswami (2001)
Earning Yield	$\mathrm{EY} = \mathrm{NI} - \mathrm{DPS/AOS}$	Fan e Palaniswami (2001); Emir, Dinçer e Timor (2012)
Equity / Assets	$\mathrm{E/TA}$	Emir, Dinçer e Timor (2012)
Equity / Tangible Assets	$\mathrm{E/TGA}$	Emir, Dinçer e Timor (2012)
Equity Growth	EG	Emir, Dinçer e Timor (2012)
Exponential Moving Average (EMA)	$EMA_{todau} = EMA_{yesterday} + \alpha(price_{todau} - EMA_{yesterday})$	Emir, Dinçer e Timor (2012)
Financial Health	FH	Huerta, Corbacho e Elkan (2013)
Fixed Assets / Assets	$\mathrm{FA/TA}$	Emir, Dinçer e Timor (2012)
Gross Profit	GP	Huerta, Corbacho e Elkan (2013)
Growth in Assets	GA	Emir, Dinçer e Timor (2012)

Growth in Net Profit	GNP	Emir, Dinçer e Timor (2012)
Income After Tax	IAT	Huerta, Corbacho e Elkan (2013)
Income Before Tax	IBT	Huerta, Corbacho e Elkan (2013)
Liabilities to Income	$\mathrm{TL/NI}$	Huerta, Corbacho e Elkan (2013)
Long Term Debt / Total Debt	$\mathrm{LTD/TD}$	Fan e Palaniswami (2001)
Market to Book Value	$\mathrm{MV/BV}$	Emir, Dinçer e Timor (2012)
Mass Index (MASS)	$\sum_{1}^{25} \frac{9-\text{day EMA of (High} - \text{Low)}}{9-\text{day EMA of a } 9-\text{day EMA of (High} - \text{Low)}}$	Emir, Dinçer e Timor (2012)
Momentum	$C_t - C_{t-4}$	Kim (2003); Emir, Dinçer e Timor (2012)
Money Flow Index (MFI)	$100 - \frac{100}{(1 + \text{Money Flow Ratio})}$	Emir, Dinçer e Timor (2012)
Moving Average Convergence and Divergence Line	$MCD = 2(DIF - DEA)$ $DIF = EMA(12) - EMA(26)$ $DEA_t = \frac{2}{10} dif + \frac{8}{10} DEA_{t-1}$ $EMA_t(n) = \frac{2}{N+1} C_t + \frac{N-1}{N} + 1MA_{t-1}(n)$	Zhang e Zhao (2009); Emir, Dinçer e Timor (2012)
Moving Average Line	$MA_t(n) = \frac{1}{N} C_t + \frac{N-1}{N} MA_{t-1}(n)$	Zhang e Zhao (2009)
Net Change In Cash	NCIC	Huerta, Corbacho e Elkan (2013)
Net Income	NI	Huerta, Corbacho e Elkan (2013)
Net Income / Sales	$\mathrm{NI/S}$	Fan e Palaniswami (2001)
Net Income / Total Capital	$\mathrm{NI/TC}$	Fan e Palaniswami (2001)
Net Income Before Extraordinary Items	NIBEI	Huerta, Corbacho e Elkan (2013)
Net Profit / Current Assets	$\mathrm{NP/CA}$	Emir, Dinçer e Timor (2012)

Net Recurring Profit Growth	NTPG	Fan e Palaniswami (2001)
Net Tangible Assets per Share	NTAS	Fan e Palaniswami(2001)
Operating Income	OI	Huerta, Corbacho e Elkan (2013)
Operating Profit Growth	OPG	Fan e Palaniswami (2001)
Profit After Tax / Current Liabilities	PAT/CL	Fan e Palaniswami (2001)
Price-Earnings Ratio	MCPS/EPS	Fan e Palaniswami (2001); Emir, Dinçer e Timor (2012)
Price Oscilator (OSCP)	$\frac{M_5 - M_{10}}{MA_5}$	Kim (2003)
Price Rate of Change (ROC)	$\frac{C_t}{C_{t-n}} \times 100$	Kim (2003)
Profit After Tax / Cash Flow	PAT/CF	Fan e Palaniswami (2001)
Profit After Tax / Equity	PAT/E	Fan e Palaniswami (2001)
Profit After Tax / Sales	PAT/S	Fan e Palaniswami (2001)
Profit Before Tax / Current Liabilities	PBT/CL	Fan e Palaniswami (2001)
Profit Before Tax / Sales	PBT/S	Fan e Palaniswami (2001)
Profit Before Tax / Total Assets	PBT/TA	Fan e Palaniswami (2001)
Profit Before Tax / Total Capital	PBT/TC	Fan e Palaniswami (2001)
Quick Ratio	(TCA − TI)/CL	Huerta, Corbacho e Elkan (2013)
Random Index	$K_t = \frac{2}{3}K_{t-1} + \frac{1}{3}RSV_t$ $D_t = \frac{2}{3}D_{t-1} + \frac{1}{3}K_t$ $J = 3D - 2K$ $RSV_t = \frac{C_t - L_n}{H_n - L_t} \times 100$	Zhang e Zhao (2009)

Receivables to Sales	$\mathrm{AR/TR}$	Huerta, Corbacho e Elkan (2013)
Relative Strength Index (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i}/n) \,/\, (\sum_{i=0}^{n-1} Dw_{t-i}/n)}$	Emir, Dinçer e Timor (2012); Kim (2003); Zhang e Zhao (2009)
Return on Assets	$\mathrm{ROA} = \mathrm{NI/TA}$	Fan e Palaniswami (2001); Emir, Dinçer e Timor (2012)
Return on Equity	$\mathrm{ROE} = \mathrm{NI/TE}$	Huerta, Corbacho e Elkan (2013); Emir, Dinçer e Timor (2012)
Sales Growth	SG	Fan e Palaniswami (2001)
Sales per Share	$\mathrm{TR/SO}$	Huerta, Corbacho e Elkan (2013)
Share Holders' Equity / Total Market Value	$\mathrm{SHE/MV}$	Fan e Palaniswami (2001)
Shareholders' Fund Growth	SFG	Fan e Palaniswami (2001)
Short Term Investiments	STI	Huerta, Corbacho e Elkan (2013)
Short Term Liabilities	STL	Huerta, Corbacho e Elkan (2013)
Slow %D	$\frac{\sum_{i=0}^{\%n-1} \%D_{t-i}}{n}$	Kim (2003)
Snapshot Accrual	$\mathrm{SAC} = \mathrm{TCA} - \mathrm{CE} - \mathrm{CL} + \mathrm{TD}$	Huerta, Corbacho e Elkan (2013)
Stochastics %K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$	Emir, Dinçer e Timor (2012); Kim (2003)
Total Assets	TA	Huerta, Corbacho e Elkan (2013)
Total Assets / Shareholders' Equity	$\mathrm{TA/SE}$	Fan e Palaniswami (2001)
Total Assets / Total Market Value	$\mathrm{TA/MV}$	Fan e Palaniswami (2001)
Total Assets Growth	TAG	Fan e Palaniswami (2001)
Total Current Assets	TCA	Huerta, Corbacho e Elkan (2013)

Total Current Liabilities	CL	Huerta, Corbacho e Elkan (2013)
Total Debt	TD	Huerta, Corbacho e Elkan (2013)
Total Equity	TE	Huerta, Corbacho e Elkan (2013)
Total Inventory	TI	Huerta, Corbacho e Elkan (2013)
Total Liabilities	TL	Huerta, Corbacho e Elkan (2013)
Total Liabilities / Shareholders' Equity	$\mathrm{TL/SE}$	Fan e Palaniswami (2001)
Total Liabilities / Total Capital	$\mathrm{TL/TC}$	Fan e Palaniswami (2001)
Total Long Term Debt	TLTD	Huerta, Corbacho e Elkan (2013)
Total Revenue	TR	Huerta, Corbacho e Elkan (2013)
Total Shares	SO	Huerta, Corbacho e Elkan (2013)
Triple Exponential Smoothing of the Log of Closing Price (TRIX)	-	Emir, Dinçer e Timor (2012)
William's %R	$\frac{H_n - C_t}{H_n - L_n} \times 100$	Emir, Dinçer e Timor (2012); Kim (2003)
Working Capital	$\mathrm{WC} = \mathrm{TCA} - \mathrm{CL}$	Huerta, Corbacho e Elkan (2013)

Fonte: Elaborado pela autora.

Apêndice B – Manual de Uso do Economática

Os dados de cotações e indicadores financeiros utilizados nesta pesquisa foram coletados na base de dados do Economática. Este manual tem por objetivo mostrar sucintamente como esta coleta foi feita, a fim de contribuir para pesquisas futuras.

Ao iniciar o Economática, o usuário encontra uma lista de janelas que compõe o sistema, como por exemplo, cotações, indicadores financeiros, indicadores de mercado, demonstrativos financeiros, entre outros. Para acessar qualquer uma dessas bases, basta clicar sobre o nome desejado.

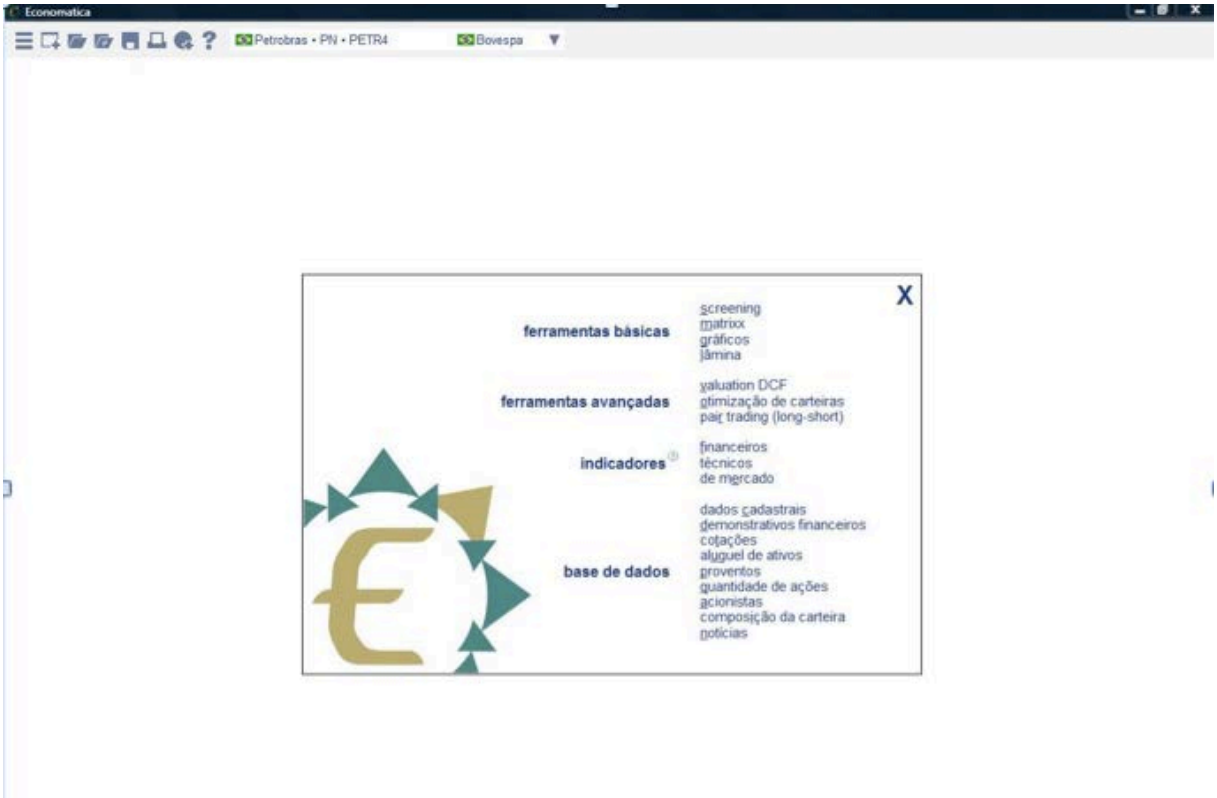


Figura 8 - Lista de janelas do Economática

Fonte: Elaborado pela autora.

Na parte superior da tela está o nome da empresa à qual as informações apresentadas nas janelas se referem. Para mudar a empresa, basta clicar neste campo para que uma janela de busca se abra. Nesta janela deve se colocar o nome da empresa ou código da ação desejada.

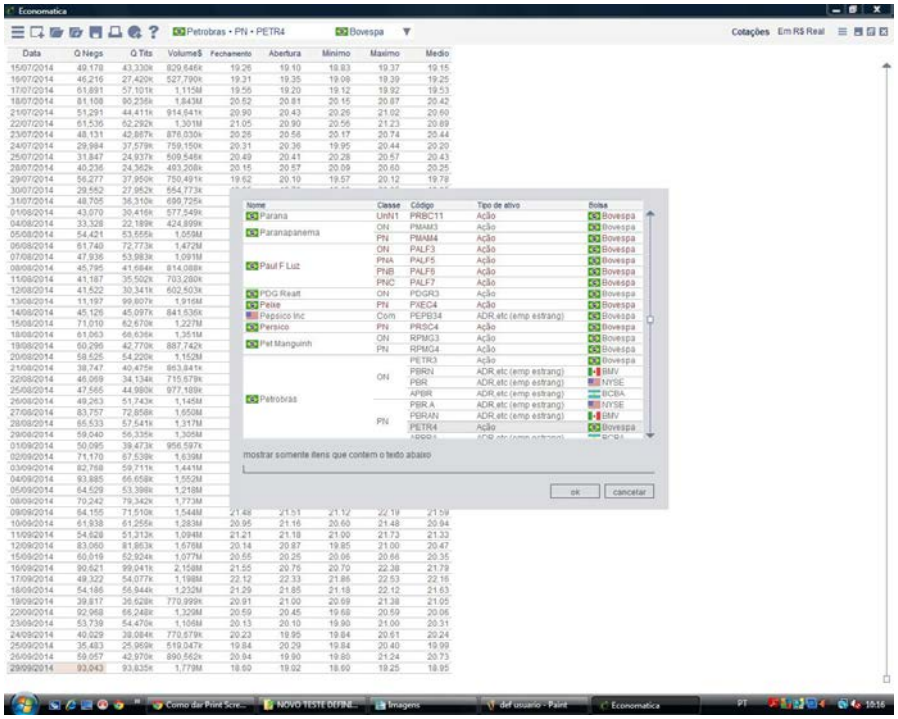
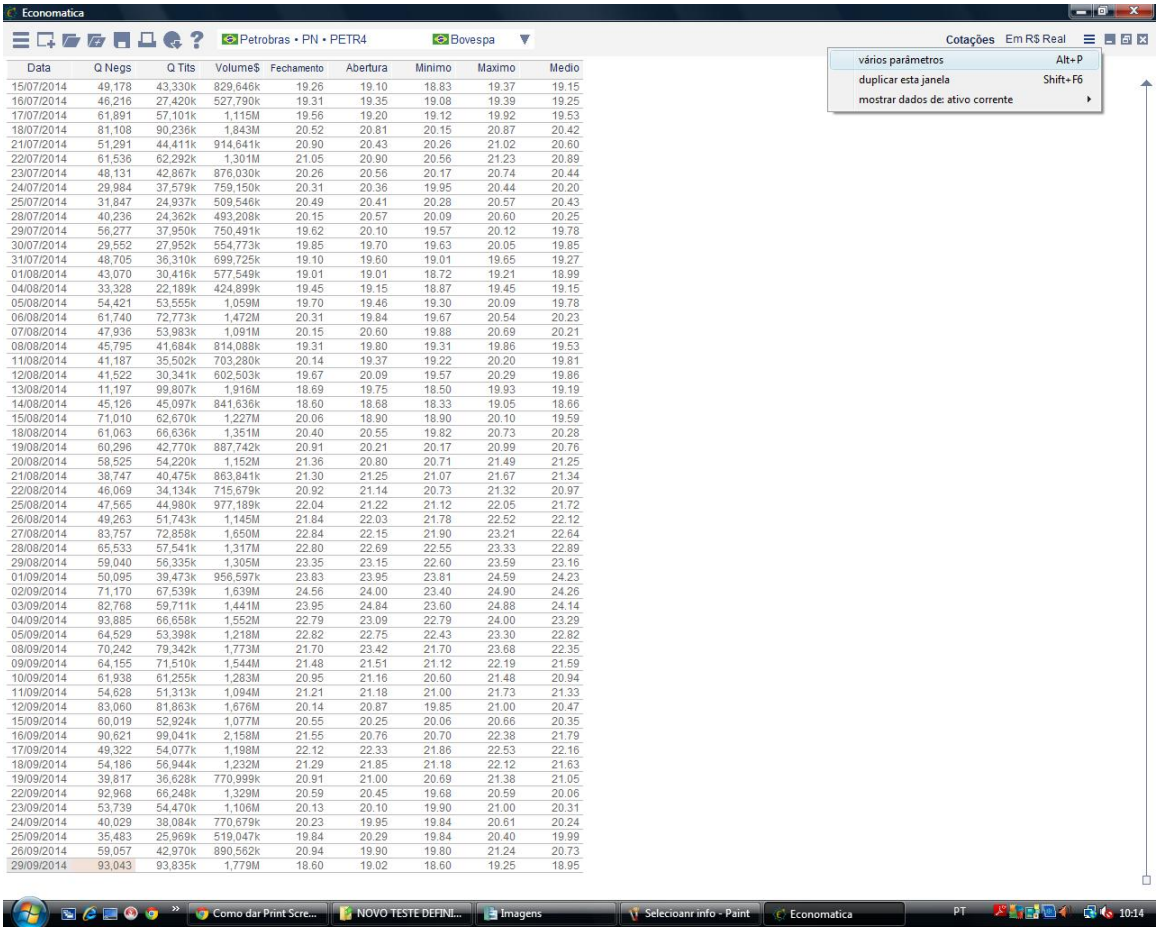


Figura 9 - Como buscar outras empresas ou ativos

Fonte: Elaborado pela autora.

Na janela de indicadores de mercado, é possível ajustar os dados de acordo com a necessidade do usuário. Para alterar qualquer parâmetro, basta clicar na opção “Vários Parâmetros”, localizada no campo superior direito.



Data	Q Negs	Q Tits	Volume$	Fechamento	Abertura	Minimo	Maximo	Medio
15/07/2014	49,178	43,330k	829,646k	19.26	19.10	18.83	19.37	19.15
16/07/2014	46,216	27,420k	527,790k	19.31	19.35	19.08	19.39	19.25
17/07/2014	61,891	57,101k	1,115M	19.56	19.20	19.12	19.92	19.53
18/07/2014	81,108	90,236k	1,843M	20.52	20.81	20.15	20.87	20.42
21/07/2014	51,291	44,411k	914,641k	20.90	20.43	20.26	21.02	20.60
22/07/2014	61,536	62,292k	1,301M	21.05	20.90	20.56	21.23	20.89
23/07/2014	48,131	42,867k	876,030k	20.26	20.56	20.17	20.74	20.44
24/07/2014	29,984	37,579k	759,150k	20.31	20.36	19.95	20.44	20.20
25/07/2014	31,847	24,937k	509,546k	20.49	20.41	20.28	20.57	20.43
28/07/2014	40,236	24,362k	493,208k	20.15	20.57	20.09	20.60	20.25
29/07/2014	56,277	37,950k	750,491k	19.62	20.10	19.57	20.12	19.78
30/07/2014	29,552	27,952k	554,773k	19.85	19.70	19.63	20.05	19.85
31/07/2014	48,705	36,310k	699,725k	19.10	19.60	19.01	19.65	19.27
01/08/2014	43,070	30,416k	577,549k	19.01	19.01	18.72	19.21	18.99
04/08/2014	33,328	22,189k	424,899k	19.45	19.15	18.87	19.45	19.15
05/08/2014	54,421	53,555k	1,059M	19.70	19.46	19.30	20.09	19.78
06/08/2014	61,740	72,773k	1,472M	20.31	19.84	19.67	20.54	20.23
07/08/2014	47,936	53,983k	1,091M	20.15	20.60	19.88	20.69	20.21
08/08/2014	45,795	41,684k	814,088k	19.31	19.80	19.31	19.86	19.53
11/08/2014	41,187	35,502k	703,280k	20.14	19.37	19.22	20.20	19.81
12/08/2014	41,522	30,341k	602,503k	19.67	20.09	19.57	20.29	19.86
13/08/2014	11,197	99,807k	1,916M	18.69	19.75	18.50	19.93	19.19
14/08/2014	45,126	45,097k	841,636k	18.60	18.68	18.33	19.05	18.66
15/08/2014	71,010	62,670k	1,227M	20.06	18.90	18.90	20.10	19.59
18/08/2014	61,063	66,636k	1,351M	20.40	20.55	19.82	20.73	20.28
19/08/2014	60,296	42,770k	887,742k	20.91	20.21	20.17	20.99	20.76
20/08/2014	58,525	54,220k	1,152M	21.36	20.80	20.71	21.49	21.25
21/08/2014	38,747	40,475k	863,841k	21.30	21.25	21.07	21.67	21.34
22/08/2014	46,069	34,134k	715,679k	20.92	21.14	20.73	21.32	20.97
25/08/2014	47,565	44,980k	977,189k	22.04	21.22	21.12	22.05	21.72
26/08/2014	49,263	51,743k	1,145M	21.84	22.03	21.78	22.52	22.12
27/08/2014	83,757	72,858k	1,650M	22.84	22.15	21.90	23.21	22.64
28/08/2014	65,533	57,541k	1,317M	22.80	22.69	22.55	23.33	22.89
29/08/2014	59,040	56,335k	1,305M	23.35	23.15	22.60	23.59	23.16
01/09/2014	50,095	39,473k	956,597k	23.83	23.95	23.81	24.59	24.23
02/09/2014	71,170	67,539k	1,639M	24.56	24.00	23.40	24.90	24.26
03/09/2014	82,768	59,711k	1,441M	23.95	24.84	23.60	24.88	24.14
04/09/2014	93,885	66,658k	1,552M	22.79	23.09	22.79	24.00	23.29
05/09/2014	64,529	53,398k	1,218M	22.82	22.75	22.43	23.30	22.82
08/09/2014	70,242	79,342k	1,773M	21.70	23.42	21.70	23.68	22.35
09/09/2014	64,155	71,510k	1,544M	21.48	21.51	21.12	22.19	21.59
10/09/2014	61,938	61,255k	1,283M	20.95	21.16	20.60	21.48	20.94
11/09/2014	54,628	51,313k	1,094M	21.21	21.18	21.00	21.73	21.33
12/09/2014	83,060	81,863k	1,676M	20.14	20.87	19.85	21.00	20.47
15/09/2014	60,019	52,924k	1,077M	20.55	20.25	20.06	20.66	20.35
16/09/2014	90,621	99,041k	2,158M	21.55	20.76	20.70	22.38	21.79
17/09/2014	49,322	54,077k	1,198M	22.12	22.33	21.86	22.53	22.16
18/09/2014	54,186	56,944k	1,232M	21.29	21.85	21.18	22.12	21.63
19/09/2014	39,817	36,628k	770,999k	20.91	21.00	20.69	21.38	21.05
22/09/2014	92,968	66,248k	1,329M	20.59	20.45	19.68	20.59	20.06
23/09/2014	53,739	54,470k	1,106M	20.13	20.10	19.90	21.00	20.31
24/09/2014	40,029	38,084k	770,679k	20.23	19.95	19.84	20.61	20.24
25/09/2014	35,483	25,969k	519,047k	19.84	20.29	19.84	20.40	19.99
26/09/2014	59,057	42,970k	890,562k	20.94	19.90	19.80	21.24	20.73
29/09/2014	93,043	93,835k	1,779M	18.60	19.02	18.60	19.25	18.95

Figura 10 - Acesso à janela de parâmetros

Fonte: Elaborado pela autora.

Nessa opção é possível mudar a escala das datas em que os dados são apresentados, ou seja, dias, semanas, meses, trimestres e anos. Além disso, os dados podem ser ajustados por proventos, isto é, splits e dividendos.

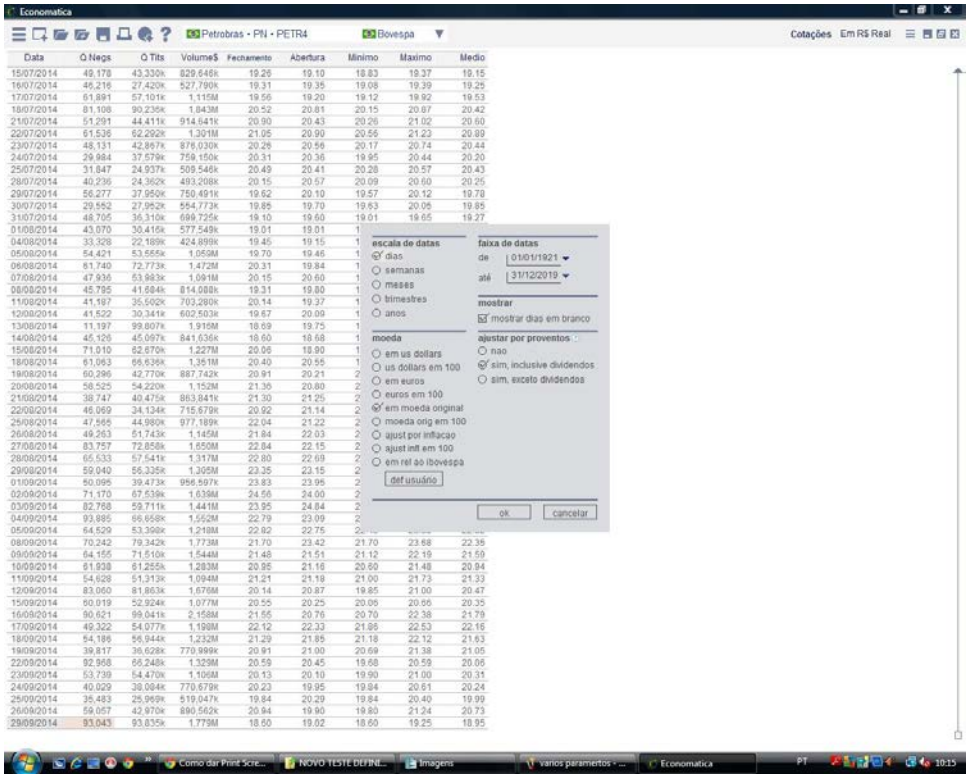


Figura 11 - Definição de parâmetros disponíveis

Fonte: Elaborado pela autora.

No caso específico da janela de cotações, os dados estão originalmente inseridos na moeda do país onde a bolsa está localizada, mas o sistema permite a conversão para outras moedas e também a atualização pela inflação. Na janela de "Vários Parâmetros", ao clicar em "Definido pelo Usuário", o usuário pode escolher o deflator que será utilizado.

Se a intenção for atualizar os valores antigos segundo um índice de inflação, a opção "Emmoeda original de dd/mm/aa" deveserselecionada. Para tal conversão, o Economática não apenas divide os valores antigos pelo índice da inflação, pois a simples divisão não é suficiente, uma vez que o resultado estaria expresso numa unidade sem significado. Então, o sistema também multiplica os valores da série, já divididos, pelo valor do índice de inflação da data para a qual se deseja atualizar os valores, data esta que deve ser informada no campo apropriado.

Por outro lado, se a opção for "Emmoeda original da última data disponível", a lógicaserá a mesma, diferindoapenas no fato de queosvaloresserãoconvertidossempreaopoderaquisitivoatual, isto é, data da informaçãomaisrecente da série do índice de inflaçãousado.

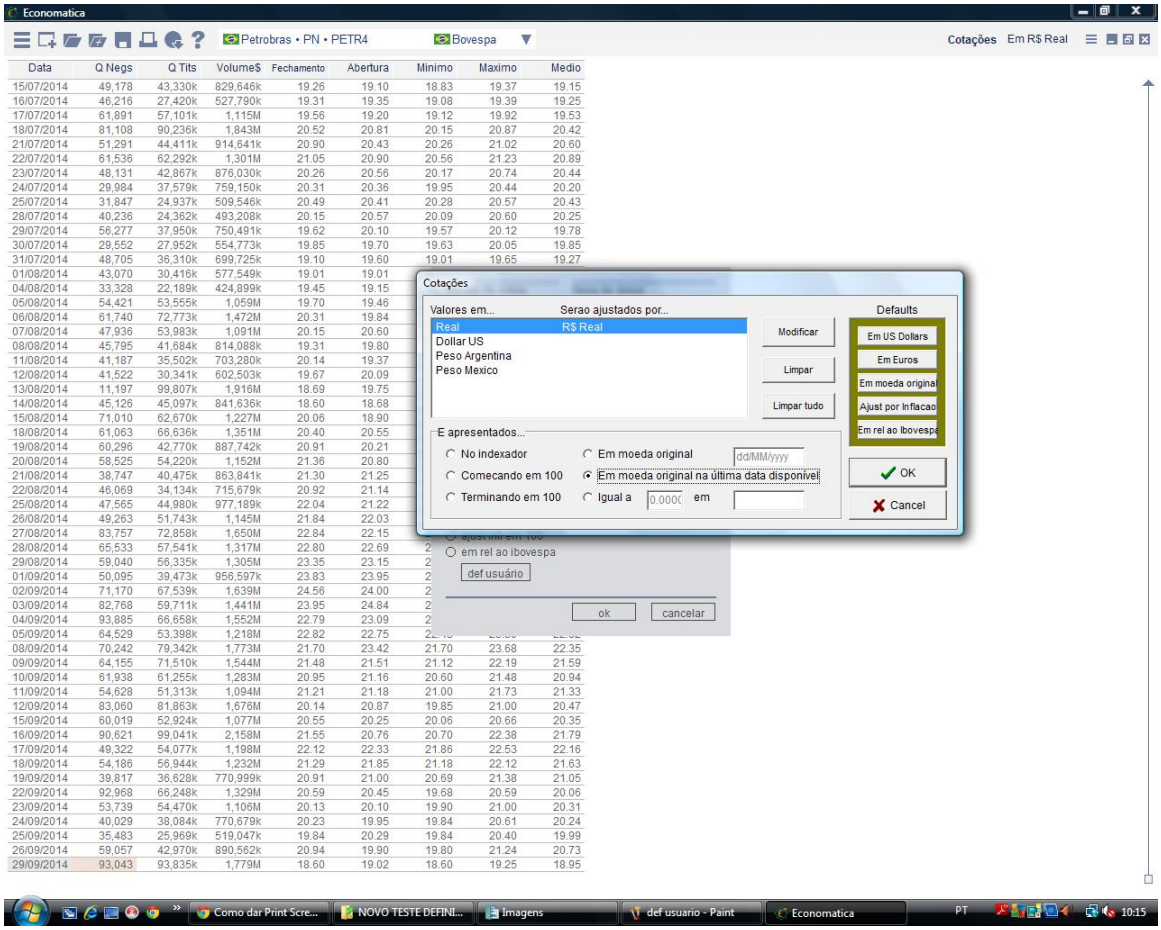


Figura 12 - Definição dos parâmetros pelo usuário

Fonte: Elaborado pela autora.

Depois de encontrar as informações desejadas, um ajuste no padrão numérico nos dados é aconselhável. Para que o separador decimal seja alterado de vírgula para ponto ou vice-versa, basta clicar no canto esquerdo do menu superior e selecionar “formato dos números”.

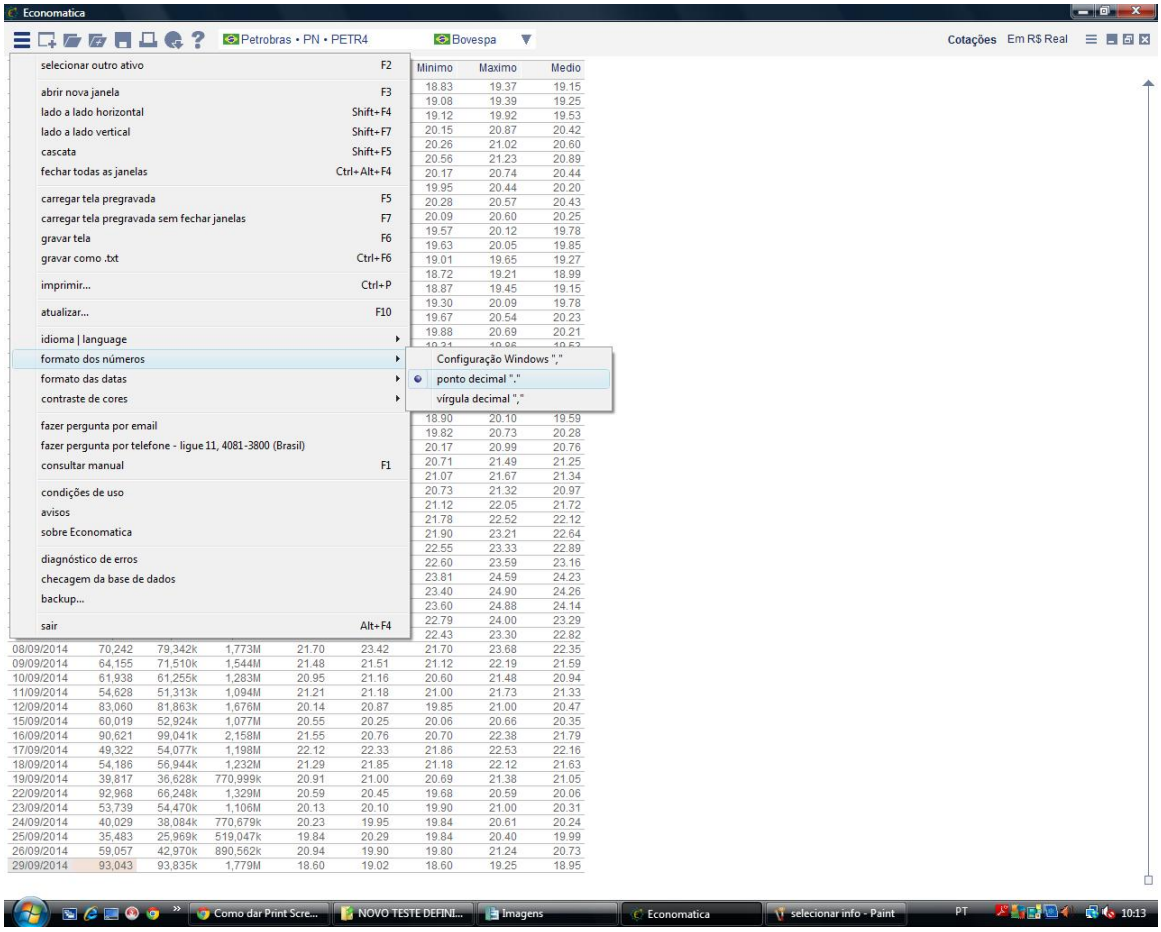


Figura 13 - Definição do formato dos números

Fonte: Elaborado pela autora.

Para coletar os dados de fato, as telas devem ser gravadas. Algumas janelas podem ser exportadas diretamente para o formato *excel*. Caso isso não seja possível, basta gravar a tela no formato *txt* e posteriormente transformá-las para os formatos *xls,xlsx* ou *csv*.

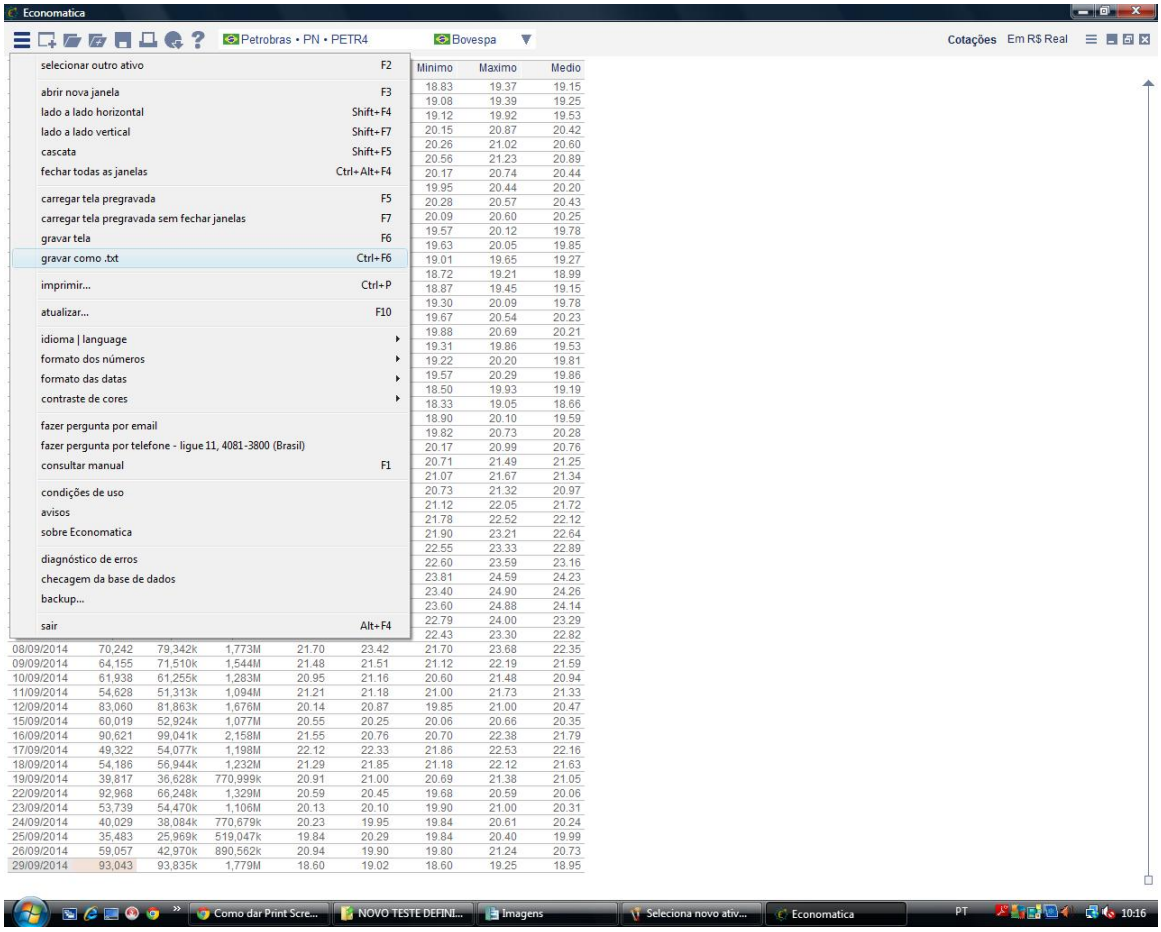


Figura 14 - Como gravar telas

Fonte: Elaborado pela autora.

Para mais informações sobre o sistema Economática, um manual oficial está disponível em http://economatica.com/support/manual/portugues/whnjs.htm.