

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA
MESTRADO EM ECONOMIA

BRUNO FERREIRA FRASCAROLI

UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS PARA
CLASSIFICAÇÃO DE *RATINGS* DE RISCO SOBERANO

João Pessoa - PB
Fevereiro 2006

BRUNO FERREIRA FRASCAROLI

UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA
CLASSIFICAÇÃO DE *RATINGS* DE RISCO SOBERANO

Dissertação apresentada ao
Programa de Pós-Graduação em
Economia da Universidade Federal
da Paraíba, em cumprimento às
exigências da obtenção do grau de
mestre em economia.

ORIENTADOR: PROF. DR. LUCIANO DA COSTA SILVA

João Pessoa - PB
Fevereiro 2006

BRUNO FERREIRA FRASCAROLI

UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS PARA
CLASSIFICAÇÃO DE *RATINGS* DE RISCO SOBERANO

Resultado: _____

BANCA EXAMINADORA

Prof. Dr. Luciano da Costa Silva
Orientador

Prof. Dr. José Antônio Rodrigues da Silva
Examinador

Prof. Dr. Mamadu Lamarana Bari
Examinador

João Pessoa - PB
Fevereiro 2006

*Ao pequeno núcleo duro:
meu Avô Ary, minha mãe
Leila e minha irmã Kelli.
Meu pai Romeu e minha
avó Ruth (in memorian).*

*Get up, stand up...
...don't give up the fight.*

AGRADECIMENTOS

Ao agradecermos às pessoas, sempre vamos nos esquecer de nomes, pois às vezes, somos traídos pela memória, porém, sabe o coração o verdadeiro valor de cada uma delas.

À minha família, meu avô Ary e meu tio Ary Júnior pelo apoio e incentivo para concluir o mestrado.

Ao professor Dr. Luciano da Costa Silva, inicialmente pelo convite, pelo incentivo, orientação e, principalmente, por saber me ouvir.

Aos professores membros da banca examinadora Dr. José Antônio Rodrigues da Silva e Prof. Dr. Mamadu Lamarana Bari que se dispuseram a contribuir no meu trabalho.

À professora Dr^a Rosemarie Bröker Bone, pelos trabalhos que foram meu fio condutor e por se dispor em tirar minhas dúvidas.

Aos professores Dr. Paulo Amilton Maia Leite Filho e Dr. Sinézio Fernandes Maia, da área de métodos quantitativos, por nos conduzirem neste desafio da vida acadêmica.

Aos demais professores do Programa em Pós Graduação em Economia CCSA/UFPA, não só pela busca em excelência de qualidade deste centro, mas acima de tudo, pelo carinho que me fizeram sentir em casa nestes dois anos distantes de minha família.

Aos funcionários Terezinha Polari e Rivonaldo por sempre nos atender prontamente, especialmente à Terezinha, tão querida, por participar em cada momento desta caminhada.

Aos colegas de turma que foram responsáveis pela criação de um ambiente tanto de estudo quanto de descontração.

Aos amigos que aqui fiz especialmente aos parceiros de guerra: Osvaldo, Vavá e Vitor.

À minha namorada Gislley, a pessoa mais especial de todas.

À CAPES pelo apoio financeiro.

SUMÁRIO

1 INTRODUÇÃO.....	11
2 LITERATURA.....	16
2.1 O papel dos <i>ratings</i> no mercado.....	18
2.2 Críticas sobre o papel dos <i>ratings</i>	19
2.3 Determinantes das atribuições de <i>ratings</i>	21
3 PROCEDIMENTOS METODOLÓGICOS.....	26
3.1 Introdução.....	26
3.2 Descrição das Redes Neurais.....	29
3.2.1 Modelo estocástico de um neurônio.....	34
3.2.2 Modelos de redes neurais estáticos e dinâmicos.....	34
3.3 Algoritmos de aprendizado.....	36
3.4 O <i>Perceptron</i>	38
3.4.1 Algoritmo de aprendizagem.....	40
3.5 <i>Multi-Layer Perceptron</i>	51
3.5.1 Teorema universal da aproximação.....	52
3.5.2 Regra de aprendizado do <i>Multi-Layer Perceptron</i>	52
3.5.3 Treinamento Levenberg-Marquardt.....	57
3.5.4 Treinamento <i>Resilient-Propagation</i> (RPROP).....	59
3.5.5 Regularização.....	61
4 SELEÇÃO E TRATAMENTO DA AMOSTRA DE DADOS.....	63
5 IMPLEMENTAÇÃO DA REDE NEURAL.....	65
5.1 Treinamento, teste e validação.....	65
5.2 Arquitetura da rede neural.....	67

5.2.1 Definição do número de camadas escondidas.....	68
5.2.2 Número de neurônios nas camadas escondidas.....	68
5.3 Treinamento da rede neural.....	69
6 ANÁLISE DE RESULTADOS.....	71
7 CONCLUSÃO.....	80
8 REFERÊNCIAS.....	83

LISTA DE TABELAS, QUADROS E FIGURAS

Tabela 1 – Sistemas de classificação de <i>ratings</i>	17
Tabela 2 – Simulações de <i>cenários</i> macroeconômicos.....	77
Quadro 1 – Contribuições no desenvolvimento de redes neurais.....	28
Quadro 2 – Exemplos de aplicações de redes neurais em economia e negócios...	29
Figura 1 – Neurônio artificial.....	30
Figura 2 – Rede neural com 2 camadas.....	31
Figura 3 – Dados linearmente separáveis pelo <i>Perceptron</i>	39
Figura 4 – Critério de definição de janelas <i>Walk-Forward</i>	66
Figura 5 – Performance do treinamento da rede neural escolhida.....	74
Figura 6 – Performance do teste da rede neural escolhida.....	74

RESUMO

O mercado financeiro é caracterizado pela presença de assimetria de informação. Assim, diante do desenvolvimento e maior integração destes, instituições e dispositivos que têm como propósito minimizar os transtornos gerados pela assimetria de informação são de muita importância. Especialmente no caso de países emergentes como o Brasil, que têm ampla necessidade de captação de recursos no mercado externo. O objeto de estudo do presente trabalho é estudar a classificação dos *ratings* de risco soberano realizadas pelas agências de *rating* e teve como propósito, a partir dos fundamentos macroeconômicos, encontrar uma classificação de risco das diversas nações utilizando o critério de risco estimado por Redes Neurais Artificiais. Vimos que quatro destes fundamentos estão mais diretamente conectados a estas atribuições. Logo após, utilizamos redes neurais para fazer simulações de *cenários* das condições externas de crédito para a economia brasileira a partir de mudanças nos fundamentos macroeconômicos com o objetivo de auxiliar os formuladores de políticas econômicas que têm como propósito melhorar estas condições de crédito no mercado internacional, a definir políticas com mais subsídio informacional.

Palavras Chave: *Rating* de risco soberano, fundamentos macroeconômicos, redes neurais artificiais.

ABSTRACT

The finance market is characterized by the presence of asymmetry of information. Then, the development and larger integration of these, institutions and devices that have as purpose to minimize the upset generated by the asymmetry of information have so much importance. Especially in the case of emerging countries like Brazil, that needs resources in the external market. The object of study of the present work is to study the classification of the ratings of sovereign risk and had as purpose, starting from the macroeconomics foundations, to find a classification of risk of the several nations using the criterion of Artificial Neural Networks. We saw that four of these foundations are more directly connected with these attributions. Therefore, we used Neural Networks to do simulations of sceneries of the external conditions of credit for the Brazilian economy starting from changes in the macroeconomics foundations with the objective of aiding the policy makers that have as purpose to improve these credit conditions in the international market, to define politics with more subsidy.

Key Words: Rating of sovereign risk, macroeconomics foundations, artificial neural networks

Classificação JEL: C45, G14, E44, D82

INTRODUÇÃO

Transações financeiras são inexoravelmente caracterizadas por assimetrias de informação entre fornecedores e tomadores de recursos. Os últimos têm necessariamente conhecimento mais amplo sobre sua própria capacidade de pagamento tanto a respeito da disposição de recursos e prazo de pagamento do que aqueles que lhes fornecem recursos. Portanto, do ponto de vista dos fornecedores de recursos, a presença de tal assimetria afetará os prêmios pelos riscos de crédito exigidos em qualquer operação de crédito e aquisição de títulos financeiros (CANUTO e SANTOS, 2003b).

Nos estudos sobre determinantes de crescimento e desenvolvimento econômico das nações é de suma importância estudar-se o mercado financeiro destas. O objeto de estudo do presente trabalho é estudar a classificação dos *ratings* de risco soberano realizadas pelas agências de *rating* e seu propósito é, a partir dos fundamentos macroeconômicos das diversas nações, encontrar uma classificação de risco utilizando Redes Neurais Artificiais. Esta classificação se baseou na localização de elementos constituintes e classificação *a posteriori* das mudanças de estado dos movimentos das séries temporais macroeconômicas determinantes das atribuições dos *ratings* de risco soberano desses países e das séries de risco soberano fornecidas pelas agências de *rating*. Logo após, utilizamos redes neurais para fazer simulações de *cenários* das condições externas de crédito para a economia brasileira a partir de mudanças em seus fundamentos macroeconômicos servindo, desta maneira, como uma ferramenta objetiva de apoio à decisão.

As redes neurais são aproximadores universais de funções. Elas são bastante similares às regressões lineares e não-lineares pelo método de mínimos quadrados ordinários (MQO), logo as

primeiras podem ser vistas como uma abordagem estatística alternativa para resolver problemas que envolvem análise recursiva de informações. Ambos métodos envolvem minimização da soma dos quadrados dos erros de estimação (KAASTRA e BOYD, 1996).

Como a economia brasileira necessita de fluxos de capitais externos para equilibrar seu balanço de pagamentos, é possível compreendermos como os fundamentos macroeconômicos e as atribuições de *ratings* afetam a economia restringindo ou ampliando sua captação de recursos no mercado financeiro internacional. Vale ressaltar que mesmo após a estabilização de algumas variáveis macroeconômicas a partir de 1994, a dificuldade de captação de recursos no mercado financeiro internacional provocou uma mudança na política cambial em janeiro de 1999.

Desta forma, torna-se claro a importância de estudos sobre as condições do mercado de crédito internacional haja vista que países emergentes como o Brasil, não são auto-suficientes em recursos financeiros tornando-se tomadores líquidos de recursos neste mercado. Reinhart (2001) salienta que, justamente para países emergentes tomadores líquidos de recursos, as condições de captação de recursos é mais crítica. Para estes países, as agências de *rating* possuem um comportamento mais severo, no sentido de, freqüentemente, penalizar estes países com *downgrades*¹ mesmo que seus fundamentos macroeconômicos estejam estáveis.

Um *rating* é uma estimativa da probabilidade de haver inadimplência futura. É importante entender que os *ratings* soberanos abordam o risco de crédito de governos nacionais (STANDARD & POOR'S, 2004). *Rating* é um ponto de vista sobre risco relativo baseado na capacidade e vontade do emissor para pagar completamente e no prazo acordado, principal e juros, durante o período de vigência do instrumento de dívida e severidade da perda, em caso de inadimplência (MOODY'S, 2003).

¹ Elevação da escala de risco soberano pelas agências de *ratings*.

Quanto maior o risco que investidores assumem em adquirir algum título de um governo soberano, menor a capacidade deste governo em tornar atraente esta aquisição e, portanto, atrair capital estrangeiro. Em consequência, maior é o prêmio remunerado aos investidores para compensá-los por assumir esse risco. Cantor e Packer (1996), confirmam que existe uma correlação inversa entre as atribuições de *ratings* realizadas pelas agências e o *spread* de prêmios pagos entre títulos dos governos soberanos e os títulos do governo norte-americano, isto é, quanto menor o conceito atribuído pela agência (maior risco) maior a remuneração paga pelo título, embora a habilidade dos *ratings* para explicar *spreads* relativos não pode ser completamente atribuída à correlação mútua com indicadores de risco soberanos.

Stiglitz e Grennwald (2004), justificam que o mercado de capitais possui características especiais pelo fato de negociar crédito, isto é, diferentemente do modelo de equilíbrio geral *Arrow-Debreu* onde os mercados possuem caráter anônimo², o mercado de capitais possui particularidades que o torna diferente dos demais. Segundo eles, existe informação assimétrica neste mercado. Este fenômeno causa problemas de risco moral e seleção adversa³. O risco moral surge pela incapacidade dos fornecedores de crédito, seja operacional ou por elevados custos, de monitorarem as ações dos tomadores de recursos após o contrato de concessão de crédito. A seleção adversa surge da informação assimétrica obtida pelos fornecedores de crédito a respeito dos projetos de investimentos dos tomadores de recursos que serão feitos com recursos emprestados.

Quando é possível diminuir o efeito maléfico das assimetrias de informação no sistema financeiro, as transações financeiras se desenvolvem. Assim, algumas medidas como coleta e processamento de informações antes das operações, firmação de contratos mais claros e monitoramento dos tomadores de recursos de modo a controlar o uso dos recursos após seu

² Anônimo no sentido de que se podem abstrair as mesmas características de todos os mercados que compõem o modelo.

³ Para uma discussão mais detalhada sobre risco moral e seleção adversa em Mas-Colell, Whinston e Green (1995).

fornecimento agem no sentido de desenvolver o sistema financeiro. Entretanto, tais medidas operam com custos que, às vezes, oneram demasiadamente as partes que negociam recursos e nem sempre essas medidas têm eficácia suficiente para suprimir essa assimetria de informação.

Neste sentido, a presença de instrumentos de ordem legal, assim como instrumentos institucionais que sustentem o cumprimento de contratos e o exercício de garantias, a assimetria de informações e os prêmios cobrados como compensação pelos riscos de crédito diminuem.

Em conformidade com Canuto e Santos (2003b), um risco em particular é o risco soberano, isto é, o risco de crédito vinculado a operações de crédito concedidas a estados soberanos. Os colaterais e o cumprimento dos contratos são distintos daqueles presentes nos casos de crédito para agentes privados. Adicionalmente, a disponibilidade da capacidade de pagamento é diferente, pois entre seus determinantes estão as variáveis macroeconômicas que indicam as expectativas a cerca da capacidade futura de pagamento daquele país que toma recursos emprestados. Entre estas, podemos destacar as reservas de divisas estrangeiras, crescimento econômico, capacidade de arrecadação tributária, etc.

Em complemento, este trabalho faz uma breve análise acerca dos principais modelos de redes neurais que foram utilizados durante as estimações experimentais. Encontramos, treinamos através de aprendizagem supervisionada e testamos uma arquitetura de redes neurais para analisar as séries temporais macroeconômicas de um conjunto de países e, a partir delas, a rede neural pôde aprender como funciona a atribuição de *ratings* de risco soberano para estes países; e, a partir da arquitetura encontrada nos experimentos, fizemos simulações sobre as condições de crédito para a economia brasileira a partir de mudanças dos comportamentos dos determinantes macroeconômicos das atribuições de *rating*.

O presente trabalho está organizado da seguinte forma. Além desta introdução, no capítulo 2, discutimos a problemática envolvendo a questão dos *ratings*, das agências e dos impactos dos *ratings* no lado real da economia. No terceiro capítulo, abordamos os procedimentos metodológicos do trabalho, aspectos históricos das redes neurais, suas aplicações no campo da economia e finanças, detalhando seus principais aspectos técnicos como, por exemplo, funções de transferência, algoritmos de aprendizagem e tipos de arquiteturas das redes neurais. Logo após, dedicamos o capítulo 4 para apresentarmos as variáveis que fazem parte da entrada da rede neural e como estas foram previamente trabalhadas para serem introduzidas no modelo. No capítulo 5, mostramos como funciona a implementação de uma rede neural, buscando enumerar passos básicos do procedimento. Após este capítulo, segue-se o capítulo 6, no qual analisamos os resultados encontrados no trabalho. Finalizando, temos o capítulo 7 conclusivo em que discutimos a ligação da problemática econômica tratada com os resultados empíricos fornecidos pela rede neural.

2 LITERATURA

O *rating* de risco soberano é um indicador que busca expressar o risco ao qual investidores estrangeiros estão submetidos ao adquirirem títulos de algum país. Os *ratings* de risco soberano são construídos com base em análises das conjunturas econômica, social e política dos países e, por este motivo, podem ser subjetivos porque envolvem julgamento não só das variáveis macroeconômicas internas e externas no presente, mas também das perspectivas das mesmas para o futuro.

As agências de *rating* são empresas independentes de qualquer interesse, quer por parte de governos ou de empresas privadas. Essa característica lhes permite ter como princípios: independência, objetividade, credibilidade e liberdade de divulgação de avaliações com relação à qualidade de crédito dos emitentes e emissões de dívida (*STANDARD & POOR'S*, 2002a). Dentre as principais agências de *rating* de risco temos, por exemplo, *Standard & Poor's*, *Moody's Investor Service* e *Fitch IBCA*⁴.

A presença de agências de *rating* como fornecedoras de mercadoria, neste caso, o processamento de informações, constitui peça-chave para que a oferta de recursos financeiros não fique restrita às instituições financeiras domésticas, permitindo tanto para o governo central destas economias quanto às empresas domésticas, ampliarem suas possibilidades de tomarem recursos.

Os *ratings* normalmente tentam refletir os fatores específicos de risco de um país, os quais podem trazer conseqüências sobre a capacidade de uma entidade para saldar integral e pontualmente sua dívida. Fazem parte destes fatores de riscos o risco de intervenção soberana, como, por exemplo, o risco de um país impor políticas cambiais ou decretar moratória da dívida. Além disso, inclui os riscos econômico ou país, que se referem ao ambiente econômico,

⁴ Segundo Bone (2005), as agências *Standard & Poor's* e *Moody's Investor Service* detêm juntas cerca de 80% do mercado de classificação de *ratings*.

político e social, o qual influencia tanto os *ratings* do próprio país quanto aquelas entidades emissoras ali presentes (*STANDARD & POOR'S*, 2001a).

A Tabela 1 traz as escalas das classificações feitas pelas principais agências de *rating* do mundo, onde temos uma escala numérica crescente de risco que vai de 1 a 24. Países que se encontram na escala de 1 a 10 pertencem ao grupo de países de menor risco e, países que se encontram na escala acima de 10, fazem parte de países de maior risco, o que configura um comportamento especulativo no mercado financeiro. Países que estão na ordem da escala de 22 a 24 são países que declararam moratória de suas dívidas. Logo, as classificações de obrigações nesta categoria são baseadas na possibilidade de recuperação parcial ou total do empréstimo.

Tabela 1 – Sistemas de classificação de *ratings*

<i>Posição</i>	<i>Empresas</i>			<i>Escala Numérica</i>
	<i>Moody's</i>	<i>S & P</i>	<i>Fitch</i>	
menor risco	Aaa	AAA	AAA	1
	Aa1	AA+	AA+	2
	Aa2	AA	AA	3
	Aa3	AA-	AA-	4
	A1	A+	A+	5
	A2	A	A	6
	A3	A-	A-	7
	Baa1	BBB+	BBB+	8
	Baa2	BBB	BBB	9
	Baa3	BBB-	BBB-	10
especulativa	Ba1	BB+	BB+	11
	Ba2	BB	BB	12
	Ba3	BB-	BB-	13
	B1	B+	B+	14
	B2	B	B	15
	B3	B-	B-	16
	Caa1	CCC+	CCC+	17
	Caa2	CCC	CCC	18
	Caa3	CCC-	CCC-	19
	-	CC	CC	20
	-	C	C	21
	Ca	SD	DDD	22
	C	D	DD	23
	-	-	D	24

Fonte: *Standard & Poor's, Fitch Ratings e Moody's*

2.1 O papel dos *ratings* no mercado

O *rating* de risco soberano exerce impacto direto ou indireto sobre a qualidade de crédito das entidades que operam em uma determinada localização de diversas maneiras. Cantor e Packer (1996), ressaltam que os *ratings* são importantes não apenas porque alguns dos maiores emissores de dívidas são governos soberanos, mas também porque de acordo com as atribuições de *ratings*, a captação de recursos por governos estaduais, municipais ou empresas privadas, por exemplo, localizados nestes países é afetada.

Na presença de informações assimétricas nos mercados financeiros mundiais, as classificações realizadas pelas agências de *rating* do risco de crédito são sinalizações para os agentes participantes do mercado financeiro.

Bone (2004) resume os motivos pelos quais, o mercado sustenta a existência das agências de *rating*:

- O *rating* deve possuir idoneidade, isto é, ele vale apenas quando é confiável. Dado que uma agência de *rating* se financia cobrando o custo de elaboração do *rating* dos emitentes de dívidas, ela não possui qualquer incentivo de produzir *ratings* viesados ou não-confiáveis;
- Um outro argumento importante do mercado para o *rating* se deve aos custos de captar e interpretar informações sobre os emitentes de dívidas. Assim, as agências de *rating* possuem retorno de escala na captação e interpretação dessas informações;
- O *rating* tem a capacidade de sintetizar as informações sobre a futura qualidade de crédito do emitente de dívidas de uma forma clara e simples para o mercado.

2.2 Críticas sobre o papel dos *ratings*

Numerosos estudos mostram que há um atraso nas mudanças nas atribuições dos *ratings* frente às mudanças ocorridas nos mercados, e estes se antecipam às mudanças nas atribuições dos *ratings* (PARTNOY, 2002). Assim, a eficácia do objetivo do *rating* seria gravemente prejudicada, pois se as agências de *rating* se propõem a avaliar a condição futura de pagamento de um emitente de dívida, e esta é realizada após a mudança de percepção do mercado, os agentes não mais se apoiariam no *rating* como uma informação fidedigna de tal condição futura de pagamento de um emitente de dívida.

Além disso, Cantor e Packer (1996), observam que há uma forte convergência das atribuições dos *ratings* realizados pelas agências *Standard & Poor's* e *Moody's*, de modo que, quando estas agências atribuem *ratings* diferentes, esta diferença frequentemente é de apenas uma posição na escala de *ratings*, o que configura uma tendência conservadora destas atribuições.

Segundo Partnoy (2002), existe um paradoxo no mercado financeiro envolvendo os *ratings* atribuídos pelas agências. Embora haja cada vez menos informações contidas nos *ratings* atribuídos pelas agências, estas parecem estar em contínua prosperidade. O autor aponta para a existência de ineficiências de ordem legal existentes nos mercados financeiros que explicam este paradoxo. Numerosas regras legais e regulamentações dependem substancialmente dos *ratings* e, particularmente, daqueles atribuídos por uma pequena parte das agências de *rating* que fazem parte da *Nationally Recognized Statistical Ratings Organizations (NRSROs)*. Contudo, as barreiras para entrar no mercado de *NRSRO* são proibitivas.

Logo, os *ratings* atribuídos pelas agências de *rating* que fazem parte da *NRSRO* são valiosos para os participantes do mercado financeiro mesmo que o conteúdo informacional deles não seja maior que a informação pública neste contida. Agentes que administram recursos de investidores institucionais, por exemplo, estão submetidos às regulamentações internas da administração dos fundos que atribuem o risco existente na construção das carteiras com base nos *ratings*. Assim, mesmo que os *ratings* não forneçam uma informação adequada sobre os riscos de determinados emittentes, os administradores desses fundos são obrigados a utilizá-los. O trabalho de Partnoy (Op. Cit.) demonstra que o aumento da utilização dos *ratings* como instrumento de regulamentação, após 1973, mudou a natureza do produto vendido pelas agências de *rating*, pois os *ratings* ao invés de fornecerem informação, determinam a conduta dos agentes no mercado financeiro.

As agências de *rating* fracassaram em antecipar o declínio rápido nos preços de centenas de papéis financeiros na crise de 1929 e, em consequência deste fato, numerosos *ratings* foram abruptamente baixados, o que pela primeira vez, demonstra uma mudança generalizada na atribuição de *ratings* pelas agências, tornando-se mais conservadoras (PARTNOY, Op. Cit.).

Mais recentemente, de acordo com Ferri, Liu e Stiglitz (1999), diversos observadores e instituições do mercado financeiro como, por exemplo, *World Bank* e *IMF*, apontaram para a falha das agências de *rating* em alertar antecipadamente o mercado sobre a crise asiática de 1997. Parecido com o que houve na crise de 1929, as agências de *rating* baixaram os *ratings* de risco soberano de alguns países envolvidos na crise configurando novamente um comportamento excessivamente conservador em vista do que os fundamentos macroeconômicos justificariam.

Segundo Ferri, Liu e Stiglitz (Op. Cit.) em reconhecimento a esta falha, uma destas agências, a *Fitch IBCA*, em comentário público em janeiro de 1998, reconhece não ter dado a devida atenção às dívidas de curto prazo em moeda estrangeira. Após a crise asiática, a *Moody's*, em declaração, também enfatiza a importância das dívidas de curto prazo em moeda estrangeira.

Bone (2002) afirma que antes da crise na Ásia em 1997 as mudanças nos *ratings* soberanos não eram significativas e se mostravam historicamente estáveis. Após a crise asiática, mais precisamente de julho de 1997 a novembro de 1998, foram observados *downgrades* abruptos não comparáveis aos que ocorriam anteriormente.

O trabalho de Ferri, Liu e Stiglitz (1999), postula a existência de uma regra procíclica das agências de *rating* o que prolongaria os efeitos de crises sobre as economias através de comportamentos excessivamente conservadores dessas agências na medida em que elas baixam excessivamente os *ratings* de risco soberano de economias que já possuem baixos *ratings* de risco soberano na escala. Ao passo que, segundo os autores, em economias com *rating* de risco soberanos mais elevados isso não acontece, pelo contrário, essas economias experimentam atribuições crescentes de *rating*.

Ferri, Liu e Stiglitz (Op. Cit.), argumentam que após as crises, as agências de *rating* se tornam mais conservadoras porque são pegas de surpresa e têm mais dificuldade de reconstruir suas reputações. Assim, como a reputação dessas agências é prejudicada também o retorno de seus serviços também é prejudicado. Logo, elas são incentivadas a se tornarem mais conservadoras para não falharem novamente.

Contudo, os *ratings* atribuídos pelas agências não devem ser atribuídos como um remédio às crises financeiras.

2.3 Determinantes das atribuições de *ratings*

Os *ratings* de risco soberano atribuídos pelas agências de *rating* para as diversas nações do mundo usam critérios que nem sempre são claros, ou seja, não há um padrão de quais são os determinantes para as atribuições destes *ratings* e nem das ponderações utilizadas para cada

determinante, pois estes *ratings* são compostos de uma combinação de determinantes qualitativos e quantitativos.

The agencies never describe their terms or analysis precisely or say, for example, that a particular rating has a particular probability of default, and they stress that the ratings are qualitative and judgmental. This secretive, qualitative process is not the type of process one would expect if the agencies had survived based on their ability to accumulate reputational capital. On the other hand, such processes make it more likely that an agency would be able to survive in a non-competitive market; if the rating process had been public or quantitative (rather than qualitative), other market entrants easily could have duplicated the rating agencies technology and methodology (PARTNOY, 2002).

Dentre os determinantes qualitativos de atribuição dos *ratings* temos determinantes de riscos políticos, como, por exemplo, a probidade das lideranças, estabilidade e transparência das instituições, que fazem parte do julgamento dos analistas sobre o comportamento destes governos frente às decisões em momentos de crises econômicas. Haque, Mark e Mathieson (1998), confirmaram empiricamente que variáveis políticas não têm nenhum ou quase nenhum impacto sobre a decisão de atribuição de *ratings* por parte das agências. Uma possível explicação apontada por eles é a de que as variáveis econômicas absorvem instantaneamente flutuações originadas no âmbito político. Diante disso, no presente trabalho utilizamos fundamentos macroeconômicos para buscarmos explicar a decisão de atribuição de *ratings* pelas agências.

Os determinantes quantitativos incluem diversas medidas de desempenho econômico e financeiro, bem como as obrigações contingenciais, embora o julgamento sobre a integridade da amostra desses dados seja uma questão mais qualitativa. Não há uma fórmula exata para se combinar os escores ao se determinar os *ratings*. As variáveis analíticas são inter-relacionadas e os pesos não são fixos, nem entre os diversos governos soberanos nem ao longo do tempo (STANDARD & POOR'S, 2004).

Cantor e Packer (1996), confirmam que, mesmo para os determinantes quantitativos, é difícil encontrar uma relação entre os pesos atribuídos aos determinantes pela *Moody's* e pela *Standard & Poor's*, por causa do grande número de critérios adotados por estas agências.

A *Standard & Poor's* divide a estrutura dos determinantes dos *ratings* soberanos em categorias, estas categorias se dividem em perspectivas de crescimento econômico, flexibilidade fiscal, estabilidade monetária e risco político, dentre outras categorias. Para atribuir *ratings* de crédito, cada governo soberano é classificado dentro de uma escala – cujas notas vão de um (a mais alta) até seis – aplicada a cada uma das categorias analíticas (*STANDARD & POOR'S*, 2004).

Algumas agências, dentre elas a *Moody's*, incluem entre os determinantes das atribuições de *ratings*, a probabilidade de *default*, isto é, a probabilidade de um governo soberano declarar moratória de sua dívida.

Dentre os determinantes, podemos enumerar alguns, na mesma linha de trabalhos como, por exemplo, Cantor e Packer (1996) e Megale (2003), Canuto, Santos e Porto (2004), Rowland (2004), que serão utilizados como entradas no processo de aprendizagem supervisionada para criação de um critério de risco soberano. Podemos dividir estes determinantes em variáveis utilizando o mesmo critério de divisão adotado por Rowland (2004), apenas adicionando mais uma categoria de variáveis determinantes do desenvolvimento e estabilidade econômica.

Variáveis de Solvência

- Taxa de crescimento das exportações de bens e serviços;
- Taxa de crescimento das importações de bens e serviços;
- Dívida de longo prazo como percentual do produto.

Variáveis de Liquidez

- Nível de reservas totais como percentual do produto;
- Dívida externa total como percentual das exportações de bens e serviços;
- Serviço da dívida total como percentual das exportações de bens e serviços;
- Balança externa de bens e serviços como percentual do produto;

Variáveis de desenvolvimento e estabilidade econômica

- Taxa de crescimento do produto *per capita*;
- Evolução do nível de preços ao consumidor;
- Grau de abertura econômica medidas pela soma do volume das exportações mais importações como percentual do produto.

O modelo econométrico de Cantor e Packer (1996) é seminal, no sentido de ajustar um bom modelo de previsão para as atribuições de *ratings* através de séries temporais de amostra de dados macroeconômicos. Porém, após a crise asiática ocorrida em 1997, o modelo econométrico destes autores não mais consegue uma boa capacidade preditiva das atribuições de *ratings*. Adicionalmente, eles chamam a atenção para a limitação de modelos quantitativos na incapacidade de explicar todos os movimentos das atribuições dos *ratings*, haja vista que estes tipos de modelos têm dificuldade de inserir entre suas variáveis explicativas, variáveis qualitativas.

O trabalho de Ferri, Liu e Stiglitz (2002), chama a atenção para o que seria o modelo utilizado pelas agências na atribuição de *ratings* após a crise asiática de 1997. Este modelo principal é dividido em dois modelos . Um modelo quantitativo em função dos determinantes de ordem macroeconômica. E um modelo de informações *ad hoc* de cada país, refletindo o julgamento das agências. O modelo pode ser expresso da seguinte maneira:

$$\begin{aligned} Ratings &= \omega_q Ratings_q + \omega_j Ratings_j \\ \omega_q + \omega_j &= 1 \end{aligned}$$

Onde

ω_q é o peso do julgamento feito pelas agências dos determinantes quantitativos;

ω_j é o peso do julgamento feito pelas agências dos determinantes qualitativos;

Temos inúmeras abordagens feitas através de modelos quantitativos ao problema de atribuição de *ratings*. Dentre os trabalhos poderíamos citar o trabalho de Rowland (2004) que utiliza a técnica de análise coletiva de dados (*pooled data*). Canuto e Santos (2004), utilizam um modelo econométrico de análise coletiva de dados em *cross-section*. Na linha do modelo econométrico de Cantor e Packer (1996), Ferri, Liu e Stiglitz (2002) estimam um modelo MQO com dois tipos de cardinalização para os *ratings*. Leichsenring (2004) estima um modelo de vetores autoregressivos (VAR) para avaliar a relação entre os *ratings* e a taxa de juros doméstica. Bone (2004) utiliza testes de estabilidade de *Chow* para medir a estabilidade das séries temporais de *ratings* soberanos emitidos pelas diversas agências.

Entre os modelos econométricos, podemos citar o problema da cardinalização. Este problema surge da tentativa de cardinalizar os *ratings* para torná-los estimáveis, pois modelos econométricos são incapazes de lidar com conceitos, assim não há como introduzirmos como variável num modelo econométrico um conceito de *rating*. Contudo, ao cardinalizar um conceito de *rating*, podemos incorrer em erros, pois cardinalizações, sejam elas lineares ou não-lineares, estabelecem uma relação quantificável entre os conceitos de *ratings*. Assim, uma atribuição de *rating* igual a dois, por exemplo, não quer dizer necessariamente que possua o dobro de risco de uma atribuição de *rating* igual a um.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Introdução

O processo de estimação utilizado no presente trabalho foi o processo chamado de Redes Neurais Artificiais. Esta metodologia é inspirada nos princípios do funcionamento do cérebro humano para resolução de determinados tipos de problemas, criando a chamada inteligência artificial. Este processo de estimação é muito útil diante de problemas de reconhecimento de padrões e, portanto, de acordo com os objetivos estabelecidos, ele se torna o estimador mais apropriado para nossa análise.

A metodologia de redes neurais possui múltiplas aplicações tanto no ambiente acadêmico como no mercado, resolvendo problemas de reconhecimento de padrões como caracteres, imagens, voz, etc. São utilizadas em sistemas de auxílio a diagnósticos como em medicina, falhas mecânicas, etc. São amplamente usadas nas áreas de robótica inteligente e sistemas de controle, além de serem também bastante utilizadas para previsões, seja previsão de séries temporais em finanças ou previsões de fenômenos naturais.

A utilização de redes neurais, assim como a utilização de outros métodos de estimação, possui vantagens e desvantagens. Entre as vantagens, podemos apontar:

- Flexibilidade na modelagem: A hipótese de linearidade não é necessária. Vários modelos não lineares podem ser construídos a partir da mesma técnica.
- Flexibilidade estatística: Nenhuma hipótese *a priori* é feita sobre as distribuições de probabilidade envolvidas.
- Categorização: As variáveis dependentes podem ser categorias qualitativas, em vez de apenas numéricas.

- Incorporação do conhecimento subjetivo: A categorização permite que o treinamento (estimação) da rede neural seja feito diretamente a partir de julgamentos subjetivos, como no reconhecimento de imagens e no diagnóstico médico.

Entre suas desvantagens podemos destacar:

- A rede neural pode apresentar dificuldade de configuração em relação à sua estrutura inicial e também no que se refere aos parâmetros dos algoritmos de aprendizado.
- Pode haver dificuldade em apresentar os resultados obtidos na forma da análise estatística tradicional, como intervalos de confiança, estímulo-resposta e relações de causalidade.
- Podem existir dificuldades de convergência e instabilidade inerentes aos algoritmos de otimização empregados;
- Pode haver lentidão do processo de aprendizado.

O Quadro 1 resume algumas das mais importantes contribuições no desenvolvimento de redes neurais.

Quadro 1 - Contribuições no desenvolvimento de redes neurais

<i>Autores</i>	<i>Contribuições</i>
McCulloch e Pitts (1943)	<i>Perceptron</i> , o primeiro neurônio artificial.
Hebb (1949)	Escreve um livro intitulado “ <i>The Organization of Behavior</i> ”, onde trata das regras de aprendizado.
Rosenblatt (1962)	Publica <i>Principles of Neurodynamics</i> .
Widrow (1962)	Cria o <i>Adaline (Adaptative Linear Element)</i> e o <i>Madaline (Multiple Adaptative Linear Element)</i> .
Minsky e Papert (1969)	Escrevem um livro intitulado <i>Perceptrons: An Introduction to Computational Geometry</i> , onde estudam as limitações do <i>Perceptron</i> .
Hopfield (1982)	Função de energia para redes recorrentes.
Rumelhart, Hinton e Williams (1986)	Criaram o algoritmo de retropropagação ou <i>Backpropagation</i> .
Carpenter e Grossberg (1987a)	Classificador de Carpenter-Grossberg.
Broomhead e Lowe (1988)	<i>Radial Basis Function</i>

Fonte: Elaborado pelo autor.

Dentre as aplicações, as redes neurais também têm sido utilizadas em economia e finanças, especialmente na previsão de séries temporais e na classificação de agentes econômicos. No Quadro 2, destacamos algumas destas aplicações que se direcionaram a estudar problemas econômicos e financeiros.

Quadro 2 - Exemplos de aplicações de redes neurais em economia e negócios

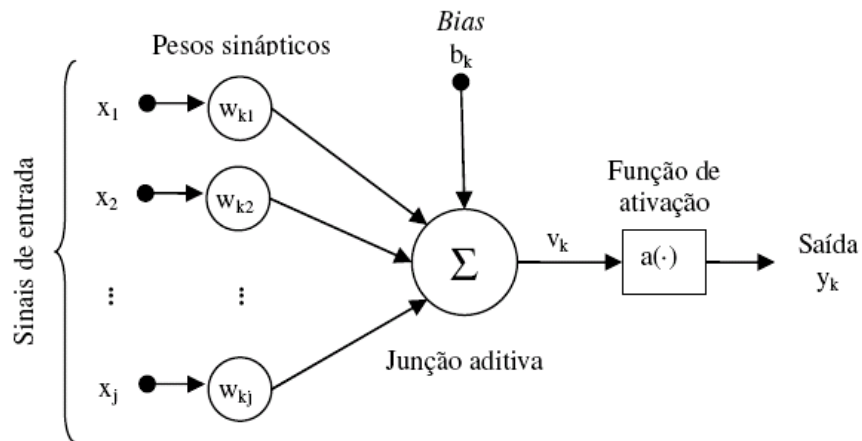
<i>Assunto</i>	<i>Estudos</i>
<i>Ciclos de Negócios</i>	Qi (2001)
<i>Previsões Financeiras</i>	Kohzadi et al. (1996), Yao, Li e Tan (2000), Braga (2004), Gately (1996)
<i>Despesas, Escolhas do consumidor e Problemas Microeconômicos</i>	Agrawal and Schorling (1996), West, Brockett e Golden (1997), Aiken and Bsat (1999), Wang (1999), Jiang, Zhong e Klein (2000), Vellido, Lisboa e Vaughan (1999), Passari (2003), Church e Curram (1996), Brondino (1999), Castro Júnior (2003)
<i>Crescimento do Produto, Inflação, Taxa de Câmbio e Problemas Macroeconômicos</i>	Tkacz (2001), Chen, Racine e Swanson, (2001a), Tseng, Yu e Tzeng (2001), Benite (2003), Zhang e Hu (1998), Leung, Chen e Daouk (2000a), Nag e Mitra (2002)

Fonte: Elaborado pelo autor.

3.2 Descrição das Redes Neurais

Redes neurais são aproximadores universais de funções, como os estimadores de regressão linear e as séries de Fourier. Seu princípio é inspirado nas redes de conexões de neurônios existentes no cérebro dos animais vertebrados. Sua unidade fundamental é o *neurônio*, cuja versão artificial é apresentada na Figura 1.

Figura 1 - Neurônio artificial



Fonte: Haykin (2001)

$\varphi(\cdot)$ é a função de transferência;

y_k é o sinal de saída do neurônio k .

Na figura, verificamos 3 elementos básicos:

- 1) Um conjunto de sinapses, caracterizadas por pesos w_{kj} , sendo k o índice utilizado para designar o neurônio e j o índice da sinapse. Um sinal x_j na entrada da sinapse j é multiplicado pelo peso sináptico w_{kj} .
- 2) Um somador, chamado na Figura 1 de *junção aditiva*, que soma os sinais de entrada, ponderados pelos respectivos w_{kj} .
- 3) Uma função de transferência para restringir a amplitude da saída do neurônio.

O *bias* b_k , é introduzido para melhorar o ajuste da função aproximada. Normalmente, o *bias*, tem um valor positivo e seu propósito é análogo ao intercepto nas regressões dos modelos econométricos.

Assim, podemos descrever um neurônio matematicamente da seguinte forma:

$$v_k = \sum_{j=1}^m w_{kj} x_j + b_k \quad (1)$$

$$y_k = \varphi(v_k) \quad (2)$$

Onde:

$x_1, x_2, x_3, \dots, x_m$ são sinais de entrada;

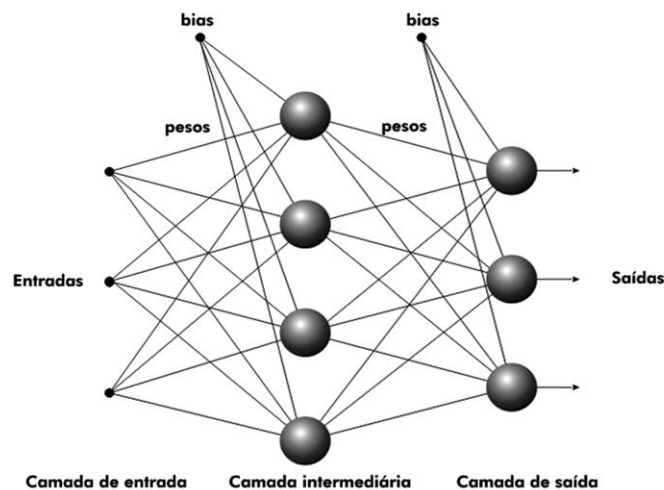
$w_{k1}, w_{k2}, w_{k3}, \dots, w_{km}$ são pesos sinápticos do neurônio k ;

v_k é a saída do combinador linear (somador);

b_k é o *bias*, ou constante do modelo;

Posteriormente, os neurônios são organizados em uma *rede*, com a saída de um servindo como entrada para outros. A Figura 2 mostra a arquitetura mais comum de rede neural, na qual os neurônios são dispostos em *camadas*. Na figura, os círculos simbolizam os neurônios e as setas representam as ligações sinápticas entre os neurônios.

Figura 2 – Rede neural com 3 camadas



Fonte: Piazzzi (2001)

A função de transferência ou de ativação dá o caráter não linear ao neurônio. Se $\varphi(v_k) = v_k$, então o modelo de neurônio é linear e rede neural funciona de modo análogo a um modelo de regressão linear. A função de transferência a ser escolhida pelo pesquisador é de suma importância para o bom desempenho do modelo a ser utilizado. Algumas formas de função são descritas a seguir:

1) Função de limiar

A saída do neurônio k que emprega esta função de limiar é expressa como:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad (3)$$

Este modelo é conhecido como Modelo de McCulloch-Pitts (1943). A saída de um neurônio assume valor 1 se a saída induzida daquele neurônio é não-negativa, e 0 caso contrário.

2) Função linear por partes:

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq +1/2 \\ v, & \text{se } +1/2 > v > -1/2 \\ 0, & \text{se } v \leq -1/2 \end{cases} \quad (4)$$

Nesta, temos uma região de operação linear entre $-1/2$ e $1/2$ e uma região de saturação fora dela. A função de transferência passa o sinal não modificado na primeira região, mas assume os extremos 0 ou 1 fora dela.

3) Função sigmóide

Esta classe de funções é a mais utilizada em redes neurais artificiais. É uma função estritamente crescente, em forma de S , que exibe um balanceamento adequado entre comportamento linear e não-linear. Como exemplo de função sigmóide, temos a função logística:

$$\varphi(v) = \frac{1}{1 + e^{(-av)}} \quad (5)$$

Onde a é o parâmetro de inclinação da função sigmóide. Quanto maior o parâmetro a mais a função sigmóide se parecerá com uma função de limiar.

Diferentemente da função de limiar, que assume apenas os valores 0 ou 1, uma função sigmóide assume valores entre 0 e 1 de forma contínua e diferenciável, o que é uma grande vantagem do ponto de vista matemático.

Algumas vezes é desejável que a função de transferência se estenda de -1 a $+1$, assumindo neste caso uma forma anti-simétrica em relação à origem. Neste caso, podemos utilizar a função tangente hiperbólica, definida por:

$$\varphi(v) = \tanh h(v) \quad (6)$$

3.2.1 Modelo estocástico de um neurônio

Neste modelo de neurônio, temos uma interpretação probabilística da função de transferência do Modelo de McCulloch-Pitts. Assim temos que um neurônio assume dois estados, +1 ou -1, por exemplo. A mudança de estado de um neurônio é probabilística. Seja x o estado de um neurônio e $P(v)$ a probabilidade de mudança de estado, onde (v) é a saída induzida do neurônio. Temos:

$$x = \begin{cases} +1 & \text{com probabilidade } P(v) \\ -1 & \text{com probabilidade } 1 - P(v) \end{cases} \quad (7)$$

Uma escolha padrão para $P(v)$ é a função de forma sigmóide:

$$P(v) = \frac{1}{1 + e^{(-v/T)}} \quad (8)$$

Onde T é um parâmetro utilizado para controlar o nível de ruído e, portanto, a incerteza de mudança de estado de um neurônio. Note que quando não há ruído sináptico, isto é, quando $T=0$, o neurônio estocástico se reduz a forma determinística.

3.2.2 Modelos de redes neurais estáticos e dinâmicos

Podemos classificar os tipos de modelos matemáticos empregados na descrição de redes neurais, segundo Corrêa (2004) em duas categorias:

1) Modelos Estáticos:

A função da rede neural é estabelecer uma relação estímulo-resposta de acordo com a entrada por ela recebida. Essas redes são denominadas Redes *Feedforward* ou de alimentação adiante. Pois, seus elementos processadores são conectados em uma só direção. Estes tipos de arquitetura de redes, desde que não possuam camadas escondidas ou ocultas, utilizem funções de transferência lineares e possuam apenas um único neurônio na saída da rede, são análogos aos modelos de regressões lineares, onde os pesos sinápticos podem ser vistos como coeficientes das regressões e o conjunto destes pesos representa a reta ajustada da regressão (KAASTRA e BOYD, 1996). Alguns exemplos de modelos estáticos de redes neurais são: *Perceptron* (ROSENBLATT, 1959); *Perceptron* de Múltiplas camadas ou *Multi-Layer Perceptrons* (RUMELHART e MCCLELLAND, 1986); Mapas auto-organizáveis ou SOM - *Self Organizing Maps* (KOHONEN, 1995); e ART ou *Adaptive Resonance Theory* (CARPENTER e GROSSBERG, 1995).

2) Modelos Dinâmicos:

Nestes tipos de modelos, a rede funciona como um sistema dinâmico, logo, seu funcionamento não é apenas de acordo com a função da entrada, mas é também função do estado em que a rede se encontra. Neste sentido, dizemos que estas redes possuem memória. Este tipo de redes neurais é denominado redes recorrentes, dinâmicas ou realimentadas, pois, possuem realimentação ou *feedback*. Alguns exemplos destes tipos de modelos são: Modelo de Hopfield (HOPFIELD, 1982); Redes Neurais Celulares ou RNCs (CHUA e YANG, 1988); e Osciladores acoplados (MIROLLO e STROGATZ, 1990).

3.3 Algoritmos de aprendizado

Refere-se ao aprendizado por meio de regras bem definidas para solução de problemas de aprendizado. Existem muitos algoritmos específicos para aprendizagem de redes neurais, diferenciando entre si pela maneira como os pesos sinápticos são modificados. Os algoritmos de aprendizado podem se dividir em:

1) Aprendizado supervisionado

Este método de aprendizagem consiste na adoção de uma função de *custo* na rede neural, onde cada vez que a rede erra a resposta desejada para o padrão de entrada, um fator externo (*professor*) indica quais devem ser os ajustes realizados nos pesos sinápticos de acordo com o erro a fim de diminuir esta resposta incorreta da rede. Logo, o objetivo deste tipo de aprendizado é, a cada seqüência de aprendizado, fazer com que a rede atinja uma solução estável, isto é, correções cada vez menores nos pesos sinápticos.

2) Aprendizado não-supervisionado

Como o próprio nome diz, é um método de aprendizado onde as redes neurais não utilizam um *professor*, ou seja, não há um algoritmo de correção das falhas das respostas da rede neural. Neste caso, a rede neural recebe a entrada de dados, e os organiza de maneira arbitrária em categorias. Na próxima entrada de dados, esta faz uma verificação se os dados correspondem a alguma categoria já existente e, em caso de não existir, ela cria uma ou mais categorias para classificar estes novos dados.

Segundo Santos (2001), dentro do tipo de aprendizado supervisionado temos:

- *Off-line*: Quando os dados do conjunto de treinamento são constantes, assim quando a rede encontra uma solução, esta permanece fixa. Contudo, se há introdução de novos dados no conjunto de treinamento, deve-se realizar um novo treinamento da rede.
- *On-line*: Quando os dados do conjunto de treinamento estão em permanente mudança, logo a rede está num processo de adaptação contínuo.

3) Aprendizado competitivo

Neste método, os neurônios competem entre si para processar os dados de entrada, comparando suas ponderações e apenas um neurônio é ativado, tratando de um caso particular de aprendizado não-supervisionado. Logo, os neurônios se especializam de maneira a organizar e criar categorias de classificação dos padrões de entrada.

De acordo com Santos (2001), um ciclo corresponde a utilização de todos N pares de entrada e saída do conjunto de treinamento, onde a correção dos pesos sinápticos num ciclo pode ser feita de duas diferentes maneiras:

- *Modo padrão*: A cada apresentação da rede de um exemplo do conjunto de treinamento, a correção dos pesos sinápticos é realizada, onde cada correção é baseada somente no erro do exemplo apresentado naquela iteração havendo em cada ciclo N correções.
- *Modo Batch*: Por ciclo, apenas uma única correção é realizada onde todos os exemplos do conjunto de treinamento são apresentados à rede, é calculado seu erro-médio e , a partir deste, são feitas as correções dos pesos sinápticos.

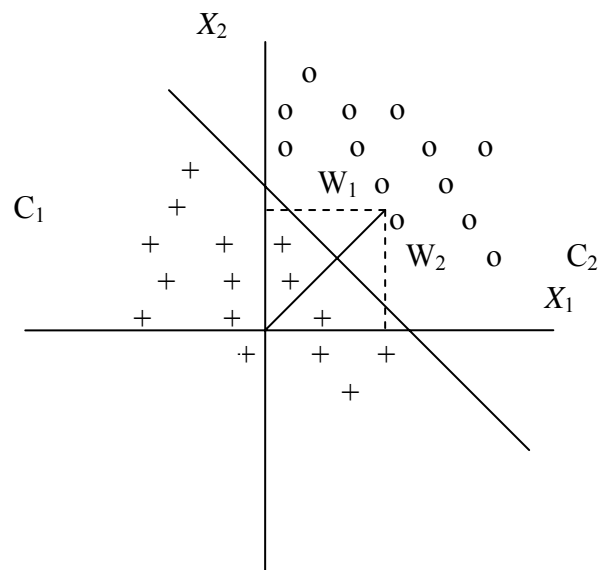
4) Aprendizado por reforço

De acordo com Santos (2001), o aprendizado por reforço constitui um caso particular do aprendizado supervisionado, pois a resposta da rede é avaliada externamente. A informação de realimentação dada à rede é se determinada saída está ou não correta, logo não se fornece a resposta correta para o padrão de entrada. Assim, não se fornece a resposta para o padrão de entrada. Portanto, a diferença deste método de aprendizado se reside no supervisionamento e na medida de desempenho ou reforço. Este método é uma forma *on-line* de aprendizado que obtém um mapeamento da entrada-saída através da busca e erro desenvolvido para maximizar o índice de desempenho escalar.

3.4 O *Perceptron*

Uma rede neural com arquitetura análoga à representada na Figura 2, com neurônios utilizando funções de ativação do tipo limiar é chamada *Perceptron*. *Perceptrons* são utilizados para reconhecimento de padrões, como o reconhecimento de escrita ou de voz. Os dados de entrada consistem de um vetor (x_1, x_2, \dots, x_m) e a saída consiste em uma de N categorias previamente determinadas (por exemplo, as letras *A*, *B*, *C*, etc). A Figura 3 ilustra um exemplo de separação de dados em dois padrões linearmente separáveis. O objetivo da rede é encontrar uma reta que separe os dois conjuntos. A equação desta reta é dada pelos pesos sinápticos mais as constantes b_k .

Figura 3 – Dados linearmente separáveis pelo *Perceptron*



Minsky e Papert (1969) analisaram matematicamente o *perceptron* e demonstraram que redes de uma camada não são capazes de solucionar problemas que não sejam linearmente separáveis. Como não acreditavam na possibilidade de se construir um método de treinamento para redes com mais de uma camada, eles concluíram que as redes neurais seriam sempre suscetíveis a essa limitação.

Contudo, o desenvolvimento do algoritmo de treinamento *backpropagation*, por Rumelhart, Hinton e Williams (1986), precedido por propostas semelhantes ocorridas a partir dos 1970's, mostrou que é possível treinar eficientemente redes com camadas escondidas, resultando no modelo de redes neurais artificiais mais utilizados atualmente, as redes *perceptron* multi-camadas (*Multi-Layer Perceptron*), treinadas com o algoritmo *backpropagation* (KUMAR, RAO e SONI, 1995).

Uma outra limitação refere-se ao tempo de treinamento de redes neurais utilizando *backpropagation*, que tende a ser muito lento. Algumas vezes são necessários milhares de ciclos

para se chegar a níveis de erros aceitáveis, principalmente se estiver sendo simulado em computadores seriais, pois a CPU deve calcular as funções para cada unidade e suas conexões separadamente, o que pode ser problemático em redes muito grandes, ou com grande quantidade de dados.

3.4.1 Algoritmo de aprendizagem

Conforme Cardoso (1999), este algoritmo é utilizado para separar classes *linearmente separáveis*. Considere o seguinte vetor de entrada do *perceptron*:

$$X = [-1, x_1, x_2, \dots, x_p]^T \quad (9)$$

O vetor de pesos sinápticos é escrito da seguinte maneira:

$$W = [\theta, w_1, w_2, \dots, w_p]^T \quad (10)$$

Escrevendo a combinação linear destes vetores na forma vetorial, obtemos:

$$u = W^T X \quad (11)$$

O objetivo do *perceptron* é encontrar um vetor de pesos sinápticos W que satisfaça a equação:

$$W^T X = 0 \quad (12)$$

A equação (12) define um hiperplano no espaço de p dimensões que representa uma superfície de separação entre duas classes diferentes que iremos chamar de C_1 e C_2 . Na Figura 3, podemos observar geometricamente este funcionamento, onde os vetores dos pesos sinápticos são sempre perpendiculares ao hiperplano definido pela equação (10).

Suponhamos agora que existam dois subconjuntos diferentes de vetores de treinamento, um pertencente à classe C_1 e outro pertencente à classe C_2 . A união entre estes dois subconjuntos forma o conjunto de treinamento.

O objetivo do treinamento do *perceptron* com os conjuntos X_1 e X_2 é encontrar um vetor de pesos W que divida as classes C_1 e C_2 . De acordo com Roseblatt (1962), o vetor W é obtido em um número finito de passos pelo algoritmo de aprendizado do *perceptron* descritos a seguir:

1. Inicializa-se aleatoriamente o vetor de pesos sinápticos com W podendo assumir qualquer valor.
2. Apresenta-se um novo vetor de treinamento X_i à rede e calcula-se o produto.
- 3.

$$u_i = W^T X_i \quad (13)$$

4. A rede verifica se o vetor X_i está corretamente classificado. No caso de uma classificação correta, não haverá modificação dos no vetor de pesos sinápticos W , isto é:

$$W(n+1) = W(n) \quad \text{se} \quad W^T X(n) \geq 0 \quad \text{e} \quad X(n) \in C_1 \quad (14)$$

e

$$W(n+1) = W(n) \quad \text{se} \quad W^T X(n) < 0 \quad \text{e} \quad X(n) \in C_2 \quad (15)$$

Onde n se refere à n -ésima iteração.

5. Caso contrário, o vetor de pesos é atualizado baseado na seguinte regra:

6.

$$W(n+1) = W(n) - \eta W(n) \quad \text{se} \quad W^T X(n) \geq 0 \quad \text{e} \quad X(n) \in C_2 \quad (16)$$

e

$$W(n+1) = W(n) + \eta W(n) \quad \text{se} \quad W^T X(n) < 0 \quad \text{e} \quad X(n) \in C_1 \quad (17)$$

Onde η é a taxa de aprendizado.

7. Retorna-se ao passo número 2 até que todos os vetores X estejam classificados corretamente.

Para esclarecer melhor a explicação da regra de aprendizagem do *perceptron*, apresentaremos um teorema descrito em Veelenturf (1995).

Teorema 1:

Durante um processo de aprendizagem, um *perceptron* nunca entrará num mesmo estado de pesos sinápticos mais de uma vez se existir um espaço de solução.

Prova:

Dado que $W(k+n)$ seja um estado dos pesos sinápticos alcançado a partir do vetor de pesos $W(k)$ depois de n adaptações, temos então os vetores somados a $W(k)$:

$$\varepsilon(\sum x_i - \sum x_j) \text{ com } x_i \in T_{W(k+p)}^+ \text{ e } x_j \in T_{W(k+p)}^- \text{ com } p = (0,1,\dots,n-1) \quad (18)$$

Onde:

x_i é o conjunto de entradas da rede;

$T_{W(k+p)}^+$ é o conjunto de respostas para os quais a saída é igual a 1, considerando que a saída atual é igual a 0;

$T_{W(k+p)}^-$ é o conjunto de respostas para os quais a saída é igual a 0 e a saída atual é 1;

ε é uma constante proporcional;

p é a entrada quando $W(0)$;

k é o número de rodadas de aprendizado do *perceptron*.

Deixando que S seja o vetor de solução, isto é, o vetor de pesos sinápticos com os quais o *perceptron* encontra a verdadeira solução, considere o produto interno dos vetores S e $W(k+n)$:

$$SW(k+n) = SW(k) + \varepsilon(\sum Sx_i - Sx_j) \quad (19)$$

$$\forall x_i \in T_{W(k+p)}^+ \text{ e } x_j \in T_{W(k+p)}^- \text{ com } p = (0,1,\dots,n-1)$$

Para cada $x_i \in T_{W(k+p)}^+$ temos que $Sx_i > 0$ e para cada $x_j \in T_{W(k+p)}^-$, temos $Sx_j \leq 0$. Assim, $SW(k+n) \neq SW(k)$ e, conseqüentemente, $W(k+n) \neq W(k)$. Talvez, na possibilidade de que $T_{W(k+p)}^+$ seja um conjunto vazio para todo p , logo $Sx_i = 0$, então $SW(k+n) = SW(k)$. Contudo, neste caso há outro vetor de solução a partir do espaço de solução tal que $Sx_j \neq 0$.

Teorema 2:

Segundo Block (1962) e Novikoff (1962), citados em Ng (2004), seja uma seqüência de exemplos $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$. Suponha que $\|x_i\| \leq D \quad \forall i$, assim, existe um vetor de comprimento u ($\|u\|_2 = 1$), tal que $y_i(u^T, x_i) \geq \gamma$, para todos os exemplos da seqüência, isto é, $u^T, x_i \geq \gamma$ se $y_i = 1$, e $u^T, x_i \leq -\gamma$ se $y_i = -1$, de forma que u separa os dados com uma margem de pelo menos γ . Então, o número total de erros que o algoritmo do *perceptron* comete nesta seqüência é no máximo de $\left(\frac{D}{\gamma}\right)^2$.

Prova:

O *perceptron* só atualiza seus pesos sinápticos nos exemplos nos quais comete erro. Deixe que W_k sejam os pesos sinápticos quando o *perceptron* cometeu seu k -ésimo erro. Assim, $W_k = \vec{0}$ e se o k -ésimo erro estivesse no exemplo (x_i, y_i) , então $g((x_i)^T W_k)$ que implica em:

$$(x_i)^T W_k y_i \leq 0 \quad (20)$$

Também, da regra de aprendizado do *perceptron*, nós temos que

$$W_{k+1} = W_k + y_i x_i \quad (21)$$

Segue que:

$$\begin{aligned} (W_{k+1})^T u &= (W_k)^T u + y_i (x_i)^T u \\ &\geq (W_k)^T u + \gamma \end{aligned} \quad (22)$$

Implicando diretamente que:

$$(W_{k+1})^T u \geq k\gamma \quad (23)$$

Logo:

$$\begin{aligned} \|W_{k+1}\|^2 &= \|W_k + y_i x_i\|^2 \\ &= \|W_k\|^2 + \|x_i\|^2 + 2y_i (x_i)^T W_k \\ &\leq \|W_k\|^2 + \|x_i\|^2 \\ &\leq \|W_k\|^2 + D^2 \end{aligned} \quad (24)$$

No terceiro passo denotado pela Equação (24), foi utilizada a equação (20). Além disso, podemos observar que a equação (24), implica em:

$$\|W_{k+1}\|^2 \leq kD^2 \quad (25)$$

Reunindo as equações (23) e (24) chegamos a:

$$\begin{aligned} \sqrt{k}D &\geq \|W_{k+1}\| \\ &\geq (W_{k+1})^T u \\ &\geq k\gamma \end{aligned} \quad (26)$$

A segunda desigualdade denotada por $\sqrt{k}D \geq (W_{k+1})^T u$ surge do fato que u é um vetor de comprimento igual a um (e $z^T u = \|z\| \|u\| \cos\phi \leq \|z\| \|u\|$, onde ϕ é o ângulo entre z e u).

Este resultado implica que $k \leq \left(\frac{D}{\gamma}\right)^2$. Conseqüentemente, se o *perceptron* cometesse um k -

ésimo erro, então $k \leq \left(\frac{D}{\gamma}\right)^2$.

Segundo Kröse e Smagt (1996), para a regra de aprendizagem do *perceptron*, há o *Teorema da Convergência*:

Teorema 3:

Se existe um conjunto de pesos sinápticos W^* que pode executar a transformação $y = d(x)$, a regra de aprendizado do *perceptron* convergirá para alguma solução (que pode ou pode não ser igual a W^*) em um número finito de passos para qualquer escolha de inicial dos pesos sinápticos.

Prova:

Levando em consideração que o comprimento do vetor W^* não importa (por causa da função de limiar), toma-se $\|W^*\| = 1$. Porque W^* é uma solução correta, o valor de $|W^* x|$ denota um escalar ou um produto interno, e será maior que 0 ou existirá um $\delta > 0$ tal que $|W^* x| > \delta$ para todas as entradas x . Definimos $\cos \alpha \equiv W W^* / \|W\|$. De acordo com a regra de aprendizagem do *perceptron*, os pesos sinápticos são modificados com a adição de uma nova entrada x , como $\Delta W = d(x)x$ e o peso sináptico depois desta modificação é $W^T = W + \Delta W$, segue-se:

$$\begin{aligned}
W^T W^* &= WW^* + d(x)W^*x \\
&= WW^* + \text{sgn}(W^*x)W^*x \\
&> WW^* + \delta
\end{aligned} \tag{27}$$

Onde sgn denota uma função de limiar.

$$\begin{aligned}
\|W^T\|^2 &= \|W + d(x)x\|^2 \\
&= W^2 + 2d(x)Wx + x^2 \\
&< W^2 + x^2 \quad \text{pois} \quad d(x) = -\text{sgn}[Wx] \\
&= W^2 + M
\end{aligned} \tag{28}$$

Depois de k rodadas de aprendizado, temos:

$$W(k)W^* > WW^* + k\delta \tag{29}$$

$$\|W(k)\|^2 < W^2 + kM \tag{30}$$

Logo,

$$\cos \alpha(k) = \frac{W^*W(k)}{\|W(k)\|} \tag{31}$$

$$\cos \alpha(k) > \frac{W^*W + k\delta}{\sqrt{W^2 + t(M)}} \tag{32}$$

Da equação (31) segue que:

$$\lim_{k \rightarrow \infty} \cos \alpha(k) = \lim_{k \rightarrow \infty} \frac{\delta}{\sqrt{M}} \sqrt{k} = \infty \quad (33)$$

Enquanto pela definição $\cos \alpha \leq 1$.

Chegamos à conclusão de que deve haver um limite superior k_{max} para k . As modificações nos pesos sinápticos do sistema possuem um número limitado de ocorrências, isto é, depois de, no máximo, k_{max} modificações dos pesos sinápticos, o *perceptron* está fornecendo corretamente as respostas. Quando $\cos \alpha = 1$, k_{max} é alcançado. Se a inicialização dos pesos sinápticos for com $W = 0$:

$$k_{max} = \frac{M}{\delta^2} \quad (34)$$

Aprendizado pela Regra Delta

De acordo com Kröse e Smagt (1996), para uma rede neural simples de uma única camada com uma saída que utiliza função de transferência linear, a saída do neurônio k é dada simplesmente por:

$$y_k = \sum_j w_{kj} x_j + b_k \quad (35)$$

Como uma simples rede neural pode representar uma relação linear entre o valor da unidade de saída e o valor das unidades de entrada através de uma função de limiar do valor da saída, um classificador pode ser construído, (como o *ADALINE*), mas aqui o objetivo é a relação linear e o uso da rede neural para uma tarefa de função de aproximação. Num espaço de entradas com muitas dimensões, a rede neural representa um hiperplano e múltiplas unidades de saída podem ser definidas.

Suponha que o objetivo seja treinar a rede neural de tal forma que um hiperplano seja calculado para um conjunto de amostras de treinamento que consistem em valores de entrada X^p e valores de saídas desejadas d^p . Para determinada amostra de entradas, a saída da rede neural se diferencia do valor desejado por $(d^p - Y^p)$, onde Y^p é a saída atual para este padrão de entradas. A regra delta se baseia num custo ou uma função de erro baseada nesta diferença para ajustar os pesos sinápticos. A função de erro ou de performance, indicada pelo nome *Least Mean Square (LMS)*, é a soma dos erros quadrados, quer dizer, o erro total E é definido por:

$$E = \sum_p E^p = \frac{1}{2} \sum_p (d^p - Y^p)^2 \quad (36)$$

Onde o índice p representa um padrão de entradas e E_p representa o erro para o padrão de entradas p .

O procedimento *LMS* encontra os valores de todos os pesos sinápticos que minimizam a função de erro por um método chamado descida de gradiente. A idéia é fazer uma mudança no peso sináptico negativamente proporcional à derivada do erro como medido no padrão atual de entradas com respeito a cada peso sináptico:

$$\Delta_p w_j = -\gamma \frac{\partial E^p}{\partial w_j} \quad (37)$$

Onde γ é uma constante de proporcionalidade, também chamada de constante de *momentum*. Derivando a equação (37):

$$\frac{\partial E^p}{\partial w_j} = \frac{\partial E^p}{\partial Y^p} \frac{\partial Y^p}{\partial w_j} \quad (38)$$

Por conta das unidades lineares, temos:

$$\frac{\partial Y^p}{\partial w_j} = x_j \quad (39)$$

e

$$\frac{\partial E^p}{\partial Y^p} = -(d^p - Y^p) \quad (40)$$

Tal que:

$$\Delta_p w_j = \gamma \delta^p x_j \quad (41)$$

Onde $\delta^p = d^p - Y^p$ é a diferença entre a saída desejada e a atual saída produzida pela rede neural para um tipo de padrão de entrada p .

A regra delta modifica o peso sináptico adequadamente da atual saída para a saída desejada tanto para entradas ou saídas contínuas ou binárias. Este algoritmo permitiu a aplicação das redes neurais para uma série de novos problemas.

3.5 Multi-Layer Perceptron

A regra de aprendizado utilizada para treinar uma rede *multilayer perceptron* é chamada Regra Delta Generalizada ou *backpropagation*. Quando um padrão de entrada é apresentado à rede pela primeira vez, esta produz uma saída aleatória. A diferença entre esta saída e a saída desejada constitui o erro. A intenção do treinamento é buscar cada vez mais diminuir o valor deste erro. O método minimiza o erro e realiza iterativamente ajustes nos pesos sinápticos, passando por todos os padrões de entradas do conjunto de treinamento. Esse método é conhecido como descida de gradiente (*gradient descent*). O processo de aprendizagem prossegue, até que algum critério seja estabelecido como, um mínimo de erro global, ou uma diferença sucessiva mínima entre erros calculados para cada iteração. O *backpropagation* é, na verdade, um método para implementar o processo de gradiente de descida no espaço dos pesos sinápticos para o treinamento de uma rede *feedforward*.

Segundo Brondino (1999), o algoritmo de *backpropagation* pode apresentar alguns problemas em relação à dificuldade de aprendizado. Ao se deparar com padrões de entradas muito complexos, o treinamento é muito lento e existe o risco da rede neural cair num mínimo local.

Paralelamente, também há desvantagem em utilizar camadas escondidas, pois o aprendizado se torna muito mais difícil. A característica principal da camada escondida é que seus elementos se organizam de tal forma que cada elemento aprenda a reconhecer características diferentes do espaço de entrada, assim, o algoritmo de treinamento deve decidir que características devem ser extraídas do conjunto de treinamento.

3.5.1 Teorema da aproximação universal

Um *Multi Layer Perceptron* trata-se de um aproximador universal de funções, assim examine o teorema da aproximação universal descrito em Cybenko (1989):

Considere $\phi(\cdot)$ como uma função contínua, não constante, limitada e monotonicamente crescente. Deixe que I_p denote o hiper cubo unitário p -dimensional. O espaço de funções contínuas em I_p é denotado por $C(I_p)$. Logo, dada qualquer função pertencente a $C(I_p)$ e $\epsilon > 0$, existe um inteiro M e um conjunto de constantes reais α_i, w_{ij} e b_{ij} , onde $i=1, \dots, M$ e $j=1, \dots, p$, de tal forma que temos como definir:

$$F(x_1, \dots, x_p) = \sum_{i=1}^M \alpha_i \phi \left(\sum_{j=1}^p w_{ij} x_j + b_{ij} \right) \quad (42)$$

como uma realização aproximada da função $f(\cdot)$, i.e., $|F(x_1, \dots, x_p) - f(x_1, \dots, x_p)| < \epsilon$ para todo $\{x_1, \dots, x_p\} \in I_p$.

3.5.2 Regra de aprendizado do *Multi-Layer Perceptron*

Em conformidade com Kröse e Smagt (1996), considerando que estamos utilizando unidades com função de transferência não-linear, temos que generalizar a regra delta para este tipo de função de transferência.

A função de transferência é uma função diferenciável da entrada total, dada por:

$$y_k^p = F(s_k^p) \quad (43)$$

Onde:

y_k^p é a saída produzida pela rede neural num certo momento do processo de aprendizado para um dado padrão de entrada p ;

s_k^p é o número de neurônios utilizados pela rede neural num certo momento do processo de aprendizado para um dado padrão de entrada p , isto é:

$$s_k^p = \sum_j w_{jk} y_j^p + b_k \quad (44)$$

Para obter a generalização correta da regra delta, é necessário definir:

$$\Delta_p w_{jk} = -\gamma \frac{\partial E^p}{\partial w_{jk}} \quad (45)$$

Onde:

Δ_p é uma constante que indica o quanto deve o peso sináptico alterar num certo momento do processo de aprendizado para um dado padrão de entrada p ;

γ é uma constante de proporcionalidade;

E^p é a soma dos erros quadrados num certo momento do processo de aprendizado para um dado padrão de entrada p .

Logo, podemos definir E^p como:

$$E^p = \frac{1}{2} \sum_{o=1}^{N_o} (d_o^p - y_o^p)^2 \quad (46)$$

Onde:

d_o^p é a saída desejada para unidade o quando o padrão de entrada p é mantido.

Continuamos definindo $E = \sum_p E^p$ como a soma dos erros quadrados. Podemos

escrever:

$$\frac{\partial E^p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}} \quad (47)$$

Pela equação (45) podemos observar que o segundo fator do lado direito da equação (47)

é:

$$\frac{\partial s_k^p}{\partial w_{jk}} = y_j^p \quad (48)$$

Podemos definir:

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} \quad (49)$$

Alterando a regra de atualização, que é equivalente à regra delta, temos que o resultado é uma descida de gradiente na superfície de erro. Assim, as alterações nas mudanças dos pesos sinápticos são feitas de acordo com:

$$\Delta_p w_{jk} = \gamma \delta_k^p y_j^p \quad (50)$$

Desta forma, o objetivo é definir qual deveria ser δ_k^p para cada unidade k da rede neural. O resultado derivado agora, é que há uma computação recursiva simples destes δ 's que podem ser implementados propagando sinais de erro para trás pela rede neural.

Para computar δ_k^p , aplica-se a regra da cadeia para escrever esta derivada parcial como produto de dois fatores, um que reflete a mudança do erro como uma função da saída da unidade k e outro que reflete a mudança da saída como uma função de mudanças na entrada. Podemos escrever:

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} = -\frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial s_k^p} \quad (51)$$

Substituindo a equação (43) no segundo fator do lado direito da equação (51), temos:

$$\frac{\partial y_k^p}{\partial s_k^p} = F'(s_k^p) \quad (52)$$

A equação (52) é a derivada da função F para a k -ésima unidade, avaliada na entrada s_k^p para a unidade k . Para substituir a equação (52) no primeiro fator do lado direito da equação (51), em primeiro lugar, tomamos que a unidade k é uma unidade de saída $k = o$ da rede neural. Neste caso, segue-se que da definição de E^p definida pela equação (46) que:

$$\frac{\partial E^p}{\partial y_o^p} = -(d_o^p - y_o^p) \quad (53)$$

Substituindo a equação (53) na equação (51), chegamos a:

$$\delta_o^p = (d_o^p - y_o^p)F'_o(s_o^p), \quad \forall o \quad (54)$$

Em segundo lugar, se k não é uma unidade de saída, mas uma unidade escondida $k = h$, não é conhecida a contribuição da unidade ao erro da saída da rede neural. Porém, a medida de erro pode ser escrita como uma função dos pesos da entrada da camada escondida para a camada de saída; $E^p = E^p(s_1^p, s_2^p, s_3^p, \dots, s_j^p)$ e utilizando novamente a regra da cadeia podemos escrever:

$$\frac{\partial E^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial s_o^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial}{\partial y_h^p} \sum_{j=1}^{N_h} w_{ko} y_j^p = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} w_{ko} = - \sum_{o=1}^{N_o} \delta_o^p w_{ho} \quad (55)$$

Substituindo a equação (55) na equação (51):

$$\delta_h^p = F'_h(s_h^p) \sum_{o=1}^{N_o} \delta_o^p w_{ho} \quad (56)$$

As equações (55) e (56) têm um procedimento recursivo para computar os δ 's para todas as k unidades da rede neural que é usada para ajustar as mudanças dos pesos sinápticos de acordo com equação (56). Este procedimento constitui a regra de delta generalizada para uma rede *feed-forward* de unidades não-lineares.

Após a definição matemática do algoritmo de aprendizagem, é possível descrever o algoritmo de *backpropagation* de uma forma mais ampla, como descrito em Ambrósio (2002):

1. Inicializar

- Definir os pesos sinápticos das conexões (através de uma distribuição uniforme aleatória com média zero).

2. Apresentar os padrões de entrada

- Reordenar aleatoriamente o conjunto de treinamento;
- Alimentar a rede com um exemplo do conjunto de treinamento;
- Para todos os exemplos do conjunto de treinamento, individualmente, executar os passos 3 e 4.

3. Propagar os sinais

- Calcular os sinais de todos os neurônios;
- Calcular os sinais de erro;
- Calcular o valor instantâneo do total de erros.

4. Retropropagar

- Calcular os gradientes locais;
- Calcular os deltas;
- Ajustar os pesos das conexões.

5. Iterar

- Após apresentação de todo conjunto de treinamento, calcular a média do erro;
- Voltar ao passo 2 até que a média do erro se encontre estabilizada no valor mínimo.

3.5.3 Treinamento Levenberg-Marquardt

Para melhorar a performance de treinamento do MLP podemos utilizar o algoritmo de treinamento de Levenberg-Marquardt (HANGAN e MENJAH, 1994). Este método realiza uma aproximação do método Newton. Logo, a forma de atualização dos pesos sinápticos será:

$$\Delta w = -[\nabla^2 E(w)]^{-1} \nabla E(w) \quad (57)$$

Onde $[\nabla^2 E(w)]^{-1}$ é a inversa da matriz Hessiana. Este método converge em um passo para funções quadráticas. Para o caso de funções não quadráticas podemos fazer a atualização como:

$$\Delta w = -\alpha [\nabla^2 E(w)]^{-1} \nabla E(w) \quad (58)$$

Onde α é um ganho não negativo. A partir da equação (53) chegamos a:

$$\nabla E(w) = J^T(w)e(w) \quad (59)$$

$$\nabla^2 E(w) = J^T(w)J(w) + S(w) \quad (60)$$

Onde $J(w)$ é a matriz Jacobiana, como apresentado na equação (61).

$$J(w) = \begin{bmatrix} \frac{\partial e_1(w)}{\partial w_1} & \dots & \frac{\partial e_1(w)}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_T(w)}{\partial w_1} & \dots & \frac{\partial e_T(w)}{\partial w_n} \end{bmatrix} \quad (61)$$

No método Gauss-Newton, assume-se que uma aproximação linear é de forma que $S(w) \approx 0$. Logo, a forma de atualização pode então ser descrita da seguinte forma:

$$\Delta w = [J^T(w)J(w)]^{-1} J^T(w)e(w) \quad (62)$$

O método de Levenberg-Marquardt é uma modificação do método de Gauss-Newton, e aquele pode ser descrito pela equação (63) como segue:

$$\Delta w = [J^T(w)J(w) + \mu I]^{-1} J^T(w)e(w) \quad (63)$$

O parâmetro μ controla o tamanho do passo do método. O método Levenberg-Marquardt é reconhecido na literatura como um dos métodos de taxa de convergência mais elevada (HANGAN e MENJAH, Op. Cit.).

3.5.4 Treinamento *Resilient-Propagation* (RPROP)

Ao contrário o algoritmo de *backpropagation* tradicional, o algoritmo *Resilient-Propagation* também chamado de RPROP, desenvolvido inicialmente por Riedmiller e Braun (1993), utiliza apenas sinais de derivadas parciais para ajustar os coeficientes dos pesos sinápticos. Este algoritmo utiliza o método de iterações e reajusta os pesos sinápticos após todos os padrões da sub-amostra de treinamento serem apresentados à rede.

O algoritmo RPROP ajusta os pesos sinápticos da rede neural seguindo a regra:

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^+ \Delta_{ij}^{(t)}, & \frac{\partial E^{(t)}}{\partial w_{ij}} \frac{\partial E^{(t-1)}}{\partial w_{ij}} > 0 \\ \eta^- \Delta_{ij}^{(t)}, & \frac{\partial E^{(t)}}{\partial w_{ij}} \frac{\partial E^{(t-1)}}{\partial w_{ij}} < 0 \end{cases} \quad (64)$$

$$0 < \eta^- < 1 < \eta^+$$

Segundo Akobir [2005?], se em determinado momento o peso sináptico correspondente da derivada parcial w_{ij} inverter seu sinal, é por que sua última modificação foi elevada e o algoritmo perdeu o mínimo local. Então, o tamanho da modificação deveria diminuir determinado pelo fator η^- e o valor prévio do coeficiente do peso sináptico restabelecido. Em outras palavras, esta modificação deveria ser desfeita.

$$\Delta w_{ij}(t) = \Delta w_{ij}(t) - \Delta_{ij}^{(t-1)} \quad (65)$$

Se o sinal da derivada parcial não mudou, então o ajuste deveria ser aumentado por $\eta+$ para alcançar mais rapidamente a convergência. Fixados os fatores de diminuição η^- e de aumento $\eta+$ podemos manter as configurações globais da rede de neural inalteradas. Logo, este fato poderia ser uma vantagem do algoritmo *Resilient-Propagation* em relação ao algoritmo de *backpropagation* tradicional (AKOBIR, Op. Cit.).

Para evitar valores de pesos sinápticos muito altos ou muito baixos, o valor de ajuste é determinado acima por Δ_{\max} e abaixo por Δ_{\min} .

A regra demonstrada pela equação (66) é utilizada para encontrar o valor de ajuste dos pesos sinápticos:

$$\Delta w_{ij}(t) = \left\{ \begin{array}{l} -\Delta_{ij}^{(t)}, \frac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, \frac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ 0, \frac{\partial E^{(t)}}{\partial w_{ij}} = 0 \end{array} \right\} \quad (66)$$

Se a derivada é positiva, isto é, o erro aumenta, o coeficiente do peso sináptico é diminuído pelo valor de ajuste, caso contrário é aumentado. Logo, a equação ilustra a forma como os pesos são ajustados:

$$w_{ij}(t+1) = w_{ij}(t) - \Delta w_{ij}(t) \quad (67)$$

3.5.5 Regularização

Além da função de performance da rede neural *LMS*, definida anteriormente para o caso do *perceptron* pela equação (36), temos modificações destas funções de performance. O objetivo destas modificações é melhorar a performance de generalização da rede neural, a qual é chamada de regularização.

Função de Performance modificada

Uma definição alternativa para a função tradicional de performance utilizada para treinar uma rede neural do tipo *feedforward* é a *LMS* descrita em Demuth e Beale (1998).

$$F = mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \quad (68)$$

É possível melhorar o poder de generalização da rede neural modificando a função de performance somando um termo que consiste na média da soma dos quadrados dos pesos e *bias* da rede neural:

$$msereg = \gamma mse + (1 - \gamma)msw \quad (69)$$

Onde γ é a razão de performance e,

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (70)$$

Ao usar esta função de performance a rede irá diminuir seus pesos sinápticos e *bias*, e isto forçará a resposta de rede a ser mais suave e menos sujeita a problemas de *overfitting*.

O problema com a regularização é que é difícil determinar um valor ótimo para o parâmetro da razão de performance. Se escolhermos um valor para este parâmetro elevado, podemos incorrer em *overfitting*. Se esta razão for muito pequena, a rede não vai generalizar adequadamente a amostra de treinamento.

4 SELEÇÃO E TRATAMENTO DA AMOSTRA DE DADOS

Como mencionado anteriormente na seção de determinantes das atribuições de *ratings*, selecionamos alguns determinantes macroeconômicos que explicam estas atribuições, os quais serão utilizados para cada país da amostra utilizada.

A amostra de *ratings* utilizada compreende os *ratings* de longo prazo em moeda estrangeira atribuídos pela agência *Standard & Poor's*, a partir de 1975, fornecidos pelo relatório de 2 de novembro de 2005. Das 23 escalas de *ratings* atribuídas por esta agência apresentadas na Tabela 1, esta amostra de *ratings* possui 18 escalas diferentes⁵. É importante ressaltar que o início da amostra temporal de *ratings* se dá a partir em 1975, porém isso não quer dizer que todos os países tenham atribuições de *ratings* se iniciando em 1975. Cada país da amostra foi incluso de acordo com a sua primeira atribuição de *rating*, valendo dizer que as redes neurais conseguem trabalhar com estes tipos de entrada de dados, o que, normalmente, não é possível nos modelos econométricos tradicionais.

Descartamos outras informações contidas no relatório da *Standard & Poor's* como os *ratings* de curto prazo tanto em moeda doméstica quanto moeda estrangeira e os *ratings* de longo prazo em moeda doméstica, pois estes não possuíam uma frequência constante de atribuição entre os países, o que prejudicaria o trabalho. Logo, utilizamos os *ratings* de longo prazo em moeda estrangeira, pois estes, além de refletir uma sinalização sobre as condições de crédito de longo prazo destes países, possuem frequência constante de atribuições entre os países, isto é, uma vez que um dos países receba uma atribuição deste tipo de *rating*, ele sempre receberá tal tipo de atribuição, ao passo que, nos demais tipos de *ratings*, isso não se verifica, o que torna a amostra destes tipos de *ratings* incompleta.

⁵ As escalas AA+, AA-, C, SD e D, não foram encontradas na amostra utilizada de atribuições de *ratings*.

Assim, começamos restringindo a amostra de países pelos que receberam atribuição de *ratings* de longo prazo em moeda estrangeira. Contudo, certos países da amostra inicial não tinham publicados entre suas estatísticas, alguns dos determinantes utilizados neste trabalho como algumas variáveis de liquidez (dívida total como percentual do produto), ou de solvência (serviço da dívida externa como percentual do produto), o que levou a restringirmos a amostra a 46 países.

As variáveis utilizadas no trabalho como determinantes macroeconômicos das atribuições de *ratings* foram coletadas nas estatísticas de indicadores de desenvolvimento dos países publicadas pelo Banco Mundial.

A publicação *Indicadores de Desenvolvimento do Mundo* realizada pelo Banco Mundial, possui observações com frequência anual e término em 2003. Como estas observações fizeram referência às estatísticas de final de período, utilizamos as atribuições de *ratings* também de final de período para o caso daqueles países que receberam mais de uma atribuição de *rating* por ano. Portanto, para os países que a agência *Standard & Poor's* atribuiu mais de uma escala de *rating* no mesmo ano, utilizamos nas estimações deste trabalho apenas observações da última atribuição de *rating* naquele ano.

Além disso, os *outliers* da amostra foram excluídos, o que não gera nenhuma perda de informação em nosso modelo e, diferentemente dos modelos econométricos onde existem tratamentos para este tipo de problema, os *outliers* em redes neurais podem e devem ser excluídos da amostra. A razão pela qual se excluiu *outliers* da amostra é simples. Na fase de treinamento das possíveis arquiteturas utilizadas na estimação final do trabalho, os *outliers* dificultam este treinamento, o que pode causar, entre outras conseqüências, perda de robustez do modelo.

5 IMPLEMENTAÇÃO DA REDE NEURAL

Este capítulo se dedica a explicar resumidamente como será implementada a rede neural utilizando uma amostra de dados de *ratings* de risco soberano e fundamentos macroeconômicos. Enumeramos alguns passos básicos depois de definidos a seleção e tratamento da amostra de dados para que o leitor consiga mais facilmente acompanhar como isto foi realizado.

Após definirmos como organizamos os dados que fazem parte da entrada da rede neural, iremos para um passo mais adiante na implementação da rede neural, constituído pela fase de treinamento, teste e validação desta.

5.1 Treinamento, teste e validação

Segundo Kaastra e Boyd (1996), uma prática comum é dividir a amostra de dados em três sub-amostras, denominadas amostra de treinamento, de teste e de validação. A amostra de treinamento é a maior em tamanho e compreende cerca de 90% do tamanho total da amostra de dados. Os 10% restantes são deixados para o teste e para a validação. O pesquisador deve utilizar o tipo de rede neural que obteve melhor desempenho na fase de treinamento e teste, isto é, a rede neural que possuir menor erro de predição. O teste final do tipo de rede a ser utilizada é a validação. Assim, é importante também a escolha do tamanho da amostra para a validação.

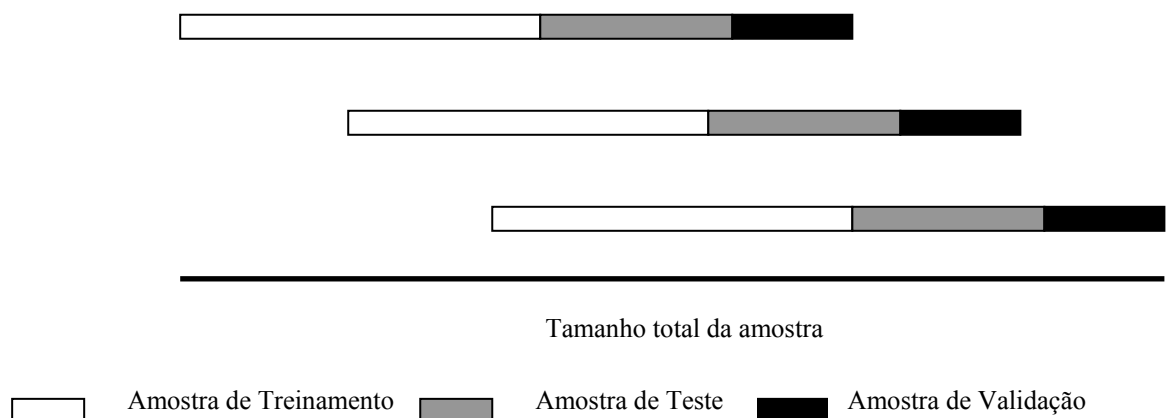
A escolha sobre qual tipo de função de treinamento utilizar na rede neural, assim como outras escolhas na elaboração de uma arquitetura de rede neural, é um processo também experimental. Logo, dado o critério de erro⁶ da rede neural na fase de treinamento, escolhe-se uma função que melhor se adaptou à amostra de dados para ela apresentada.

⁶ O critério a ser adotado para se estabelecer desempenho da rede neural pode ser o *MSE*, *SSE* ou *MSEREG*.

A amostra de teste pode ser definida a partir de duas abordagens diferentes. Ela pode ser escolhida aleatoriamente a partir da amostra de treinamento ou pode ser definida com observações mais recentes que sucedem as observações da amostra de treinamento (para o caso de séries de tempo). Uma vantagem do uso da segunda abordagem é testar a rede com observações mais recentes que podem ser informações mais relevantes para a explicação do problema a ser tratado pelas redes neurais (KAASTRA e BOYD, Op. Cit.).

Uma abordagem mais rigorosa, do ponto de vista da robustez estatística da rede neural na escolha de amostras para o treinamento, teste e validação, seria usar o critério do teste *walk-forward*. Este critério consiste em definir janelas ou sub-amostras a serem utilizadas no treinamento, teste e validação de modo que estas se movam para frente ao longo da amostra total de dados. Assim, definem-se sub-amostras para treinamento, teste e validação da rede neural de modo que, após realizados estas três etapas a primeira vez, as janelas são novamente redefinidas utilizando dados mais recentes e, assim por diante. A Figura 4 ilustra como é o procedimento do critério *walk-forward*.

Figura 4 - Critério de definição de janelas *Walk-Forward*



Fonte: Kaastra e Boyd (1996)

Embora treinar uma rede neural pelo método *walk-forward* aumente o custo computacional, o que dispende mais tempo, devemos chamar a atenção sobre a vantagem que ele possui, pois a rede neural escolhida por este critério tem um maior poder de generalização, logo esta rede neural seria um aproximador mais capacitado a lidar com variações nas observações da amostra de dados.

Entre outras características do método *walk-forward*, está o fato de que este método se adequa a séries de tempo. Assim, como na amostra de dados utilizada neste trabalho temos observações tanto temporais quanto longitudinais, adotamos um outro método de escolha das sub-amostras de treinamento, teste e validação. Dada a amostra total de dados, escolhemos por processo aleatório as três sub-amostras definindo *a priori* apenas o tamanho destas sub-amostras. O objetivo de tomarmos uma escolha aleatória dos dados consiste em desagruparmos estes de modo que o processo de aprendizado da rede possua observações de atribuições de *ratings* e fundamentos macroeconômicos de todos os países que compõem a amostra.

5.2 Arquitetura da rede neural

A arquitetura da rede neural, isto é, o número de camadas de neurônios e o total de neurônios por camada, assim como qual tipo de interconexão entre estes, depende da natureza e do tamanho da amostra de dados que se pretende utilizar na rede neural. O número de neurônios utilizados na entrada e na saída da rede neural é, respectivamente, o número de variáveis independentes e dependentes do modelo a ser estimado.

A arquitetura ideal da rede neural deve ser tal que ela seja tão grande quanto necessária para conseguir obter as representações necessárias, mas ao mesmo tempo pequena o suficiente para se obter um treinamento mais rápido. Não existem regras claras para se definir quantas

unidades devem existir nas camadas escondidas, quantas camadas, ou como devem ser as conexões entre essas unidades (SIU et al., 1995).

5.2.1 Definição do número de camadas escondidas

Teoricamente, uma rede neural com uma camada de neurônios escondida com um número suficiente de neurônios nesta camada, seria capaz de aproximar qualquer tipo de função contínua. Na prática, redes neurais com uma ou duas camadas de neurônios escondidas são amplamente utilizadas e possuem performances bastante satisfatórias.

Deve-se ressaltar que aumentar tanto o número de camadas como também o número de neurônios da rede neural, além de aumentar o custo computacional, também expõe a arquitetura da rede neural a problemas de *overfitting*.

Segundo Kaastra e Boyd (1996), o problema de *overfitting* ocorre quando um modelo de predição, como o de redes neurais, possui muito poucos graus de liberdade. Este problema causa perda de poder de generalização das redes neurais fazendo com que esta ganhe em memorização de idiossincrasias de observações individuais e perda do poder predição. Logo, um modelo de redes neurais robusto deve, analogamente aos modelos econométricos, seguir o critério da parcimônia, o que, além de diminuir o custo computacional da rede neural, aumenta sua capacidade de generalização.

5.2.2 Número de neurônios nas camadas escondidas

Embora não haja qualquer regra sobre quantos neurônios devam ser utilizados por camada escondida numa rede neural, temos algumas *regras de bolso*, que devem servir como orientação inicial no processo experimental desta escolha. Alguns autores mencionados por

Kaastra e Boyd (1996), utilizam pequenas fórmulas para se fazer este cálculo. Por exemplo, para uma rede neural de três camadas, n neurônios de entrada e m de saída, temos que o número de neurônios utilizados na camada escondida deve ser de $\sqrt{n \times m}$ neurônios. Outros autores, como Baily e Thompson (1990), sugerem que numa rede neural de três camadas, o número de neurônios utilizados na camada escondida deva ser 75% do número de neurônios utilizados na camada de entrada da rede.

Como já exposto anteriormente, este é um processo totalmente experimental que depende do tipo de problema tratado e da base de dados utilizada. Neste sentido, Kaastra e Boyd (1996) destacam dois métodos para se decidir qual o número de neurônios que devem ser utilizados na camada escondida. Ambos os métodos lidam com mudanças do número de neurônios na camada escondida durante o treinamento da rede neural. O primeiro método, chamado de *construtivo*, consiste em adicionar neurônios na camada escondida até que o erro comece a crescer piorando a performance da rede neural. Ao contrário da primeira metodologia, o método *destrutivo*, consiste em retirar neurônios na camada escondida até que o erro também cresça.

Embora estas metodologias apresentadas possuam vantagens, certos *softwares* disponíveis no mercado para redes neurais, não nos permitem adicionar ou retirar neurônios durante o treinamento da rede neural. Este é o caso do *software* utilizado neste trabalho. Contudo, esta impossibilidade não comprometeu a estimação realizada, apenas aumentou seu custo computacional.

5.3 Treinamento da rede neural

A respeito de qual o número ótimo de iterações utilizadas para o treinamento da rede neural, temos duas abordagens, como exposto pelo trabalho de Kaastra e Boyd (1996). A primeira abordagem chama a atenção para a dificuldade e perigo da rede neural não encontrar

seu mínimo global, justificando que o treinamento somente deve ser interrompido quando não houver mais nenhum decréscimo do erro. A segunda abordagem adota uma série de interrupções na fase de treinamento da rede neural. Nesta abordagem, o treinamento é avaliado após um número predeterminado de iterações e a rede neural é testada sem se preocupar unicamente com a convergência do erro. Como o objetivo da fase de treinamento da rede neural é encontrar-se o mínimo global, na segunda abordagem, o pesquisador não tem como saber se um treinamento adicional pode aumentar a capacidade de generalização da rede neural, especialmente se levarmos em consideração que os pesos sinápticos, a cada processo de treinamento, se iniciam por uma seleção aleatória. Logo, na prática, o pesquisador enfrenta muitos *trade-offs*, entre escolher qual a melhor rede neural a se utilizar e o respectivo custo computacional para implementá-la.

6 ANÁLISE DE RESULTADOS

Este trabalho se destinou a fazer inferências sobre o comportamento dos *ratings* de risco soberano a partir dos fundamentos macroeconômicos. Desta forma, neste capítulo, detalhamos inicialmente os aspectos técnicos da aplicação dos procedimentos metodológicos sobre os dados empíricos e, logo após, procuramos, à luz da teoria econômica, discutir estes resultados.

Nossas análises foram feitas a partir de variações do modelo de redes neurais artificiais *Multi Layer-Perceptron* (MLP). O motivo pelo qual foi utilizado tal procedimento justifica-se pelo tipo de problemática, neste caso uma classificação das atribuições de *ratings* de risco soberano, onde problemas envolvendo classificação constituem a principal aplicação da MLP.

Aplicamos no processo de treinamento da rede neural a aprendizagem do tipo supervisionada. Testamos diversas redes neurais e, de acordo com o caráter experimental das estimações por este procedimento metodológico, começamos com uma arquitetura de rede neural composta por onze neurônios na primeira camada, do que iniciamos o experimento com onze variáveis macroeconômicas que compunham o *input* da rede. A saída ou *output* da rede neural foi composto inicialmente por 18 neurônios, pois na amostra de *ratings* da *Standard & Poor's* utilizada nas estimações, tínhamos 18 escalas de *ratings*.

Especificamente, testamos arquiteturas de redes neurais do tipo *feedforward* e *cascade-forward*⁷. Foi observado que as arquiteturas *feedforward* obtiveram desempenho ligeiramente superior ao desempenho das arquiteturas *cascade-forward*, assim optamos por definirmos os resultados finais a partir da primeira.

De acordo com o princípio da parcimônia, para encontrarmos uma rede neural robusta e evitarmos problemas como, por exemplo, a presença de *overfitting*, construímos inicialmente uma rede neural com apenas uma camada escondida e com 10 neurônios nesta camada,

⁷ Ver mais sobre arquiteturas do tipo *cascade* em Arbib (2003) p. 648.

aumentando gradativamente o número de neurônios de acordo com os testes de performance da rede. Além disso, testamos também redes com mais de uma camada escondida o que não aumentou a performance da rede. A rede neural utilizada nas estimações do trabalho foi uma rede com uma camada escondida composta por 40 neurônios, uma camada de entrada composta por 4 neurônios e uma camada de saída composta por 8 neurônios.

Das 18 escalas de *ratings* trabalhadas inicialmente na rede neural, optamos por deixar apenas 8, pois, observamos que, com este número de escalas, a performance da rede foi mais robusta. A simbologia escolhida, por ordem crescente de risco, da escala de *ratings* utilizada na rede foi: A, AA, AAA, B, BB, BBB, CC e CCC.

Foi realizada uma análise de sensibilidade da rede aos *inputs* que acrescentaram pouca informação ao modelo, análise esta presente nos modelos econométricos como o método *stepwise*⁸. Desta forma, as variáveis escolhidas para serem utilizadas na estimação foram:

- Nível de reservas totais como percentual do produto;
- Dívida externa total como percentual das exportações de bens e serviços;
- Balança externa de bens e serviços como percentual do produto;
- Taxa de crescimento do produto *per capita*.

O tipo de função de aprendizado constitui também uma peça importante da rede neural e a função utilizada na estimação deste trabalho foi a aprendizagem chamada *backpropagation* com uma constante de *momentum*, descrita na seção de procedimentos metodológicos.

Testamos diversos tipos de algoritmos de treinamento, onde os mais importantes foram Levenberg-Marquardt⁹, *backpropagation* com uma constante de *momentum* e *Resilient-Propagation*¹⁰. O tipo de treinamento de melhor performance seja pelo critério de melhor

⁸ Este método consiste em reduzir a necessidade de computação intensiva para selecionar o melhor subconjunto de regressores, adicionando ou reduzindo algum regressor através de critérios estatísticos. Ver mais em Greene (1997).

⁹ Descrito pela equação (63).

¹⁰ Descrito pela equação (67).

predição ou por possuir menor custo computacional, foi o *Resilient-Propagation* o que nos levou a utilizá-lo na estimação do trabalho.

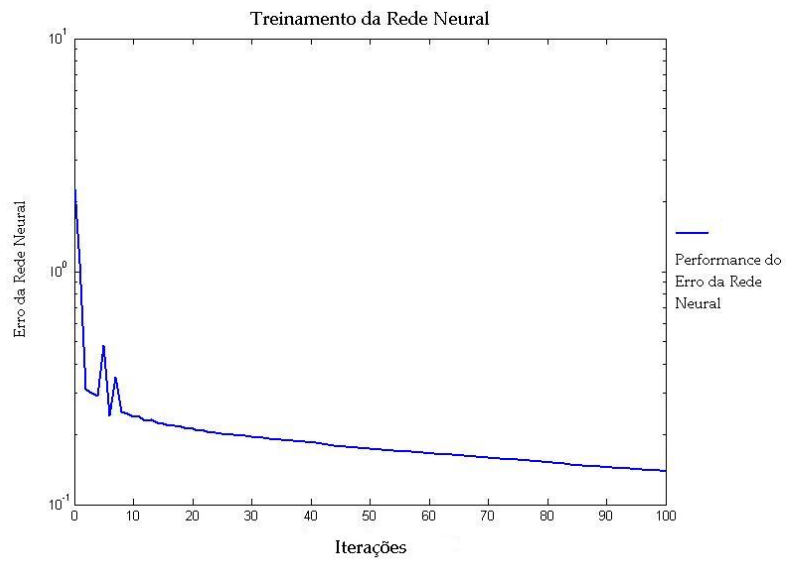
A mensuração da performance da rede neural foi realizada pela função *msereg* definida pela equação (69) que permitiu a regularização da performance ótima da rede. O valor para o parâmetro γ escolhido onde a rede neural obteve uma performance satisfatória foi $\gamma = 1,0$.

Após testes, definimos a função de transferência ou ativação entre os neurônios da rede neural como uma função do tipo tangente hiperbólica definida pela equação (6).

Adicionalmente, testamos qual o número ótimo de iterações a partir da performance do erro-quadrado-médio apresentado pelas diferentes redes neurais testadas e, de acordo com esta performance, escolhemos não mais que cerca de cem iterações necessárias, pois nota-se claramente que a partir de cinquenta iterações o erro-quadrado-médio converge para um valor mínimo.

As Figuras 5 e 6 ilustram, respectivamente, as performances do treinamento e teste realizados na rede neural escolhida. Na Figura 5, em azul encontra-se a performance de treinamento de rede, onde podemos observar que seu erro atinge um nível satisfatório em torno de $0,13944$ na centésima iteração realizada pelo treinamento.

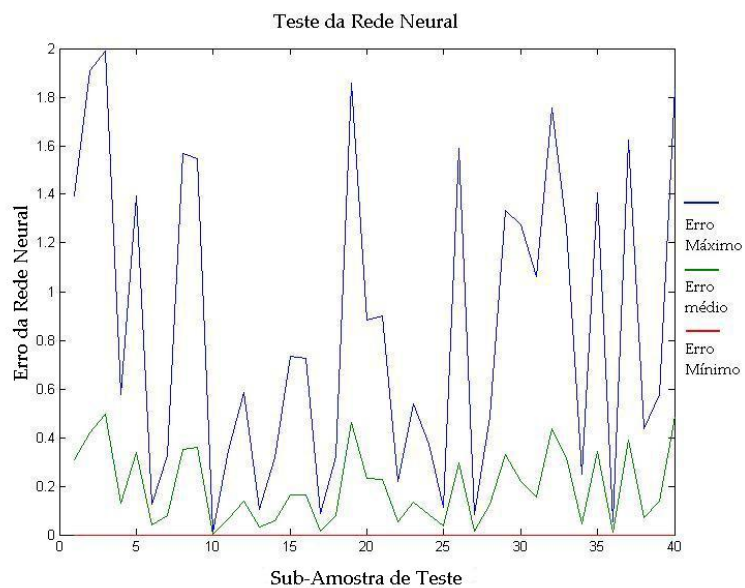
Figura 5 – Performance do treinamento da rede neural escolhida



Fonte: Elaboração própria a partir de dados da *Stancard e Poor's e World Bank*.

Na Figura 6, em verde temos a média de erros da rede, em azul temos seu erro máximo e em vermelho seu erro mínimo.

Figura 6 – Performance do teste da rede neural escolhida



Fonte: Elaboração própria a partir de dados da *Stancard e Poor's e World Bank*.

O erro de predição da rede neural ilustrado pelos pontos mais elevados das linhas de erro, constituem a falha da rede neural em predizer, dados os fundamentos macroeconômicos apresentados à ela, qual deveria ser a atribuição de *rating*. Assim, observamos que a rede consegue com certa robustez predizer as atribuições de *ratings* a partir dos 4 tipos de variáveis macroeconômicas apresentados a ela.

Teoricamente, de acordo com Kaastra e Boyd (1996), uma rede neural com uma camada escondida consegue aproximar qualquer tipo de função. Contudo, ao aplicarmos redes neurais dos tipos descritas anteriormente neste capítulo, aos determinantes quantitativos empíricos dos *ratings* de risco soberano de longo prazo, observamos certo grau de dificuldade que nos levou a diminuir o número de escalas de *ratings* da saída da rede. Embora este problema não tenha prejudicado significativamente a análise deste trabalho ele surgiu por que a rede escolhida não conseguiu, a partir dos dados macroeconômicos, fazer distinção entre 18 escalas diferentes de *ratings*.

Possíveis explicações surgiram para tal problema. Como apontado por diversos estudos empíricos citados ao longo do trabalho, por exemplo, Partnoy (2002), Ferri, Liu e Stiglitz (1999), Bone (2004), Cantor e Packer (1996), Canuto, Santos e Porto (2004), mesmo para os determinantes quantitativos, é difícil encontrar uma relação direta entre estes e os *ratings* atribuídos pela *Standard & Poor's*, por causa do grande número de critérios adotados por esta agência. Pois eles são construídos com base em análises das conjunturais sobre a vida econômica, social e política dos países e, por este motivo, podem ser subjetivos porque envolvem julgamentos sobre uma grande quantidade de variáveis.

Além disso, estes trabalhos tratam de inúmeros problemas envolvendo os *ratings* como problemas de ordem legal dos mercados financeiros como discutido no trabalho de Partnoy (2002), ou as antecipações das mudanças das variáveis econômicas às mudanças nas atribuições dos *ratings* como discutido neste mesmo trabalho e nos trabalhos de Bone (2004) e Ferri, Liu e

Stiglitz (1999). Este último trabalho, defende a idéia da existência de uma regra procíclica das agências de *rating* prolongando os efeitos de crises sobre as economias através de comportamentos excessivamente conservadores caracterizados por atribuírem *ratings* como remédio às crises financeiras. Isto é, quando as agências não conseguem antever uma crise sinalizando ao mercado o perigo desta, elas atribuem *ratings* na tentativa de remediá-la, porém o que elas conseguem é prolongá-la, pois os novos *ratings* atribuídos fazem com que os investidores componham uma carteira de investimentos de menor risco, vendendo os títulos dos países envolvidos nesta crise, provocando uma perda de liquidez destes fazendo com que esta se prolongue.

Estas possíveis explicações surgem em trabalhos elaborados dentro deste tipo de temática, onde a aplicação de diversos tipos de modelos quantitativos refinados a determinantes quantitativos empíricos dos *ratings* de risco soberano freqüentemente não encontram resultados com precisão apurada.

Mesmo com todos os *trade-offs* encontrados por um pesquisador que utiliza redes neurais somados aos problemas envolvendo a questão dos *ratings*, conseguiu-se chegar a um tipo de rede neural que generaliza relativamente bem 8 escalas de *ratings* a partir de 4 variáveis macroeconômicas como mencionado anteriormente.

A partir da escolha da rede neural de melhor poder de generalização, simulamos *cenários* para a economia brasileira, a partir de mudanças nos comportamentos dos fundamentos macroeconômicos que compõem os *inputs* da rede neural, nos quais o risco soberano nestes casos diminui atingindo um patamar menor saindo da escala de risco *B* para *BB*. Mais precisamente, tomamos as quatro variáveis macroeconômicas citadas anteriormente para escolha da rede neural e simulamos mudanças nestas variáveis para o ano de 2003¹¹.

¹¹Diante da indisponibilidade de dados macroeconômicos dos anos 2004 e 2005 pelo Banco Mundial, tomamos o ano de 2003 para fazermos nossas análises.

Estas simulações consistiram na geração aleatória pela rede neural escolhida, com distribuição uniforme, de 1.000 *cenários* que gerariam um *rating* de risco igual a *BB* a partir dos valores observados das 4 variáveis macroeconômicas no ano de 2003.

Inicialmente, simulamos modificações nas observações de apenas uma variável macroeconômica de cada vez. Contudo, observamos que o *rating* de risco se manteve inelástico a esta modificação. Isto pode demonstrar que, para que haja uma nova atribuição de *rating* de risco soberano, seja necessário um conjunto de modificações nos fundamentos macroeconômicos e não apenas a modificação de apenas um destes fundamentos.

Diante disso, a partir dos *cenários* simulados, obtivemos uma média de possíveis realidades macroeconômicas para a economia brasileira em 2003. A Tabela 2 resume estes tipos de *cenários* simulados, onde na primeira coluna de dados, temos os valores observados das variáveis macroeconômicas em 2003 e, nas demais colunas, temos os tipos de *cenários*.

Tabela 2 – Simulações de *cenários* macroeconômicos

Variáveis (em %)	Observações/2003	<i>Cenário</i> 1	<i>Cenário</i> 2	<i>Cenário</i> 3	<i>Cenário</i> 4
Balança externa / Produto	6	6,4	7,6	7,6	7,6
Dívida externa total / Exportações	265	235	235	235	265
Crescimento do produto <i>per capita</i> (%)	-1	2	1	2	2
Nível de reservas totais / Produto	9	15,75	15,75	13,75	15,75

Fonte: Elaboração própria a partir de dados da *Stancard* e *Poor's* e *World Bank*.

A partir das simulações destes 4 diferentes *cenários*, percebemos inicialmente que, para o Brasil alcançar uma nova atribuição de *rating* de risco, as 4 variáveis econômicas que fazem parte dos fundamentos macroeconômicos dos *ratings* devem se alterar.

No *cenário 1*, temos que a balança externa de bens e serviços como percentual do produto, aumenta aproximadamente 6,7% em relação a 2003, enquanto a dívida externa total como percentual das exportações de bens e serviços diminui aproximadamente 11,32%. Já a taxa de crescimento do produto *per capita* se eleva para 2% enquanto a razão do nível de reservas totais como percentual do produto aumenta 75%.

O *cenário 2* é composto da seguinte forma: a balança externa de bens e serviços como percentual do produto, aumenta aproximadamente 26,7% em relação a 2003, enquanto a dívida externa total como percentual das exportações de bens e serviços diminui, como no *cenário 1*, aproximadamente 11,32%. Temos que a taxa de crescimento do produto *per capita* se eleva para 1% enquanto, da mesma forma que no *cenário 1*, a razão do nível de reservas totais como percentual do produto aumenta 75%.

Já no *cenário 3*, a balança externa de bens e serviços como percentual do produto, aumenta, como no *cenário 2*, aproximadamente 26,7% em relação a 2003, ao passo que a dívida externa total como percentual das exportações de bens e serviços também diminui aproximadamente 11,32%. A taxa de crescimento do produto *per capita* se eleva para 2% enquanto a razão do nível de reservas totais como percentual do produto aumenta 52,78%.

E, finalizando a análise descritiva dos *cenários*, no *cenário 4*, temos que a balança externa de bens e serviços como percentual do produto também aumenta aproximadamente 26,7% em relação a 2003, enquanto a dívida externa total como percentual das exportações de bens e serviços se mantém constante. Já a taxa de crescimento do produto *per capita* também se eleva para 2% enquanto a razão do nível de reservas totais como percentual do produto aumenta 75%.

Observamos que, para que haja uma mudança da atribuição de *rating* de risco para a economia brasileira, há uma combinação das mudanças dos fundamentos macroeconômicos. Desta forma, o crescimento do produto *per capita* e a razão do nível de reservas totais como

percentual do produto se mostram variáveis importantes, pois em todos os *cenários*, estas se elevam mais acentuadamente.

Com exceção do *cenário* 4, onde todas as variáveis que se relacionam diretamente com a escala de *rating* de risco soberano, aumentam mais acentuadamente, a dívida externa total como percentual das exportações de bens e serviços diminui. Logo, esta variável também se mostra importante para que as condições de captação de recursos pelo Brasil nos mercados internacionais melhorem.

O aumento da razão do nível de reservas totais como percentual do produto também se mostra sempre crescente de acordo com a melhora de atribuição de *rating*. Observamos que o aumento do nível de reservas é importante para melhorar as condições de crédito no mercado internacional.

Uma das características dos fundamentos macroeconômicos de todos os *cenários* simulados, é que todos são possíveis de serem alcançados no curto prazo através de instrumentos de políticas econômicas. Contudo, não faz parte dos objetivos do presente trabalho discutir quais as políticas econômicas ou quais os instrumentos destas, devem ser utilizados.

De acordo com Mankiw (2004), questões do tipo *como aumentar a taxa do produto da economia*, ou *até que ponto a dívida do governo pode se transformar em um grave problema*, fazem parte de importantes questões onde os macroeconomistas freqüentemente discordam. Assim, baseados na amostra escolhida e nos resultados obtidos, verificamos quais as variáveis que exercem forte interferência no processo de atribuição de *rating* soberano. E, no caso do Brasil, apresentamos possíveis modificações dos fundamentos macroeconômicos que poderiam levar à melhor atribuição de *rating* risco soberano.

7 CONCLUSÃO

Os problemas em torno da assimetria de informação na economia, e suas conseqüências diretas sobre as vidas das pessoas, podem ser mensurados pela quantidade e qualidade dos estudos produzidos em torno deste tipo de temática.

Diante do desenvolvimento e maior integração entre os mercados financeiros mundiais, instituições e dispositivos que têm por objetivo minimizar os transtornos gerados pela assimetria de informações são de muita importância. Desta forma, a magnitude da importância dos *ratings* na economia, como síntese clara e simples de informações sobre as futuras condições de pagamento de um emitente de dívida, também é incomensurável. Haja vista que as condições de captação de recursos no mercado estão fortemente atreladas a eles como foi discutido por Partnoy (2002).

Uma determinada atribuição de *rating* para um emitente de dívida soberano como um país, por exemplo, pode levá-lo a conseqüências permanentes e desastrosas ou levá-lo a uma marcha rumo ao desenvolvimento sócio-econômico via aumento da sua taxa de investimentos. No caso de países emergentes como o Brasil, que têm ampla necessidade de captação de recursos no mercado externo para concretizar projetos de investimentos, sejam privados ou estatais, esta atribuição pode conter um peso substancialmente maior, pois eleva o custo de captação destes recursos. Isso pode acontecer porque dada uma atribuição de *rating* de risco soberano elevada, os dirigentes da economia brasileira serão obrigados a aumentar o prêmio de risco pago por eles, isto é, a taxa de juros paga pelos papéis e títulos financeiros emitidos pelo governo brasileiro.

Em outras palavras, o governo será obrigado, dada uma elevação do risco de crédito concedido para o Brasil no mercado internacional, a aumentar o *spread* de prêmio pago entre

títulos do governo brasileiro e os títulos do governo norte-americano, causalidade esta apresentada pelo trabalho de Cantor e Packer (1996).

Mesmo levando em consideração características das agências de *ratings* que as justificam como fornecedoras deste tipo de serviço na economia, críticas tanto do meio acadêmico quanto do mercado, têm surgido no sentido de atacar sua capacidade preditiva frente a eminências de crises financeiras. Logo, estudar-se uma ferramenta objetiva de apoio à decisão em torno deste tipo de problema se torna fundamental.

Assim, a utilização de redes neurais artificiais no aprendizado da *regra* que rege o processo de atribuição de *rating* de risco soberano objetivou elucidar um processo decisório que, como levantado por Ferri, Liu e Stiglitz (1999), funciona de maneira vaga. Além disso, utilizamos esta *regra* na construção de cenários para a economia brasileira a fim de sabermos como se modificariam as variáveis macroeconômicas determinantes destas atribuições.

Vimos que 4 variáveis macroeconômicas estão mais diretamente conectadas a estas atribuições. São elas: nível de reservas totais como percentual do produto, dívida externa total como percentual das exportações de bens e serviços, balança externa de bens e serviços como percentual do produto e taxa de crescimento do produto *per capita*.

Das 18 escalas diferentes de *ratings* que inicialmente faziam parte do *output* da rede neural, conseguiu-se através de uma arquitetura do tipo *feedforward*, com uma camada escondida composta por 40 neurônios, chegar-se a 8 escalas diferentes de *ratings*. Isto se deu em parte à dificuldade de implementação de métodos quantitativos à problemática envolvendo *ratings* de risco soberano como amplamente relatado na literatura e, em parte, a maior refinamento da rede neural. Embora, como foi descrito no capítulo 6, foram testadas diversas variações de composições de arquiteturas de redes.

Os *cenários* construídos nos mostraram que combinações de modificações dos fundamentos macroeconômicos foram necessárias para que o Brasil saltasse da escala de

atribuição de *rating B* para *BB*. E, além disso, observamos que o sentido das modificações das variáveis foi de acordo com a teoria econômica e que a magnitude destas modificações podem ser alcançadas no curto prazo.

Este estudo é precursor ao aplicar este tipo de procedimento metodológico ao problema estudado. Dado o processo experimental de construção de uma arquitetura de rede neural adequada, podemos concluir que a busca de maior refinamento deste, se torna uma preocupação com a qual o pesquisador que o aplica terá sempre que lidar. Devemos ressaltar que o melhor modelo de redes neurais será sempre tangenciado.

Logo, o trabalho buscou alcançar seu principal propósito no sentido de auxiliar os formuladores de políticas econômicas que têm por objeto melhorar as condições de crédito no mercado internacional, a definir políticas com mais subsídio informacional, pois a comprovação científica é uma importante informação para processos de decisão.

REFERÊNCIAS

- AMBRÓSIO, Paulo Eduardo. **Redes neurais artificiais no apoio ao diagnóstico diferencial de lesões intersticiais pulmonares**. Dissertação de Mestrado – Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras de Riberão Preto, 2002.
- AGRAWAL, D.; SCHORLING, C. **Market share forecasting**: an empirical comparison of artificial neural networks and multinomial logic model. In: Journal of Retailing, v. 72, 1996, p. 383–407.
- AIKEN, M.; BSAT, M. **Forecasting market trends with neural networks**. Information Systems Management, 1999, p. 42–48.
- AKOBIR, Shahidi. **Mathematical apparatus of the RPROP learning algorithm**. BaseGroup Labs. Disponível em: <<http://www.basegroup.ru/neural/rprop.en.htm>> Acesso em: jan. 2006.
- ARBIB, Michael A. **The handbook of brain theory and neural networks**. 2 ed. London, England: MIT Press, 1995.
- BAILY, D. ; THOMPSON, D. M. **Developing neural network applications**. Al Expert. set. 1990, p. 33-41.
- BENITE, Maurílio. **Aplicação de modelos de redes neurais na elaboração e análise de cenários macroeconômicos**. Dissertação de Mestrado – Universidade de São Paulo, Escola de Engenharia de São Carlos, 2003.
- BLOCK, H.D. **The Perceptron**: A Model for Brain Functioning, In: I Review of Modern Physics, v. 34, 1962, p. 123-135.
- BONE, Rosemarie Bröker. **Análise crítica dos fundamentos dos ratings soberanos: o caso da Argentina**. Banco Central del Uruguay. Montevideo, Uruguai, 2002. Disponível em: <<http://bcu.gub.uy/autoriza/peiees/jor/2002/iees03j060702.pdf>> Acesso em: out. 2005.
- BONE, Rosemarie Bröker. **Ratings soberanos e corporativos: o rompimento do teto soberano pela Petrobrás e REPSOL-YPF**. Tese de Doutorado. Universidade Federal do Rio de Janeiro. Instituto de Economia. Rio de Janeiro, 2004.
- BONE, Rosemarie Bröker. **A importância dos fundamentos nos ratings soberanos brasileiros, 1994-2002**. In: XXXIII Encontro Nacional de Economia, Dez. 2005, Natal **Anais...** Natal: XXXIII Encontro Nacional de Economia, 2005. CD-ROM.
- BRAGA, Marcus Vinícius Diniz Barros. **Redes neurais aplicadas à previsão de ações da bolsa de valores**. Monografia - Faculdade Integrada do Recife, Recife, 2004.
- BROOMHEAD, D. S.; LOWE, D. **Multivariable functional interpolation and adaptive networks**. Complex Systems. 1988, p. 321-335.

BRONDINO, Nair Cristina Margarido. **Estudo da influência da acessibilidade no valor de lotes urbanos através do uso de redes neurais**. Tese de Doutorado - Universidade de São Paulo, Escola de Engenharia de São Carlos, 1999.

CANTOR, R.; PACKER, F. **Determinants and Impact of Sovereign Credit Ratings**. Economic Policy Review, v. 2, nº 2, October 1996. Disponível em: <www.ny.frb.org/rmaghome/eco_pol/1096cant.html>. Acesso em: out. 2005.

CANUTO, O.; SANTOS, P. F. dos; PORTO, Paulo C. de Sá. **Macroeconomics and sovereign risk ratings**. World Bank Discussion papers. Washington D. C., Jan. 2004.

CANUTO, O.; SANTOS, P. F. dos. **Risco-soberano e prêmios de risco em economias emergentes**. Ministério da Fazenda, Secretaria de Assuntos Internacionais, Temas de economia internacional, Jan. 2003b.

CARDOSO, Ademar de Azevedo. **Redes neurais artificiais na avaliação de concentração de tensões em juntas tubulares soldadas**. Tese de Doutorado - Universidade de São Paulo, Escola Politécnica, 1999.

CARPENTER, G. A.; GROSSBERG, S. **A massively parallel architecture for a self-organizing neural pattern recognition machine**. Computer Vision, Graphics, and Image Processing, 1987a.

CARPENTER, G. A.; GROSSBERG, S. **Adaptive resonance theory**. Arbib M. ed., Handbook of brain theory and neural networks. London, England: MIT Press, 1995.

CASTRO JR., F. F. **Previsão de Insolvência de Empresas Brasileiras Usando Análise Discriminate, Regressão Logística e Redes Neurais**. Dissertação de Mestrado – Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade, São Paulo, 2003.

CHEN, X.; RACINE, J.; SWANSON, N. R. **Semiparametric ARX neural-network models with an application to forecasting inflation**. In: IEEE Transactions on Neural Networks, 2001, p. 674–683.

CHUA; YANG. **Cellular neural networks: theory**. In: IEEE transactions on systems and circuits, 1988.

CHURCH, K. B.; CURRAM, S. P. **Forecasting consumers' expenditure: a comparison between econometric and neural network models**. In: International Journal of Forecasting, 12, 1996, p. 255–267.

CORREIA, Leonardo Garcia. **Memória associativa em redes neurais realimentadas**. Dissertação de Mestrado – Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, 2004.

CYBENKO, G. **Approximation by superpositions of a sigmoid function**. Mathematics of Control Signals and Systems, 1989, p. 303–314.

DEMUTH, H.; BEALE, M. **Neural network toolbox**. For use with MATLAB. Computation, visualization, programming. User's guide. Version 3. The Math Works. 1998. Disponível em: <<http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/>> Acesso em: jan. 2006.

FERRI, G.; LIU, L-G.; STIGLITZ, J. E. **The procyclical role of rating agencies**: evidence from the East Asian crisis. *Economic Notes*, v. 28, n. 3, 1999, p. 335-355.

GATELY, E. **Neural Network for Finance Forecasting**. New York: John Wiley & Sons, 1996.

GREENE, William H. **Econometric Analysis**. 5th ed. Upper Saddle River, New Jersey: Prentice Hall, 2002.

HANGAN, M. T.; MENJAH, M. B. **Training feedforward network with hte marquardt algorithm**. In: *IEEE Trans. on Neural Networks*, 1994, p. 989–993.

HAQUE, Nadeem U.; MARK, Nelson; MATHIESON, Donald J. **The relative importance of political and economic variable in creditworthiness ratings**. IMF: working paper 98/46. Abr. 1998.

HAYKIN, S. **Redes Neurais: Princípios e Práticas**. 2 ed. São Paulo: Bookman, 2001.

HEBB, D. O. **The organization of behaviour**. New York: Wiley, 1949, p. 18, 121.

HOPFIELD, J. J. **Neural networks and physical systems with emergent collective computational abilities**. *Proceedings of the National Academy of Sciences*, 1982.

JIANG, J. J.; ZHONG, M.; KLEIN, G. **Marketing category forecasting**: An alternative of BVAR—Artificial neural networks. *Decision Sciences*, 2000, p. 789–812.

KAASTRA, Iebling; BOYD, Milton. **Designing a neural network for forecasting financial and economic time series**. *Neurocomputing*. v. 10, n. 3, jan. 1996.

KOHONEN, T. **Self-organizing maps**. Springer, 1995.

KOHZADI, N. et al. **A comparison of artificial neural network and time series models for forecasting commodity prices**. In: *Neurocomputing*, v. 10, 1996, p. 169–181.

KRÖSE, Ben; SMAGT, Patrick van der. **An Introduction to Neural Networks**. Universidade de Amsterdã. 8 ed., Amsterdã, nov. 1996.

KUMAR, A.; RAO, V. R.; SONI, H. **An empirical comparison of neural network and logistic regression models**. *Marketing Letters*, 1995.

LEICHSENRING, Daniel. **Endogeneidade e mecanismos de transmissão entre a taxa de juros doméstica e o risco soberano**: uma revisita aos determinantes do Risco-Brasil. Dissertação de Mestrado – Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade, Departamento de Economia, São Paulo, 2004.

LEUNG, M. T.; CHEN, A. S.; DAOUK, H. **Forecasting exchange rates using general regression neural networks**. *Computers and Operations Research*, 2000.

MANKIW, N. Gregory. **Macroeconomia**. 5 ed. Tradução por A. B. Pinheiro de Lemos. Tradução de Macroeconomics. Rio de Janeiro: LTC , 2004.

MAS-COLLEL, A.; WHINSTON, M D.; GREEN, J. R. **Microeconomics theory**. New York: Oxford University Press, 1995.

McCULLOCH, W.; PITTS, W. **A Logical Calculus of the Ideas Imminent in Nervous Activity**. In: Bulletin of Mathematical Biophysics, v. 5, 1943.

MEGALE, Caio. **Fatores Externos e o Risco País**. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. Rio de Janeiro, 2003.

MINSKY, M.; PAPERT, S. **Perceptrons: An Introduction to Computational Geometry**. London, England: MIT Press, 1969.

MIROLLO, R.; STROGATZ, S. **Synchronization of pulse-coupled biological oscillators**. SIAM J. Appl. Math, 1990.

MOODY'S INVESTORS SERVICE. **Moody's rating symbols & definitions**. Ago. 2003. Disponível em: <www.moodys.com>. Acesso em: set. 2005.

NAG, A. K.; MITRA, A. **Forecasting Daily Foreign Exchange Rates Using Genetically Optimized Neural Networks**. Journal of Forecasting, 2002, p. 501–512.

NG, Andrew. **The perceptron and large margin classifiers**. Lectures notes. Stanford University. Califórnia, EUA, 2004. Disponível em: <<http://www.stanford.edu/class/cs229/notes/cs229-notes6.pdf>> Acesso em: set. 2005.

NOVIKOFF, A. B. J. **On Convergence Proofs of Perceptrons**. Proc. Symp. on Math. Theory of Automata. Polytechnic Institute of Brooklyn. Brooklyn, NY, 1962, p. 615-622.

PARTNOY, F. **The paradox of credit ratings**. In: Levich; Majnoni e Reinhart, Ratings, Rating agencies and the global financial system. Boston – Kluwer, 2002.

PASSARI, A. F. **Exploração de Dados Atomizados para Previsão de Vendas no Varejo Utilizando Redes Neurais**. Dissertação de Mestrado – Universidade de São Paulo, Faculdade de Economia, Administração e Contabilidade, Departamento de Administração, São Paulo, 2003.

PIAZZI, Fábio Silva. **Introdução às redes neurais**. Centro Brasileiro de Pesquisas Físicas. Seminários Internos. Rio de Janeiro: Novembro, 2001. Disponível em: <www.cbpf.br/cat/download/seminarios/internos/Fabio.ppt> Acesso em: set. 2005.

QI, M. **Predicting US recessions with leading indicators via neural network models**. In: International Journal of Forecasting, 2001, p. 383–401.

REINHART, Carmen M. **Sovereign credit ratings before and after financial crises**. University of Maryland and NBER. Working paper. fev. 2001.

RIEDMILLER, M.; BRAUN, H. **A direct adaptive method for faster backpropagation learning**: The RPROP algorithm. In: Proceedings of the IEEE International Conference on Neural Networks, San Francisco, 1993.

ROSENBLATT, F. **Principles of neurodynamics**. New York: Spartan Books, 1959, p. 23-26.

ROWLAND, Peter. **Determinants of spread, credit ratings and creditworthiness for emerging market sovereign debt**: a follow-up study using pooled data analysis. Banco de la Republica de Colombia. Bogotá, Colômbia: 2004. disponível em: <<http://www.banrep.gov.co/docum/ftp/borra296.pdf>> Acesso em: dez. 2005.

RUMELHART, D.; HINTON, G.; WILLIAMS, R. **Learning Internal Representations by Error Propagation**. Rumelhart e McClelland: Parallel Distributed Processing - In: Explorations in the Microstructure of Cognition - v. 1. Foundations. London, England: MIT Press, 1986.

RUMELHART, D. E.; MCCLELLAND, J. L. **Parallel distributed processing**: explorations in the microstructure of cognition. London, England: MIT Press, 1986.

SANTOS, Cláudia Cristina dos. **Modelagem de bacias urbanas com redes neurais artificiais**. Dissertação de Mestrado - Universidade de São Paulo, Escola Politécnica, 2001.

SIMON, Carl P.; BLUME, Lawrence, **Matemática para Economistas**. São Paulo: Bookman, 2004.

SIU, Roychowdhury et al. **Toward massively parallel design of multipliers**. (context), 1995.

SMAGT, Patrick Van der; KRÖSE, Ben. **An introduction to neural networks**. 8 ed. Amsterdam: University of Amsterdam Press, 1996.

STANDARD & POOR'S. **Sovereign Ratings History Since 1975**. New York, USA: nov. 2005.

STANDARD & POOR'S. **Corporate ratings**. criteria. 2002a. Disponível em: <www.standardandpoors.com>. Acesso em: set. 2005.

STANDARD & POOR'S. **Critérios - comentários**: ratings de crédito soberano: principais conceitos. 2004. Disponível em: <www.standardandpoors.com>. Acesso em: set. 2005.

STANDARD & POOR'S. **Risco soberano e ratings acima do soberano**. critérios – comentários. *ratings* de crédito: soberanias. 23 jul. 2001a. Disponível em: <www.santandarandpoors.com.br>. Acesso em: set. 2005.

STIGLITZ, Joseph; GREENWALD, Bruce. **Rumo a um novo paradigma em economia monetária**. Tradução Laura Knapp e Cecília Camargo Bartalotti. São Paulo: Francis, 2004. Tradução de: Towards a new paradigm in monetary economics.

THE 2005 WORLD BANK DEVELOPMENT INDICATORS. Disponível em: <<http://ddp-ext.worldbank.org/ext/DDPQQ/member.do?method=getMembers&userid=1&queryId=19>> Acesso em: out. 2005.

TKACZ, G. **Neural network forecasting of Canadian GDP growth**. *International Journal of Forecasting*, 2001, p. 57–69.

TSENG, F. M.; YU, H. C.; TZENG, G.-H. **Applied hybrid grey model to forecast seasonal time series**. In: *Technological Forecasting and Social Change*, 2001, p. 291–302.

VEELENTURF, L. P. J. **Analysis and applications of artificial neural networks**. United Kingdom: Prentice Hall International, 1995.

VELLIDO, A.; LISBOA, P. J. G.; VAUGHAN, J. **Neural networks in business: a survey of applications (1992–1998)**. *Expert Systems with Applications*, 1999, p. 51–70.

WANG, S. **An adaptive approach to market development forecasting**. In: *Neural Computation & Applications*, 1999, p. 3–8.

WEST, P. M.; BROCKETT, P. L.; GOLDEN, L. L. **A comparative analysis of neural networks and statistical methods for predicting consumer choice**. In: *Marketing Science*, v. 16, 1997, p. 370–391.

WIDROW, B. **Generalization and information storage in networks of adaline neurons**. In: Yovits, M. C., Jacobi, G. T., and Goldstein, G. D., editors, *Self-Organizing Systems*. Chicago, Illinois. 1962, p. 435-461.

YAO, J.; LI, Y.; TAN, C. L. **Option price forecasting using neural networks**. *Omega*, 28, 2000, p. 455–466.

ZHANG, G. P.; HU, M. Y. **Neural network forecasting of the British Pound/US dollar exchange rate**. *Omega*, 1998.