Association for Information Systems

AIS Electronic Library (AISeL)

AMCIS 2020 Proceedings

Select Recordings from Conference

Aug 12th, 12:00 AM

Workshop: Al and Deep Learning Using SAS Viya

Vijayan Sugumaran
Oakland University, sugumara@oakland.edu

James Harroun SAS Institute, james.harroun@sas.com

Follow this and additional works at: https://aisel.aisnet.org/amcis2020

Recommended Citation

Sugumaran, Vijayan and Harroun, James, "Workshop: Al and Deep Learning Using SAS Viya" (2020). *AMCIS 2020 Proceedings*. 8.

https://aisel.aisnet.org/amcis2020/select_recordings_from_conference/select_recordings_from_conference/8

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.





SIGODIS/SIGDSA Workshop: Al & Deep Learning Using SAS Viya

Vijayan Sugumaran, Ph.D.

Chair, Department of Decision and Information Sciences
School of Business Administration
Oakland University

sugumara@oakland.edu

James Harroun

Data Science Initiatives Manager SAS Institute

James.Harroun@sas.com









Agenda

- Welcome and General Introduction/Procedures
- First Hour Machine Learning and SAS Viya for Learners
 - Overview of SAS Viya and SAS Viya for Learners
 - Basics of Neural Networks
 - Demo
 - SAS Visual Data Mining and Machine Learning Neural Network
 - Model Comparison, and Model Deployment
- Break (30 second)
- Second Hour Deep Learning and Open Source Integration
 - Deep Learning and Basic Architectures
 - Overview of Convolutional Neural Network
 - SAS and Open Source Integration
 - Demo
 - Python Jupyter Notebook and Convolutional Neural Network Example
- Academic Resources Available







Source-based **Engines**

In-Stream



In Hadoop

Parallel and Serial, Pub / Sub,

Web Services, MQs

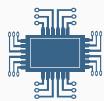


In-Database

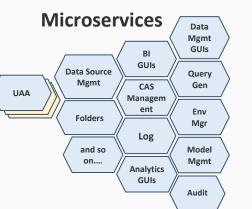


SAS° Viya

In-Memory Run-Time Engine



Cloud Analytics Services (CAS)



Ssas

Customer Intelligence



Analytics

Risk Management

Solutions

APIs



Business Visualization

Fraud and Security Intelligence











































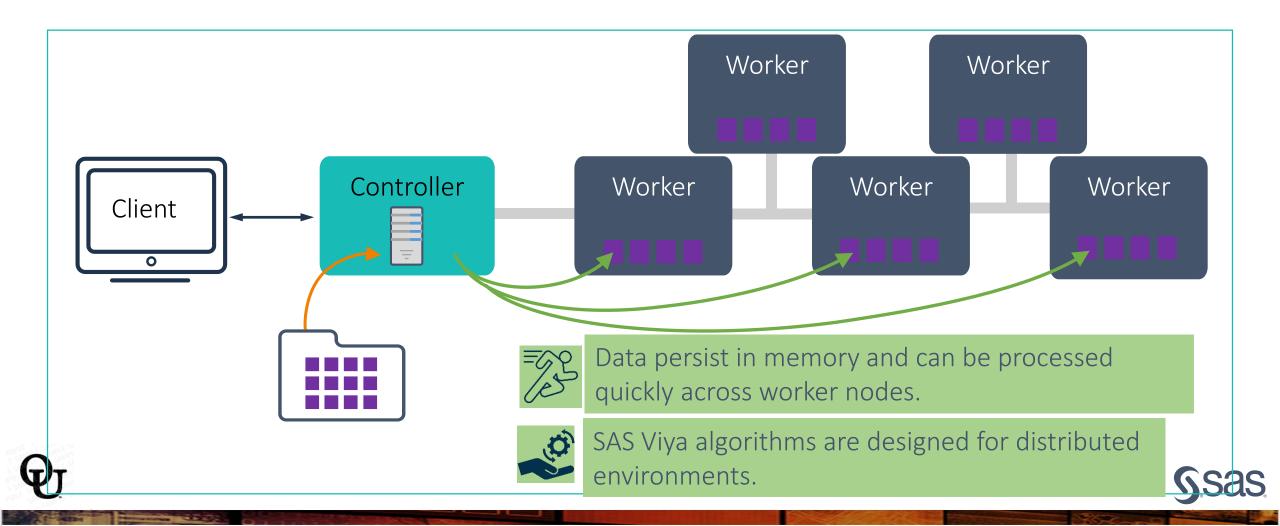








CAS Distributed Environment





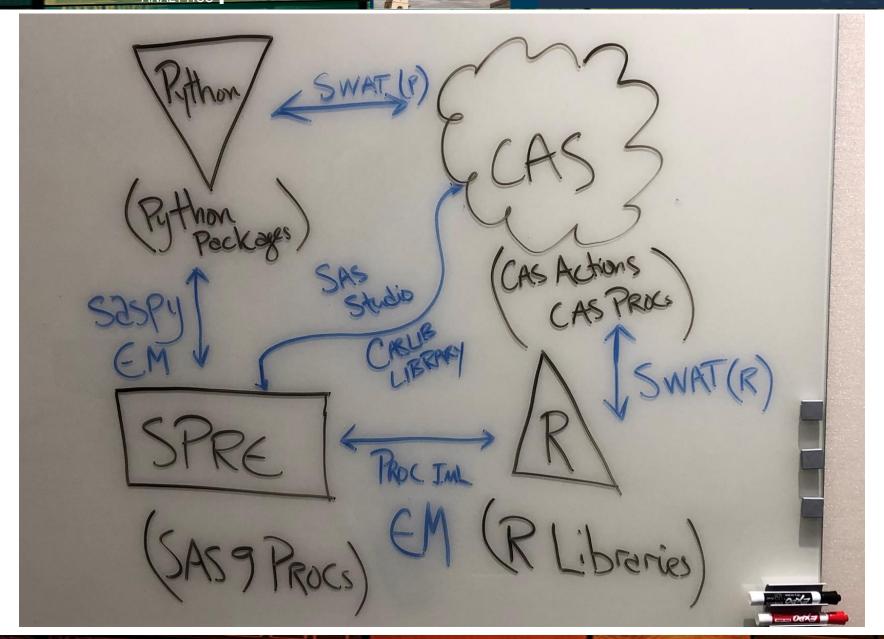


What is SAS® Viya® for Learners (VFL)?

- SAS® Viya® for Learners is a cloud-based offering that provides Academic Institutions and On-Line Learning Platforms access to advanced analytics SAS® Viya® software coupled with Jupyter Notebook to enable coding in their language of choice: R, Python, SAS.
- No downloads.
- No cost.
- For academic, noncommercial use only.
- Educators are provided access to SAS courses and materials for use in their classrooms.







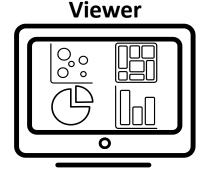


Ssas



SAS Visual Analytics Applications



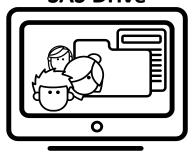


SAS Theme

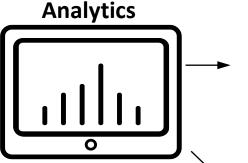
Designer

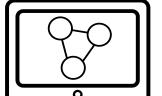
0

SAS Drive



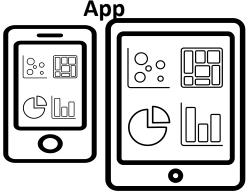
SAS Visual





SAS Visual Statistics

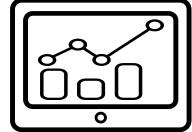




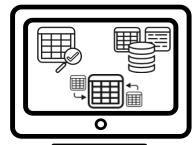


SAS Cloud Analytic

Services (CAS)
SAS Graph Builder











Sas







SAS Visual Data Mining and Machine Learning

Programming Approach



Visual Drag-and-Drop Approach





SAS Studio



SAS Visual Analytics



Model Studio









SAS Viya Functionality



SAS Visual Analytics

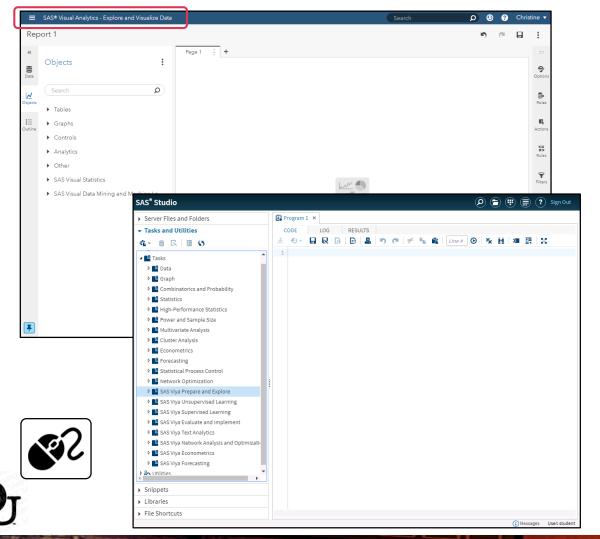
- Explore data and discovery relationships
- Examine distributions and summary statistics
- Perform post-model analysis and reporting
 - SAS Visual Statistics
 - Build unsupervised and supervised models
 - Interactively refine candidate models
 - Compare models and generate score code
 - SAS Visual Data Mining and Machine Learning
 - Six additional machine learning models

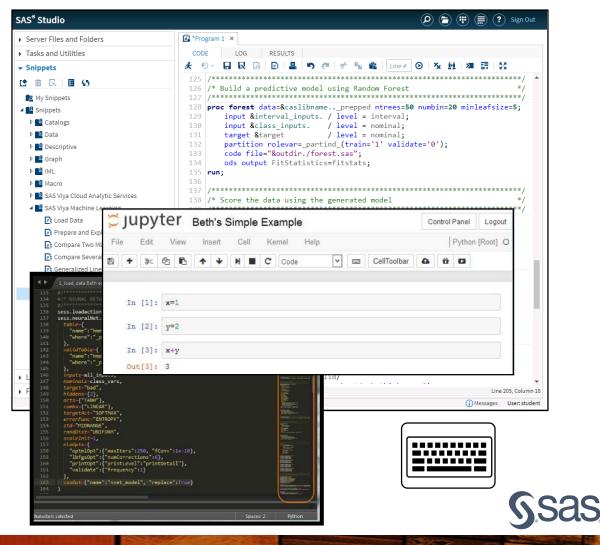




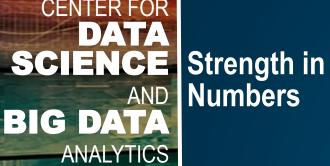


SAS Viya Interfaces Multiple Interfaces, Including Visual and Programmatic













Neural Network Basics



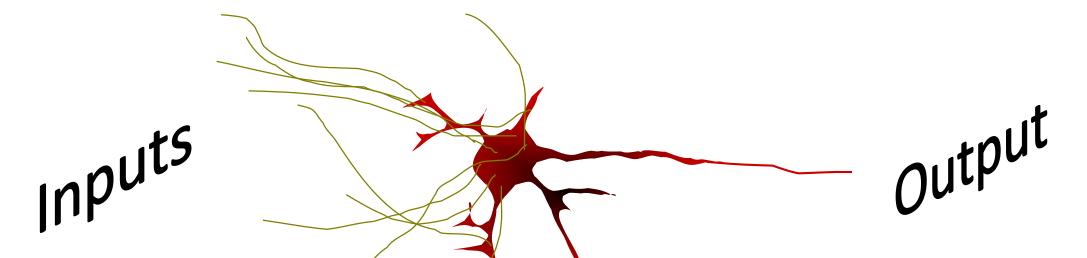








The Biological Neuron



- Inputs arriving through *dendrites* are first weighted by adaptable synapses before being summed.
- If the sum is greater than an adaptable **bias**, the neuron sends a signal down its axon to other neurons.





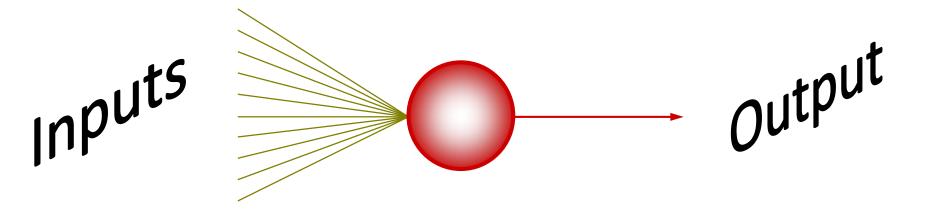




The Mathematical Neuron

- The mathematical neuron has two parts:
 - combination function (pre-activation)
 - activation function

$$H = f \left(w_0 + \sum_{i=1}^d w_i x_i \right)$$
Bias Value



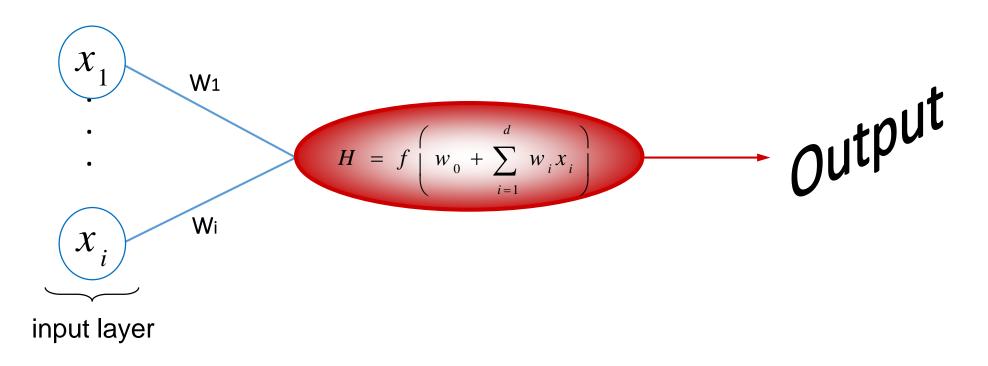






A Single Layer Perceptron

 A single layer perceptron with a linear activation function resembles a generalized linear model.









Perceptron

A simplified representation is,

$$ext{output} = \left\{ egin{array}{ll} 0 & ext{if } w \cdot x + b \leq 0 \ 1 & ext{if } w \cdot x + b > 0 \end{array}
ight.$$

- 1. w and x are vectors whose components are the weights and inputs, respectively
- Moving the threshold to the other side of the inequality in the previous equation and replacing it by perceptron's bias, b≡-threshold

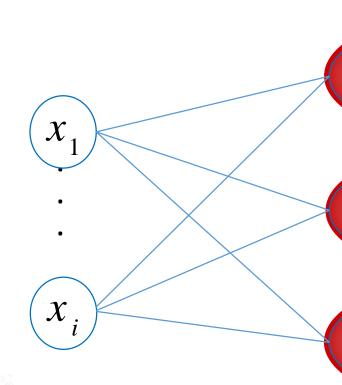






Multilayer Perceptron

$$g^{-1}(\hat{y}) = w_0 + \sum_{i=1}^h w_i f_i \left(w_{0i} + \sum_{j=1}^d w_{ij} x_j \right)$$



$$H = f \left(w_{01} + \sum_{i=1}^{d} w_i x_i \right)$$

$$H = f \left(w_{02} + \sum_{i=1}^{d} w_i x_i \right)$$

$$H = f \left(w_{03} + \sum_{i=1}^{d} w_i x_i \right)$$

$$g^{-1}(\hat{y}) = w_0 + \sum_{i=1}^n w_i H_i$$







Where Are Activation Functions Typically Used?

Answer:

Input-to-hidden unit weights

$$\begin{bmatrix} x_1 \end{bmatrix}$$



input layer

$$H = f \left(w_{01} + \sum_{i=1}^{d} w_i x_i \right)$$

$$H = f \left(w_{02} + \sum_{i=1}^{d} w_{i} x_{i} \right)$$

$$H = f \left(w_{03} + \sum_{i=1}^{d} w_{i} x_{i} \right)$$

hidden layer

$$g^{-1}(\hat{y}) = w_0 + \sum_{i=1}^h w_i H_i$$
output layer

output layer













Activation Functions

	Activation Function	Function	Range	Plot	Typical Usage
	IDENTITY	t	$(-\infty, +\infty)$		Output Layer
	Exponential	exp ^t	$(0,\infty)$		Output Layer
Default	SOFTMAX	$\frac{e^{t}}{\sum exponentials}$	(0,1)	none	Output Layer
Sciadic	TANH (hyperbolic tangent)	$1-\frac{2}{1+e^{(2t)}}$	(-1,1)		Hidden Layer
Фт	ReLU (rectified linear)	$\begin{cases} t & if \ t \ge 0 \\ 0 & if \ t < 0 \end{cases}$	$(0,\infty)$		Hidden Layer
	ELU (exponential linear)	$\begin{cases} t & if \ t > 0 \\ \alpha (\exp(t) - 1) & if \ t \le 0 \end{cases}$	$(-\alpha,\infty)$		Hidden Layer
	SOFTPLUS	$\ln\left(1+e^{t}\right)$	$(0,\infty)$		Hidden Layer
	SINe	$\sin(t)$	(-1,1)		Hidden Layer

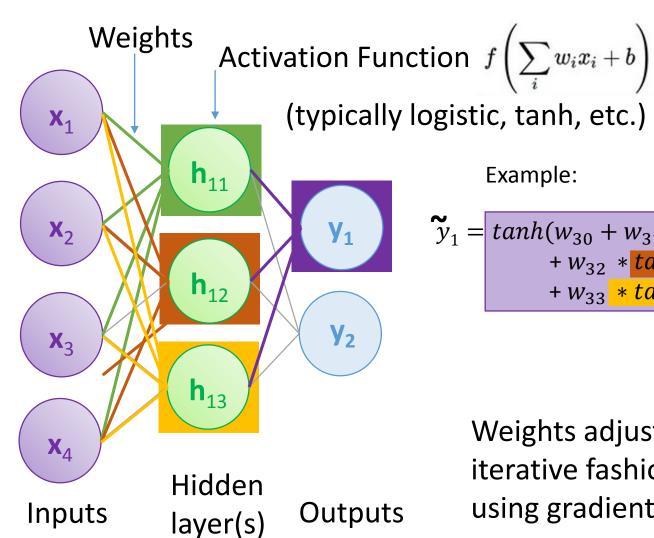








Neural Networks





Example:

$$\mathbf{\tilde{y}}_{1} = tanh(w_{30} + w_{31} * tanh(w_{10} + w_{11}x_{1} + w_{12}x_{2} + w_{13}x_{3} + w_{14}x_{4}) + w_{32} * tanh(w_{20} + w_{21}x_{1} + w_{22}x_{2} + w_{23}x_{3} + w_{24}x_{4}) + w_{33} * tanh(w_{30} + w_{31}x_{1} + w_{32}x_{2} + w_{33}x_{3} + w_{34}x_{4}))$$

$$\tilde{y}_1 \neq y_1$$
?

Weights adjusted to train the network in an iterative fashion through **back-propagation** (ie, using gradient of error with respect to weights) Sas







Assessment Methods for Classification

- Predictive accuracy
 - Hit rate
- Speed
 - Model building versus predicting/usage speed
- Robustness
- Scalability
- **Interpretability**
 - Transparency, explainability









Accuracy of Classification Models

In classification problems, the primary source for accuracy estimation is the confusion matrix

Accuracy: Overall, how often is the classifier correct?

Misclassification: Overall, how often is the classifier wrong?

True Negative Rate: When it is actually negative, how often does it predict negative?

F Score: Harmonic mean of Precision and Recall

True Positive Rate: When it is actually positive, how often does it predict positive?

Precision: When it predicts positive, how often is it correct?

False Positive Rate: When it is actually negative, how often does it predict positive? Prevalence: How often does the positive condition actually occur in the sample?

True/Observed Class					
	,	Positive	Negative		
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)		
Predicte	Negative	False Negative Count (FN)	True Negative Count (TN)		

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

True Positive Rate =
$$\frac{TP}{TP + FN}$$
(Sensitivity, Recall)

True Negative Rate =
$$\frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$
 $Re \ call = \frac{TP}{TP + FP}$

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

True Negative Rate =
$$\frac{TN}{TN + FP}$$
 $FScore = \frac{TP}{TP + \frac{FN + FP}{2}} = \left(2 \times \frac{Precision \times Recall}{Precision + Recall}\right)$

$$Precision = \frac{TP}{TP + FP}$$
 $Re\ call = \frac{TP}{TP + FN}$ $Prevalence = \frac{TP + FN}{TP + TN + FP + FN}$







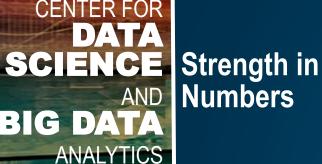
Demo Using SAS Viya for Learners

https://vfe.sas.com/portal/home











Stretch Break (30 Sec)

After the Break:

Moving Beyond Basic Neural Networks: Deep Learning and Open Source Integration (Python API)











Deep Learning vs. Conventional Neural Network

- Differences in methodologies
 - strengths and weaknesses of conventional NN
 - strengths and weaknesses of Deep Learning
- Differences in Applications objectives
 - Neural Network: input and output provided
 - Deep Learning, identifying new features and feature engineering occur within the archetecture
- Various Architectures:
 - CNN
 - RNN
 - LSTM





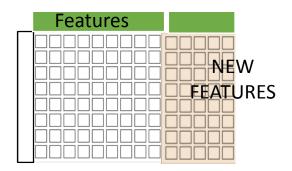






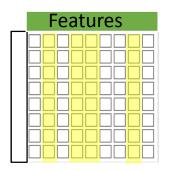
Types of Feature Engineering

Feature Construction



- Cleansing
- Brainstorming
- Aggregation
- Decomposition

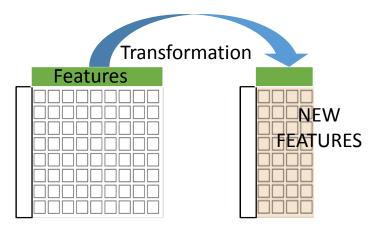
Feature Selection



- MIC, information gain, chisquare...
- Stepwise regression, LASSO, elastic net...
- Decision tree

Consider using several methods and create a pool of features chosen by many

Feature Extraction



- PCA, SVD
- Nonnegative matrix factorization
- Autoencoding neural networks

Loss of interpretability



Simpler models, shorter training times, improved accuracy







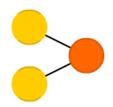


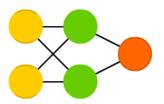


SAS Neural Network Types

Perceptron

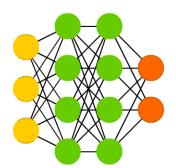
Feed Forward Networks (FF)

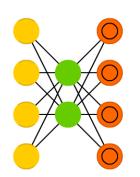


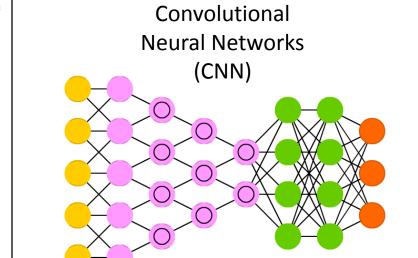


SAS Deep Learning Architecture Types

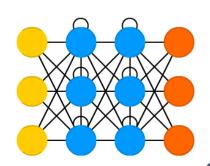
Deep FF Neural Network (DNN) Auto Encoder*
(AE)







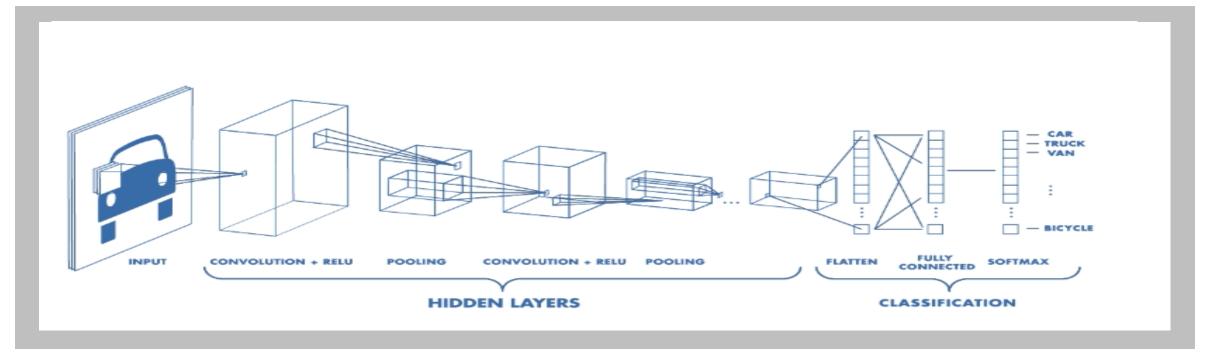
Recurrent Neural Networks (RNN)







Convolutional Neural Networks



Input Image
Pixel values



54 46 2 2 3 47

Fi	Ite	r		D .	
1	1	1		Bias	
0	0	0	+	10	=
-1	-1	-1			

Feature map

44	187	







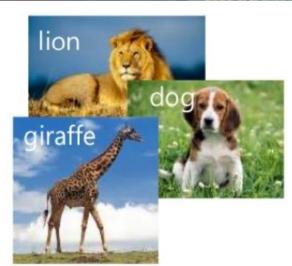


Image Classification

















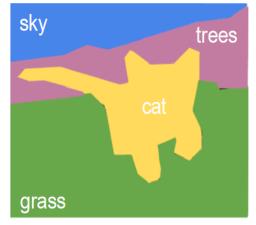


Face Recognition and Key Point Detection

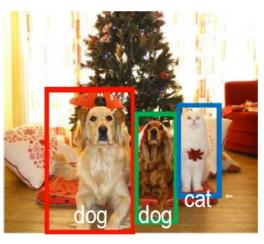
Computer Vision



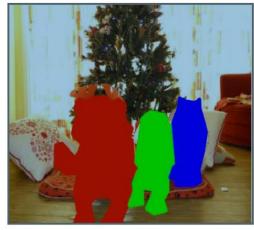
Object Localization



Semantic Segmentation



Object Detection



Object Segmentation



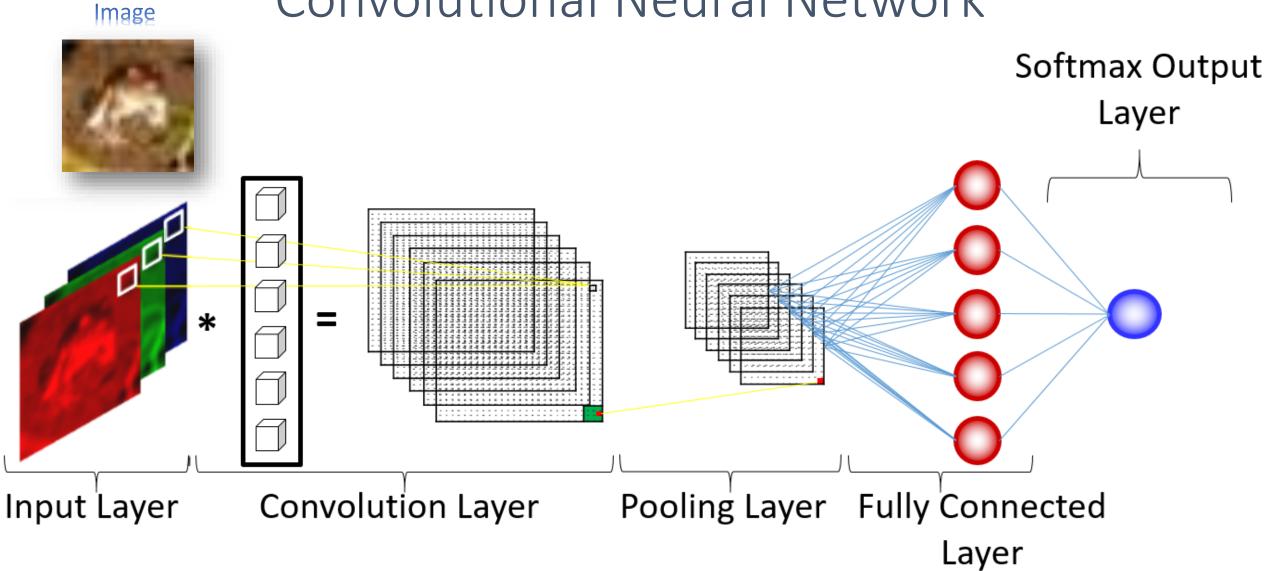








Convolutional Neural Network

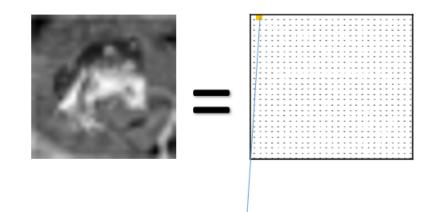






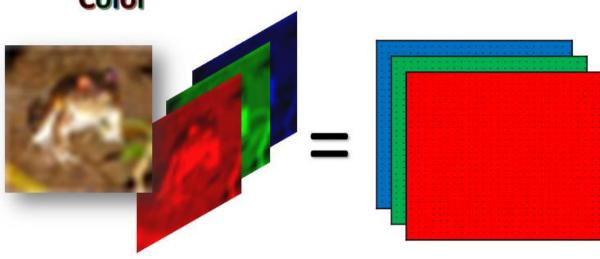
Input Layer

Grayscale



Label	Column 1	Column 2	 Column <i>N</i>
Frog	223	225	 82





Labo	All the second	olum1 C	Service III	1000	imn Column
Fi	Label	Column 1	And the control of th		Column N
	Frog	144	37	****	180









SAS Deep Learning Actions



```
BuildModel / modeltable={name="< Model table name >"}
type = "DNN";
```

Convolutional Neural Network (CNN)

```
BuildModel / modeltable={name="< Model table name >"}
type = "CNN";
```

Recurrent Neural Network (RNN)

```
BuildModel / modeltable={name="< Model table name >"}
type = "RNN";
```











Build a Deep Neural Network

```
PROC CAS < exc > < noqueue >;
    BuildModel / modeltable={name="< model table name >"} type ="DNN";
AddLayer /
   modeltable="< model table name >"
   name="< name of layer >"
   layer={type="layer type"
            n="< number of hidden units >"
            act="< type of activation transformation >"
            init="< weight initialization method >"
   srcLayers={"< previous layer name >"};
```

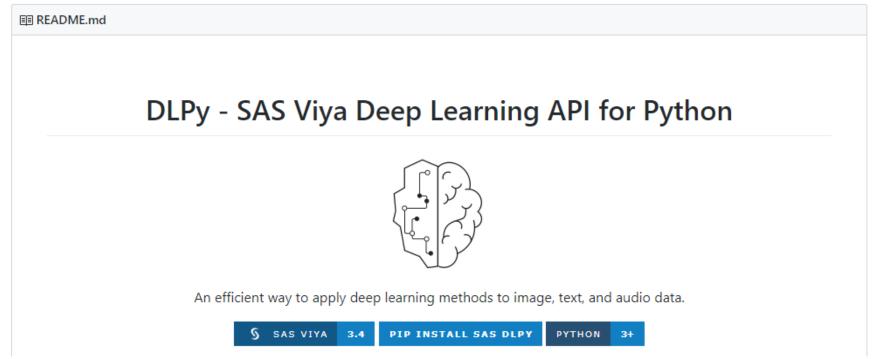




SAS Deep Learning

- DeepLearn action set for programming in CAS
 - Custom-built, VGG, ResNet, InceptionV3, YOLOv2, etc.
 - Transfer learning, GPU training/scoring
- DLPy easy-to-use python wrapper https://github.com/sassoftware/python-dlpy













SAS and Open Source Integration

Base SAS

 Execute open source code using system commands via the DATA step.

Interactive Matrix Language (IML)

- Submit R code within IML from SAS Code Editor.
- Pass data between R and SAS.

Enterprise Miner

- Execute R code from the open source integration node.
- Execute Python code using a system command.

SAS Viya

 Use open source software to take control of analytical tools.











CAS Actions

- The tools used to interact with data on the CAS server
- Act as wrappers for parallel processing algorithms
- Load data, transform data, compute statistics, perform analytics, and create output

Open Source Functions \equiv SAS Procedures \equiv CAS Actions

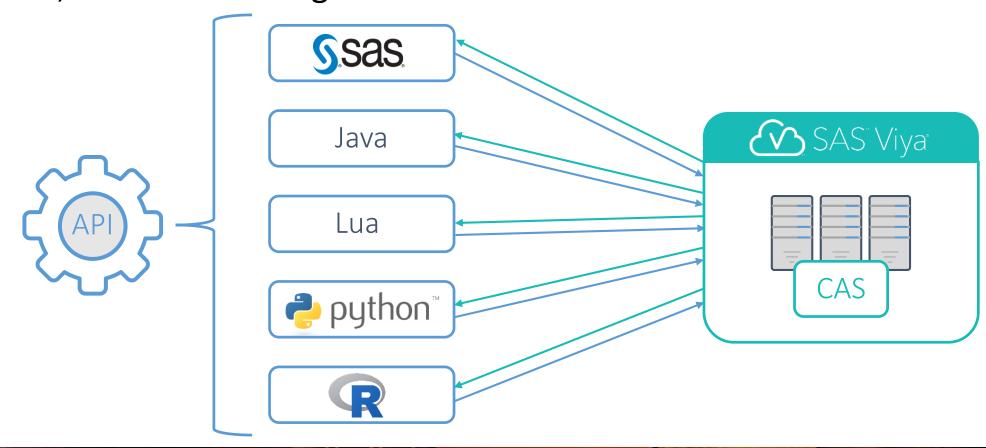






Interfaces to SAS Viya

• Although SAS Viya can be used by various SAS applications, it also enables you to access analytic methods from SAS, Python, Lua, and Java, as well as through a REST interface that uses HTTP or HTTPS.









R

₽ RO

Jupyter Notebook Jupyter Untitled Last Checkpoint: 2 hours ago (unsaved changes)





Code Cell

Linear Model Example with R and Jupyter Notebook

2000

CellToolbar

Equations

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

H ■ C Code

$$\hat{\sigma^2} = \frac{(Y - X\hat{\beta})^T (Y - X\hat{\beta})}{n - p}$$





Results

Results



Jupyter Notebook Results

```
In [1]: ## Create Linear Model Data ##
        set.seed(802)
        n = 50
        x = runif(n, min=10, max=30)
        beta0 = 2
        beta1 = 3
        error = rnorm(n, mean=0, sd=5)
        y = beta0 + x*beta1 + error
In [2]: ## Fit Model ##
        mod1 = lm(y \sim x)
        summary(mod1)
        Call:
       lm(formula = y \sim x)
        Residuals:
                    1Q Median 3Q Max
        -9.3360 -2.7005 -0.2983 2.9151 8.9333
        Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
        (Intercept) -0.6977
                              2.5214 -0.277 0.783
                     3.1214
                              0.1183 26.390 <2e-16 ***
       Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
        Residual standard error: 4.446 on 48 degrees of freedom
       Multiple R-squared: 0.9355, Adjusted R-squared: 0.9342
        F-statistic: 696.4 on 1 and 48 DF, p-value: < 2.2e-16
In [3]: ## Plot Model ##
        plot(y~x, col="blue")
        abline(mod1, col="red", lwd=2)
               90
```

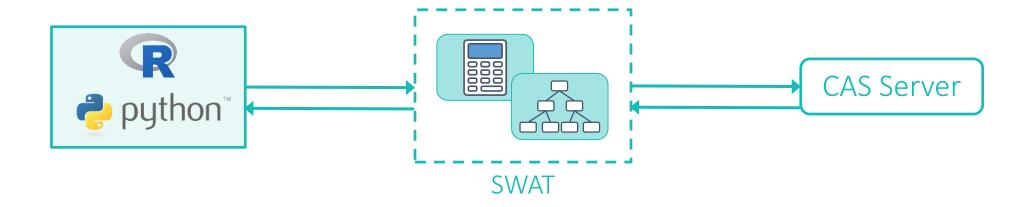






Scripting Wrapper for Analytics Transfer (SWAT)

- The SWAT package enables you to interface with CAS from R or Python.
- You can write an R or Python program that connects to a CAS server, load data into CAS, analyze large in-memory data sets quickly and efficiently using CAS actions, and work with results of your analyses using familiar data wrangling techniques in the open source language.





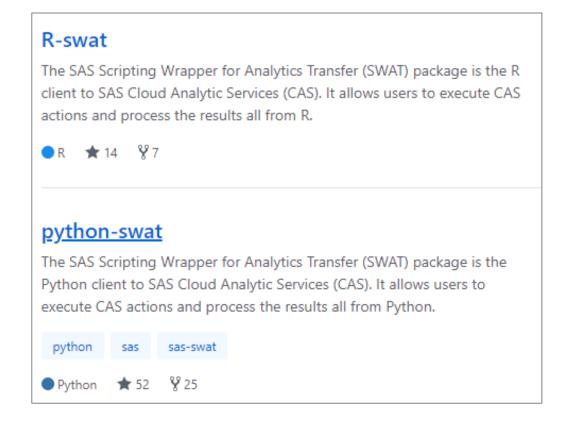






SAS GitHub Page

- Download the SWAT package from the SAS GitHub page.
- View system requirements, prerequisite packages required, and example notebooks.

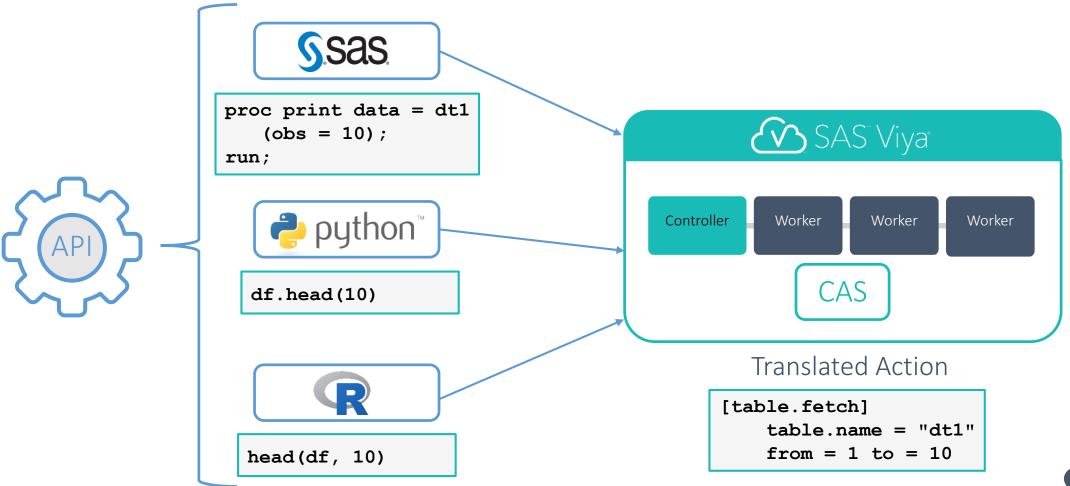








Same CAS Action, Different Interface









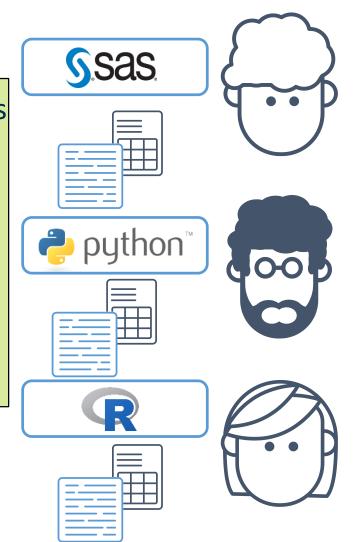


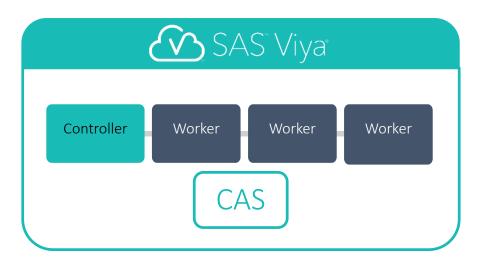




Open Architecture Advantage

- The same CAS action is used regardless of the interface and API.
- Results are equivalent (given seeds).
- Information can easily be passed between different languages.













SAS Viya for Learners Image Classification Jupyter Notebook Demo

https://v4e008.vfe.sas.com/JupyterHub/user/sugumara@oakland.edu/lab?redirects=1









Academic Collaboration with SAS, Resources for Academics, Syllabus

https://vfe.sas.com/portal/home







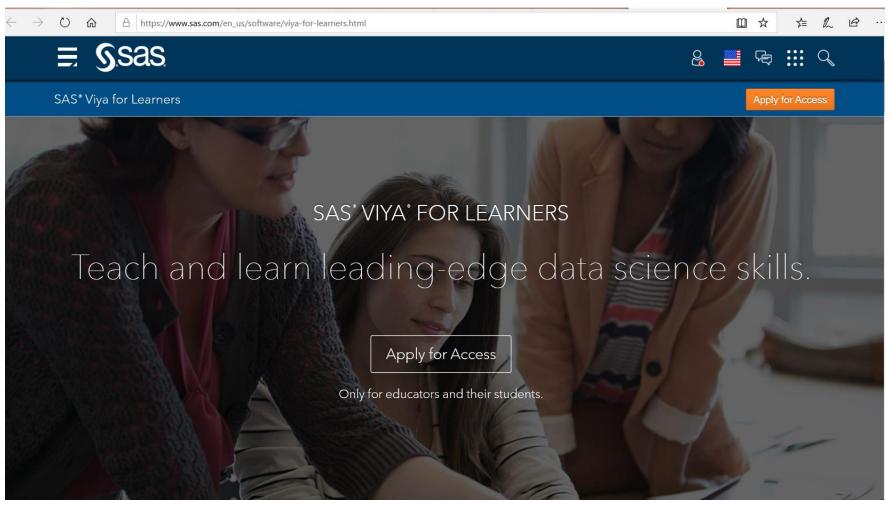






Accessing VFL

https://www.sas.com/en_us/software/viya-for-learners.html









Access to VFL

Educators

- Submit an online <u>Educator Access</u>
 Request Form
- Global SAS Education Approvers verify and approve the Educators.

Students

 Gain access when their instructor adds them to their course.















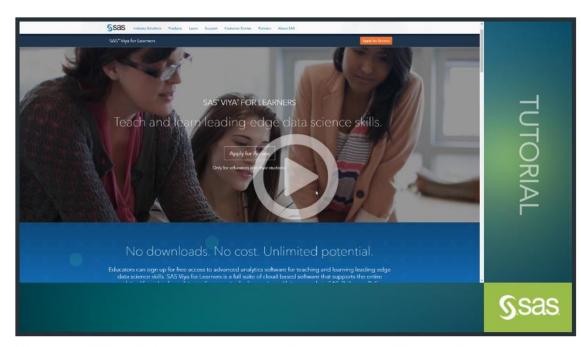
Why SAS® Viya® for Learners?

SAS® Viya for Learners

Apply for Access

For Educators

For Students



Discover how SAS Viya for Learners makes it easy for you to create and manage your course section so you can spend more time working with and teaching students.

Advantages for Educators

- Free suite of SAS Analytics software. Build a complete advanced analytics program from entry-level to doctorate with a single software environment.
- More time to spend teaching analytics. We'll provide the tools your students need to learn the software - including SAS courses online chat, web tutorials, e-learning opportunities, documentation, communities and world-class technical support - leaving you free to actually teach analytics and critical thinking skills.
- Curriculum support. Get curriculum support for using SAS in the classroom and building an advanced analytics curriculum.

We'll help you build an advanced analytics curriculum and delivery platform that will boost graduate career placement rates.







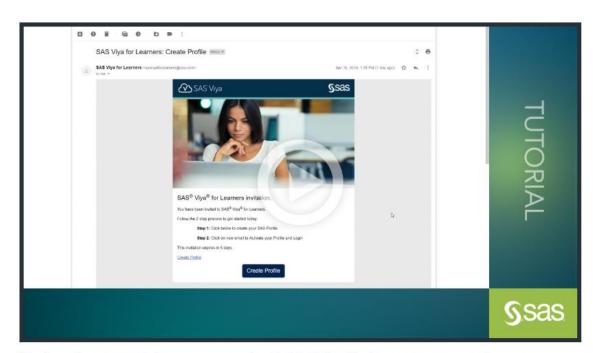
Why SAS® Viya® for Learners?

SAS® Viya for Learners

Apply for Access

For Educators

For Students



Find out how easy it is to get started with SAS Viya for Learners as part of a course requirement.

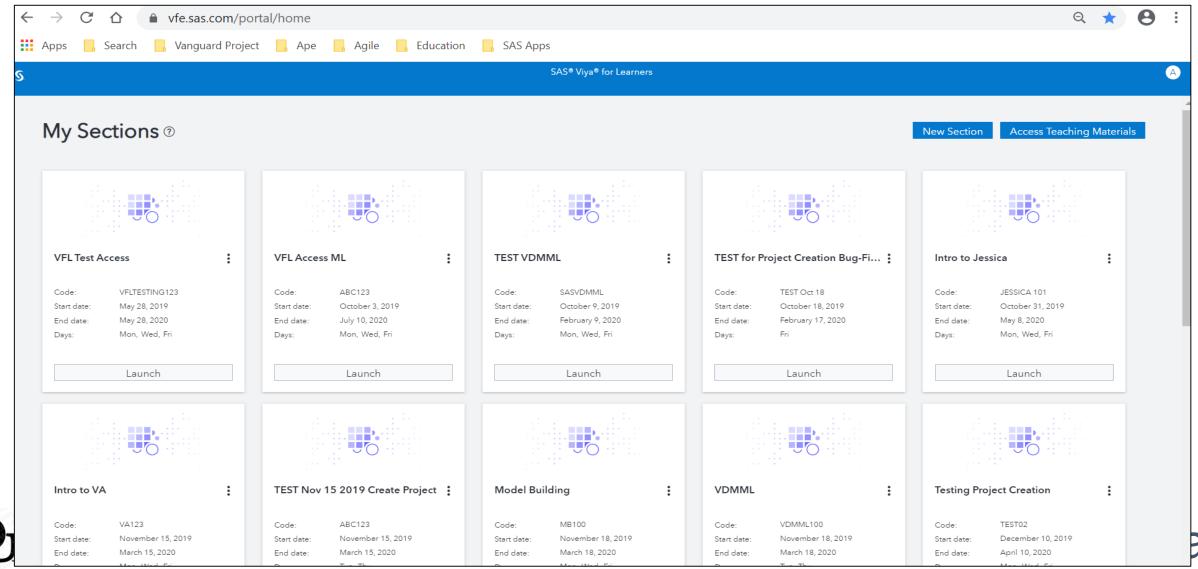
Advantages for Students

- Free SAS advanced analytics software. Program in your language of choice SAS, Python, R. Easily create visualizations to explain your work.
- Marketable skills. Explore data, discover insights and deploy models within a single, free software offering. And show off your skills with badges and certification opportunities.
- Real-world practice. Work with true business use cases using the same SAS software used by 92 of the top 100 companies on the Fortune Global 500.
- Free help when you need it. Access easy-to-digest learning materials - from basic programming through advanced deep learning, artificial intelligence and machine learning.

3



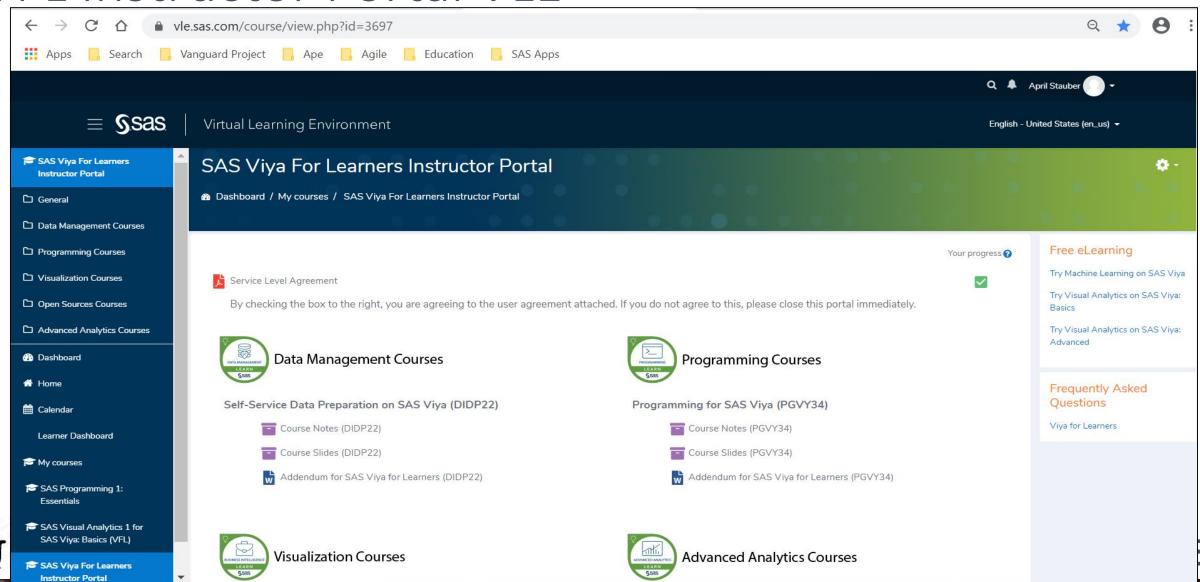
VFL Portal





VFL Instructor Portal VLE

https://vle.sas.com/course/view.php?id=3697















SAS Courses Available in VFL



Data Management Courses

Self-Service Data Preparation on SAS Viya (DIDP22)



Visualization Courses

- SAS Visual Analytics 1 for SAS Viya: Basic (YVA13V)
- SAS Visual Analytics 2 for SAS Viya: Advanced (YVA23V) added (1/8/20)



Open Source Courses

SAS Viya and Open Source Integration with R and Python (EVMLOPR)



Programming Courses

Programming for SAS Viya (PGVY34)



Advanced Analytics Courses

- SAS Visual Statistics on SAS Viya: Interactive Model Building (SVSO83)
- SAS Visual Data Mining and Machine Learning on SAS Viya: Interactive Machine Learning (VDMO83)
- Supervised Machine Learning Procedures Using SAS Viya in SAS Studio (DMML34)
- Machine Learning Using SAS Viya (CPML83)
 - SAS® Visual Text Analytics in SAS® Viya® (SVTA34) SAS®





SAS Academic Specialization

https://www.sas.com/en_us/learn/academic-programs/specializations/academic-specializations.html









Academic Specialization in SAS Analytics

- Started in Fall 2019
- Three Courses to be taken where SAS is used
 - MIS 5560 Introduction to Data Science
 - MIS 6900 Deep Learning and Text Analytics
 - QMM 5520 Forecasting
- Capstone Project
 - Use SAS in capstone project for analysis in addition to other tools









Additional Details

SAS - Oakland University Academic Specialization in SAS Analytics

Issued by SAS

Graduates of the Master of Science in Information Technology Management with Business Analytics concentration (MSITM-BA) at Oakland University earn a Tier 2 Academic Specialization in SAS Analytics recognizing their ability to leverage SAS analytical tools for statistical analysis, data mining, forecasting, machine learning, text mining, and visual data analytics.

Skills







Questions/Comments

Thank You for your Attention!

Vijayan Sugumaran: sugumara@oakland.edu

James Harroun: James. Harroun@sas.com

OU School of Business Administration

(http://www.oakland.edu/business)



