SAS | Python | R

** Best-of-all-Worlds Workshop **

Larry Orimoloye
Global Technology Practice
Agenda
AGENDA

• SAS Big data analytics platform (9:15am – 10:30am)
  • Introduction
  • SAS Analytics Lifecycle process
  • SAS Viya
  • SAS Viya Data Mining and Machine Learning
  • Demo

Break(30 Minute)

• SAS Viya API (Python) – (11am – 12.15pm)
  • Python crash course (Optional)
  • How to connect to CAS & Load data using jupyter notebook
  • Working with CasTable using jupyter notebook
    • Using CASTable objects like a DataFrame
  • Data exploration and summary statistics
  • SAS VIYA & Python model: Best of both worlds

Break(15 Minute)
AGENDA

• SAS Viya API (R) – (12.30pm – 1.30pm)
  • R crash course (Optional)
  • How to connect to CAS & Load data using R studio
  • Working with CasTable using R Studio
    • Using CASTable objects like a DataFrame
  • Data exploration and summary statistics
  • SAS VIYA & R model: Best of both worlds

• Lunch – (90 Minutes)

• Advanced topics (Optional) - (3pm – 4pm)
  • AI / NLP

• Questions – (4pm – 5pm)
The Analytical Lifecycle

Prepare
- Data Science
- Explorative
- Innovation
- New Data
- Experiments

Explore
- Ask
- Model
- Evaluate

Deploy
- Operations
- Automation
- Robust
- Actions
- Decisions

Data Warehouse / Data Lake

DATA

Data Scientist, Business Units

IT, Business Analyst, Business Units
Architecture Revolution: key drivers

**Powerful**
- Improve time-to-answer with faster deployment and provisioning
- Self-Service with API and APPS
- Innovative algorithms
- Optimised for in-memory, in-stream, in-Hadoop, in-DB, in-cloud, in-device

**Adaptive**
- Built to be elastic and scalable
- Support industry open standards for Cloud and on-prem deployment
- Consumption based pricing for selected product & solutions

**Open**
- Access through SAS languages and non-SAS languages
- 3rd party applications integration via API & Services

**Unified**
- One centralised analytic environment
- Analytical lifecycle end-to-end
- Administration, management, delivery and execution all integrated
SAS Visual Data Mining and Machine Learning is an end-to-end machine learning solution on the most advanced analytics platform to date – SAS Viya.
Visual Data Mining and Machine Learning

Visual Analytics
Visual Statistics
Visual Data Mining and Machine Learning

Baseline Procedures
VS Procedures
VDMML Procedures

Baseline Action sets
VS Action sets
VDMML Action sets

Requires Visual Analytics
Requires Visual Statistics

Python, Java, Lua, REST APIs

Copyright © SAS Institute Inc. All rights reserved.
SAS® Visual Data Mining and Machine Learning

- Decision Trees
- Generalized Linear Models
- K-means and K-modes Clustering
- Linear Regression
- Logistic Regression
- Nonlinear Regression
- Ordinary Least Squares Regression
- Partial Least Squares Regression
- Principal Component Analysis
- Quantile Regression

- Factorization Machines
- Gradient Boosting
- Neural Networks
- Random Forest
- Support Vector Machines

- Network Analytics/Community Detection
- Boolean Rules
- Text Mining
- Autotuning

- Assess Supervised Models
- Creates score code

- Multi Threaded Data Step
- DS2
- SQL
- Variable Binning
- Variable Cardinality Analysis
- Sampling and Partitioning
- Missing Value Imputation
- Variable Selection
- Transpose
- Image processing
SAS Viya Data Mining and Machine Learning

New Capabilities

• Gradient Boosting
  • New distributed algorithm – all new implementation

• Factorization Machines
  • New distributed algorithm for large scale sparse data
  • “Recommendation Engine” - primary business driver

• Auto-tuning capabilities
  • Available for complex algorithms
    Decision Trees, Random Forests, Neural Networks, Gradient Boosting, Support Vector Machines, Factorization Machines

• High Frequency Analytics
  Support Vector Data Description, Robust PCA, Moving Windows PCA, Short Time Fourier Transform

• Image Analysis
AUTOTUNING HYPERPARAMETERS

- Highly data dependent
- Related to model complexity

Auto Tuning:
- Automate hyperparameters search and find the optimal set
- Maximize predictability on independent data set
- Aims to avoid over-fitting by controlling model complexity
- Creates more accurate models faster vs hand tuning
- SAS auto tuning leverages world class SAS optimization engines

- Decision Trees
- Neural Networks
- Gradient Boosting
- Random Forest
- Support Vector Machines
- Factorization Machines
SAS supports four methods: Random search, Latin hypercube sampling (LHS), Bayesian search, Genetic algorithm (LHS + Optimization).
SAS Viya
For Many Users

Visual Interfaces

Programming Interfaces

API Interfaces

GUI user

SAS Coder Non-SAS Coder

Developer
SAS® Visual Analytics

**Reporting**
Build and share reports and dashboards that help your organization make intelligent decisions.

**Discovery**
Explore your data with visualizations that help you understand even the most complex data.

**Self-Service Analytics**
The power of easy-to-understand analytics, in the hands of more users and provide path towards maturity.
Data Preparation

- Access to different data sources
- Table and column profiling
- Filter data and column transformations
- Visual joins
Visual Exploration

• Discover relationships, trends, outliers
• Smart auto charting
• Analytics driven visualizations
• Publish any visualization as a report object.
Self-Service Analytics

- Descriptive statistics
- Forecasting and scenario analysis
- Supervised and unsupervised learning
- Text analytics
Advance Analytics

- Highly scalable, in-memory analytical processing
- Integrated data mining and machine learning
- Modern machine learning algorithms
- Model development
- Model assessment and scoring
SAS® Visual Data Mining and Machine Learning

GUI Interface – SAS Visual Analytics

- **Machine Learning**
  - Forest
  - Gradient Boosting
  - Neural Networks
  - Support Vector Machines
  - Factorization Machines

- **Statistics**
  - Linear Regression
  - Logistic Regression
  - GLM Regression
  - Clustering (k-means)
  - Decision Tree
SAS® Visual Data Mining and Machine Learning
Programmatic Interface - SAS Studio

- Web Interface
- Interactive Program Editor
- Snippets
- Tasks
- Program Generator
### Programming Interfaces – Open APIs

- Access to all analytical actions for programmers
- Script-Save-Share-Schedule
- Open API Integration
- Integration from Python, Java, Lua, R
SAS® Viya™ and Python
SAS Scripting Wrapper for Analytics Transfer (SWAT) for Python

• Access to SAS Viya from Python
• Integration of SAS Analytics in Python code
• Jupyter Notebook support
• Issue tracking and collaboration in development through GitHub project
SAS® Viya™ and R
SAS Scripting Wrapper for Analytics Transfer (SWAT) for R

- Access to SAS Viya from R
- Integration of SAS Analytics in R code
- R Studio and Jupyter Notebook support
- Issue tracking and collaboration in development through GitHub project
What Does SWAT Do in a Nutshell

• SAS Scripting Wrapper for Analytics Transfer (SWAT) packages are open source interfaces to CAS

• Python coders can have access to the SAS Cloud Analytic Services (CAS) engine (the centre piece of the SAS Viya framework)

• You can load and analyse large data sets using processing power of CAS engine (either on a physical server or on cloud) and execute workflows of CAS analytic actions from Python on the client side.

• The SWAT package mimics much of the API of the Pandas package so that using CAS feels familiar to Pandas users.

• Install client components from GitHub

  #pip install https://github.com/sassoftware/python-swat/releases/download/vX.X.X/python-swat-X.X.X-platform.tar.gz

• Dependencies:
  • 64-bit Python 2.7, 3.4, or 3.5 on Linux
  • SAS Viya
SAS® Viya™ Programming
As viewed through Legos...

SAS Studio can include one or more procedures to create a task.

Stack one or more action sets to create procedures

Stack one or more actions to create action sets

Each action has one or more parameter settings.

Actions are at the heart of every CAS procedure
SAS® Viya™: New Open Architecture
Different Languages – Same Power

proc print data = hmeq (obs = 10);
run;

df = s.CASTable('hmeq')
df.head(10)

df <- defCasTable(s, 'hmeq')
head(df, 10)

[table.fetch]
    table.name = "hmeq"
    from = 1 to = 10
Calling VDMML Analytics from Python

### Machine Learning Workflow

**Set Up the Jupyter Notebook for Analysis**

Note: We have our brand new package called `swat - SAS Scripting Wrapper for Analytics Transfer` - available on GitHub via pip install

```python
# Import necessary packages and modules
import swat
import pandas as pd
import collections
from matplotlib import pyplot as plt
from swat.render import render_html

# Define options for print message
has_print_messages = False
hasPlotLibInline = True

# Define directory and data file name
indata_dir = "C:\opt\main\inside\newsdata"
indata = "news"
target = "type"

# CAS Server connection details
cas_host = "sampleserver.demos.sas.com"
cas_port = 5000

cas_user = "admin"
cas_password = "casuser"
cas_servername = "sample"
caslib = "casuser"

# Load actionsets for analysis (for data prep, modeling, assessing)
actionsets = ['table', 'cardinality', 'sampling', 'foSQL', 'decisionTree', 'nearMiss', 'svm', 'naive']

# LoadDataActionset([i] for i in actionsets)
```

```python
    # Added action set 'table'.
    # Added action set 'cardinality'.
    # Added action set 'sampling'.
    # Added action set 'foSQL'.
    # Added action set 'decisionTree'.
    # Added action set 'nearMiss'.
    # Added action set 'svm'.
    # Added action set 'naive'.
```

```python
# Build actionset([i] for i in actionsets)
```

```python
    # Build action set 'table'.
    # Build action set 'cardinality'.
    # Build action set 'sampling'.
    # Build action set 'foSQL'.
    # Build action set 'decisionTree'.
    # Build action set 'nearMiss'.
    # Build action set 'svm'.
    # Build action set 'naive'.
```
SAS Viya with SWAT
Open API for Python for SAS Viya (CAS Engine)
https://github.com/sassoftware/python-swat

SAS Viya with SWAT R
Open API for R for SAS Viya (CAS Engine)
https://github.com/sassoftware/R-swat
SAS® ENTERPRISE MINER VIYA™ BRIDGE NODE  
Increase the Reach

- Model comparison
- Testing new SAS Viya algorithms
- Simultaneous processing of all model trainings
- Model ensembles
- Integration with SAS Enterprise Miner Metadata
- Management, deployment and monitoring of SAS Viya models
Get Started with SAS® Viya™ Enablement

SAS Viya™ Virtual Learning

Support your adoption

• Free resources
  • Getting Started training
  • Video Libraries
  • e-Learning

• Topics Include
  • Administration
  • Programming & Analytics
  • SAS Visual Analytics
  • Open Source Integration
SAS Visual Data Mining and Machine Learning

Some highlights

- Custom pipelines
- Parallel execution
- SAS best practice pipelines (i.e. Rapid Predictive Modeler)
- Defining and sharing custom reusable nodes
- Ability to deploy models into databases directly
- Score code generation and score code APIs
- Import SAS 9.4 score code into pipeline comparison
- SAS Code integration
DEMO

Visual Data Mining and Machine Learning
Break – 30minute

(10:30-11AM)
SAS Viya API (Python)

Timing (11-12:15PM)

Hands-on session
• Python crash course (Optional)
• How to connect to CAS & Load data using jupyter notebook
• Working with CasTable using jupyter notebook
  • Using CASTable objects like a DataFrame
• Data exploration and summary statistics
• SAS VIYA & Python model: Best of both worlds
Decentralized
Centralized
Shared Data, Shared Computing
Demo Assets
Click these!
Resources
Break – 15minute
(12:15-12:30PM)
SAS Viya API (R)

Timing (12:30-1:30PM)

Hands-on session
• R crash course (Optional)
• How to connect to CAS & Load data using R Studio
• Working with CasTable using R Studio
  • Using CASTable objects like a DataFrame
• Data exploration and summary statistics
• SAS VIYA & R model: Best of both worlds
Demo Assets

Click these!

---

**Data Mining Workflow**

**Set Up the Jupyter Notebook for Analysis**

Note: We have our brand new package called swat - SAS Scripting Wrapper for Analytics Transfer - available on GitHub via pip install.

Copyright © 2017 SAS Institute Inc.

```
import necessary packages and modules
import swat
import pandas as pd
import collections
from matplotlib import pyplot as plt
swat.options.cas.print_messages = False
matplotlib inline

# Define directory and data file name
indata_dir = '~/sasviyafiles'
indata = 'money'
```

---

**Data Mining Workflow**

**Set Up the R Notebook for Analysis**

```
# Load necessary packages
library('swat')

# SWAT 1.0.0
library('ggplot2')
library('reshape2')
options(cas.print.messages = FALSE)
```
Lunch – 90 minute
(13:30-15:00PM)
Artificial Intelligence/NLP
(15:00 - 16:00PM)
Artificial Intelligence is the science of training systems to emulate human tasks through learning and automation.
ARTIFICIAL INTELLIGENCES CHALLENGES

CHALLENGE

- Lack of Customer insights
- Human-like tasks are not automated
- Technology and supplier not reliable
- Lack of Banking Tailored applications
- Loads of Unused Unstructured data such as Text and Images

Business Executive
ARTIFICIAL INTELLIGENCE CHALLENGES

CHALLENGE

- Lack of Customer insights
- Human-like tasks are not automated
- Technology and supplier not reliable
- Load of Unused Unstructured data such as Text and Images
- Large unstructured data, poor quality & governance
- Increased integration complexity with open source
- Data Scientists/Researchers are a scarce resource
- Time consuming analytical processes

Analytics Practitioner
ARTIFICIAL INTELLIGENCE CHALLENGES

**CHALLENGE**

- Lack of Customer insights
- Human-like tasks are not automated
- Technology and supplier not reliable
- Lack of Banking Tailored applications
- Loads of Unused Unstructured data such as Text and Images
- Large unstructured data, poor quality & governance
- Increased integration complexity with open source
- Data Scientists/Researchers are a scarce resource
- Time consuming analytical processes
- Difficult and costly to develop and maintain
- No real-time capabilities
- Lack of Automation and Repeatability

IT Executive
ADVANCING WITH ARTIFICIAL INTELLIGENCE: CHALLENGES & VISION

**CHALLENGE**

- Lack of Customer insights
- Lack of Banking Tailored applications
- Increased integration complexity with open source
- Difficult and costly to develop and maintain
- Human-like tasks are not automated
- Loads of Unused Unstructured data such as Text and Images
- Data Scientists/Researchers are a scarce resource
- No real-time capabilities
- Time consuming analytical processes
- Lack of Automation and Repeatability
- Large unstructured data, poor quality & governance

**STRATEGY**

- Repetitive learning and discovery through data
- Develop Governed and Tailored Applications
- Scale Analytics (deep learning) process to cope with big data
- Get more out of data with increase accuracy

**VISION**

- Automate manual tasks
- Add new features to existing solutions
- Better value from investments in Analytics
- Improve customer insight
- Competitive advantage through the best models and data usage

Copyright © SAS Institute Inc. All rights reserved.
Some application of AI in banking

- Anti-fraud
- Credit Application
- Real time Transaction Analysis
- Customer Optimization Journey
- Chatbot
- Personalized Recommendation
- Next Best Action
- Anti-money Laundering
- Algorithmic Trading
Insurance industry

- Standardize image size
- Convert from color to grayscale
- Reduce noise and create binary data
- Recognize contours
Cancer treatment
Wild-Life conservation
Performance assessment
Movie Industry
Customer journey optimization - Reinforcement learning
Open discussion

What are your AI use cases?
Deep Learning
- Deep forward networks
- Auto encoders
- Convolutional networks
- Recurrent networks

Cognitive Computing
- Speech to Text
- Natural language interaction
- Natural language generation
- Biomedical image processing

Powered by Viya®
- MPP & GPU processing
- Python, Lua, Java, CASL and REST

AI METHODS
Deep Learning Toolkit
Ships with VDMML license as CAS actions

CAS Action, but also built into MS pipeline for Neural Nets with more than 5 layers

Deep Forward

Convolutional

Autoencoders

Recurrent

CAS Actions Only
SAS deep learning toolkits

- Platform

- CASL, Python, R, DL API

- SAS Deep Learning Actions
  - Training and Scoring

- ESP Engine
  - Scoring

- CAS Platform
  - Hadoop Storage Engine
  - GPU
Image Actions
processImages action

Contour image is combination of:
- Convert_color
- Bilateral_filter
- Threshold
- Laplacian
- Contours
BioMedical
Deep Learning Python APIs

Deep Learning

DL APIs

Image

Read/Write

Processing

Data Manipulation

Construction

Training

Scoring

Deploy

Keras

TensorFlow

Caffe

Theano

mxnet

SAS Viya

torch

Caffe2

Copyright © SAS Institute Inc. All rights reserved.
Design Strategy
Streaming A.I.

Sample Images
Profile Labels: gender / outfit / company

Model Dev / Execute / Monitor

Model Training

Alerts / Reports / Decisioning

Marketing Automation

Channels

IoT Data (camera)

Model Deployment @ Streaming Data

Copyright © SAS Institute Inc. All rights reserved.
Streaming AI

Train At-Rest & Score In-Stream

SAS Machine Learning Models
- Forest, Gradient Boosting
- Factorization Machine
- Support Vector Machine
  - Robust PCA
- Support Vector Data Description
- Text Analytics (ASTORE only)
  - Deep Learning
  - Bayesian networks

Train & Score In-Stream

SAS Machine Learning Models
- Streaming K-Means
- Streaming DBSCAN
- Streaming Linear Regression
- Streaming Logistic Regression
- Streaming Support Vector Machine

Streaming Data Prep for AI

SAS Algorithms (among others)
- Streaming Text Tokenization and Text Vectorization
- Streaming Receiver Operating Characteristic Information
- Streaming Image Processing

$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$
Reinforcement Learning in CI360
Coming Soon in 2018

Create relevant, satisfying, valued customer experiences.

Allow marketers to automatically learn what sequences of content maximize conversion goal on a web page, also taking into account different user attributes.
Cognitive Computing
Enable more “human-like” interactions with our software solutions. Reason on input. Explain on output.

Cognitive computing is based on self-learning systems that use machine learning techniques to perform specific, human-like tasks in an intelligent way.
Detecting Emotions
Validate Safety Compliance

Hard Hat Demo
Image Classification

Classifying images in real time
Thank You!
sas.com