How does SAS® Support Machine Learning (ML)?

Getting the most out of your data
How does SAS support Machine Learning?

Agenda

• What is Machine Learning?

• Terminology and key characteristics

• How you can use machine learning in SAS

• Examples in SAS 9 and SAS Enterprise Miner

• Examples in SAS Viya
Machine Learning & Artificial Intelligence
What is Machine Learning?

Definition

- Automatic
- Adaptive

Using iterative processes, machine learning builds models that automatically adapt with little or no human intervention.
Why is it so important now?

Data

Computing Power

Algorithms
Terminology
Terminology
Machine learning terms versus inferential statistics terms

What are all these archaic, outmoded and confusing terms?

- Feature
- Input
- Target
- Object

What are all these new fangled and confusing terms?

- Variable
- Independent Variable
- Dependent Variable
- Observation
Terminology
What are Machine Learning terminology?

• In statistics we predict a Y or a dependent variable.
• In data mining, Y is called a target.
• In machine learning, a target is called a label.
• In statistics and data mining our inputs are called X’s.
• In machine learning our inputs are called features.
• In statistics and data mining we transform our X’s.
• In machine learning we do feature creation.
How Does Machine Learning Work?
Distinguish apple from orange
How Does Machine Learning Work?

Distinguish Granny Smith apple from Fuji apple
How Does Machine Learning Work?
Finding the rotten apple
How Does Machine Learning Work?

Supervised Learning

Trained on labeled examples
How Does Machine Learning Work?

Unsupervised Learning

Trained on unlabeled examples
How Does Machine Learning Work?

Semi-Supervised Learning

Use labeled and unlabeled observations
How Does Machine Learning Work?

Semi-Supervised Learning

Use labeled and unlabeled observations
How Does Machine Learning Work?

Not New for SAS

Machine Learning has been available in both SAS/STAT and Enterprise Miner for decades

- Neural Networks
- Decision Trees
- Random Forests
- Clustering
- Gradient Boosting
- Text Analytics
- Regression
**PROC DISCRIM** (K-nearest-neighbor discriminant analysis)  
– James Goodnight, SAS founder and CEO, 1979

*1950’s:*
  - Samuel’s checker program

SAS Data Mining Primer Course  
SAS Institute, 1998

Neural Networks and Statistical Models,  
SAS Institute, 1994
SAS Machine Learning

ALGORITHMS

• Neural networks
• Decision trees
• Random forests
• Associations and sequence discovery
• Gradient boosting and bagging
• Support vector machines
• Nearest-neighbor mapping
• k-means clustering
• Self-organizing maps
• Local search optimization techniques such as Genetic algorithms

• Regression
• Expectation maximization
• Multivariate adaptive regression splines
• Bayesian networks
• Factorization Machines
• Kernel density estimation
• Principal components analysis
• Singular value decomposition
• Gaussian mixture models
• Sequential covering rule building
• Model Ensembles
• And More......
accessed named SAS Institute Inc. All rights reserved.

Did you know?

HPSplit Procedure Documentation

HPSplit Procedure Documentation

proc hpsplit data=sashelp.hmeq maxdepth=7 maxbranch=2;
  target BAD;
  input DELINQ DEROG JOB NINQ REASON / level=nom;
  input CLAGE CLNO DEBTINC LOAN MORTDUE VALUE YOJ / level=int;
  criterion entropy;
  prune misc / N <= 6;
  partition fraction(validate=0.2);
  rules file='hpsplhme2-rules.txt';
  score out=scored2;
run;
SAS Enterprise Miner
Algorithms – basic and advanced

SAS Enterprise Miner
- Linear & Logistic Regression
- Decision Trees
- Random Forest
- Gradient Boosting
- Support Vector Machines
- Neural Networks
- Clustering
- Bayesian Networks
- Principal Components
- Open Source Models
Classification

Our example today

- The dataset is from a financial institution with customer demographics and loan/credit behavior.
- The goal of this modeling exercise is to predict which people are likely to default on a home equity loan.
- The data are at the customer-level (subject-level).
- n=5960
- columns = 13
SAS Enterprise Miner
Machine Learning and Artificial Intelligence Demo
SAS Viya
SAS Viya Products

• SAS Viya is an underlying foundation for additional products that will take advantage of a cloud-enabled, open platform. Most offerings include both a coding interface as well a visual interface.
  • SAS Visual Analytics
  • SAS Visual Statistics
  • SAS Visual Data Mining and Machine Learning
  • SAS Visual Forecasting
  • SAS Visual Text Mining
  • SAS Optimization
  • SAS Econometrics
  • SAS Visual Investigator
SAS® Visual Data Mining and Machine Learning

Visual "drag & drop" Interface

Programming Interface

Data Preparation

Visual Exploration

Model Studio

Model Deployment

Machine Learning
**SAS® Viya™ Based Product offerings**

### Interfaces

- **SAS Visual Analytics**
  - Visual Analytics (VA) Interface
- **SAS Visual Statistics**
  - Visual Statistics (VS) add-on to VA Interface
- **SAS Visual Data Mining and Machine Learning**
  - Visual Data Mining and Machine Learning (VDMML) add-on to VA Interface

#### Visual approach

- **SAS Visual Analytics**
- **SAS Studio**
  - CAS actions, PROCS related to VA capabilities

#### Programmatic approach

- **SAS Studio**
  - CAS actions, PROCS related to VS capabilities
- **SAS Visual Data Mining and Machine Learning**
  - CAS actions, PROCS related to VDMML capabilities
How Does Machine Learning Work?
Example Algorithms in SAS VS/VDML

**SUPERVISED LEARNING**
- Tree-based models (Decision Trees, Random Forests, Gradient Boosting)
- Regressions
- Neural Networks
- Support Vector Machines
- Recommender Systems

**UNSUPERVISED LEARNING**
- Clustering
- Principal Component Analysis
- Text Topic Detection

**SEMI-SUPERVISED LEARNING**
- Clustering
- Factorization Machines
Introduction: Software Overview

Multiple Interfaces Target Different Users
Interfaces
Building a Model from Scratch in the Visual Reporting Interface
Interfaces

Building a Model Using SAS Studio Tasks
Interfaces

Building a Model Using SAS Studio Snippets
Interfaces

Building a Model Using Python

```python
import sas
from sas import *
from pprint import pprint
from matplotlib import pyplot as plt
import pandas as pd
import sys
import matplotlib

cashost='gatekrbhdap01.gatehadoop.com'
casport=5570
casauth='/home/sasdemo/.sauthinfo'
indata_dir='/ugt/sas/data'
indata='hmeq'
sess = sas.CAS(cashost, casport, authinfo=casauth, caslib="mycaslib")
sess=sess.table.addcaslib(name="mycaslib", path="/opt/sas/data/", datasource=[results=sess.table.caslibInfo(caslib="mycaslib"), verbose=True])
info=sess.table.fileinfo(caslib="mycaslib")
res=sess.table.loadtable(path="hmeq.sas7bdat", caslib="mycaslib", casout="name")
```

BAD: BAD indicates whether or not a home equity loan recipient defaulted (BAD = 1) or not (BAD = 0).

1. Set up (same as previous exercise)
SAS Viya Key Features

Visual Interface

• Modeling Techniques
  • Clustering (k-means)
  • Linear Regression
  • Logistic Regression
  • GLM Regression
  • Decision Trees
  • Forest
  • Factorization Machine
  • Gradient Boosting
  • Neural Network
  • Support Vector Machine

• Common Features
  • Training & Validation partitioning
  • Variable Importance / Profile
  • Model Assessment
  • Model comparison
  • Derivation of predictive outputs
  • Ability to export model statistics into Excel
  • Score Code or Astore Table
Includes algorithms in the visual interface plus

- Unsupervised Learning
  - Moving Window PCA
  - Robust PCA
  - Support Vector Data Description
  - Text Parsing and Topic Discovery
- Supervised Learning
  - Boolean Rules
Visual Interface Demo

SAS Visual Analytics
Visual Interface

Pipelines
SAS® Viya
Pipelines

Drag-and-drop pipelines including preprocessing and machine learning techniques

Customizable and portable nodes and SAS best practice pipelines (Toolbox)

Support for SAS coding (macro, data step, procs, batch Enterprise Miner) within pipelines

Collaboration through the use of the “Toolbox” – a collection of SAS Best Practice Pipelines, in addition to user-generated templates

Example Code for Pipeline
SAS® Viya
Pipelines

Data Mining Preprocessing
- Anomaly Detection
- Clustering
- Feature Extraction
- Filtering
- Imputation
- Manage Variables
- Replacement
- Text Mining
- Transformations
- Variable Clustering
- Variable Selection

Supervised Learning
- Batch Code
- Bayesian Network
- Decision Tree
- Forest
- GLM
- Gradient Boosting
- Linear Regression
- Logistic Regression
- Neural Network
- Quantile Regression
- Score Code Import
- SVM

Postprocessing
- Ensemble

Miscellaneous
- Data Exploration
- Open Source Code
- SAS Code
- Save Data
SAS® Viya
Pipelines
Automated API generation for retraining and scoring
Ability to deploy models into databases directly
Assessment against imported Test datasets
Integration with Model Manager for versioning, tracking, and deployment
Integration with SAS 9.4 Enterprise Miner score code & Batch Code
Integration with Open Source (R and Python)
Visual Interface Demo

Pipelines
Autotuning
Automating Autotuning: Hyperparameters

- Training a model involves using an algorithm to determine model parameters or other logic to map inputs to a target.
- Tuning a model involves determining the algorithm hyperparameters (tuning options) that result in the model which maximizes predictability on an independent data set.

Very data/problem dependent!
SAS Programming Interface - AutoTuning Demo
SAS Visual Analytics & SAS Studio
Resources
Where can I learn more?
SAS Viya - Ask the Expert

http://support.sas.com/training/askexpert.html
Recommended Resources

An Overview of SAS® Visual Data Mining and Machine Learning on SAS® Viya

Video - Automated Machine Learning at Scale

Machine learning - what it is and why it matters (reading)

Live web and classroom training - Big Data, Data Mining, and Machine Learning
Big Data course
SAS® Visual Statistics
Try it before you buy!

Apply now for a free 14-day trial of SAS Visual Statistics running on SAS Viya. Experience the power and ease of building and refining descriptive and predictive models to quickly surface valuable insights.

About the Trial
• There's nothing to download; a web browser is all you need (we recommend Google Chrome 64-bit for the best experience).
• Use your own data with the trial to see how it works. Or if you prefer, you can use our sample data and demos provided.

You'll need to sign in to your SAS Profile to request the trial.

Sign in

Don't have a SAS Profile? Create one now.
SAS® Visual Data Mining and Machine Learning

Try it before you buy!

SAS® VISUAL DATA MINING AND MACHINE LEARNING

Everything you need to solve the most complex analytical problems – in a single, integrated, collaborative solution.

Try it for free

Copyright © SAS Institute Inc. All rights reserved.
Communities
Questions?

Thank you for your time and attention!

Connect with me:
LinkedIn: [https://www.linkedin.com/in/melodierush](https://www.linkedin.com/in/melodierush)
Twitter: @Melodie_Rush

sas.com