

Demo: SVM

Dr. Goutam Chakraborty

SAS® Professor of Marketing Analytics

Director of MS in Business Analytics and Data Science* (<http://analytics.okstate.edu/mban/>)

Director of Graduate Certificate in Business Data Mining (<http://analytics.okstate.edu/certificate/grad-data-mining/>)

Director of Graduate Certificate in Marketing Analytics (<http://analytics.okstate.edu/certificate/grad-marketing-analytics/>)

- *Name change pending internal approval.
- Note some of these slides are copyrighted by SAS® and used with permission. Reuse or redistribution is prohibited

1

1

Objectives

- Introduce the concept of high-performance computing.
- Discuss support vector machines (SVM) in SAS Enterprise Miner.
- Demonstration of SVM using SAS EM

2

2



Why the Need for High Performance?

- Very computationally intensive model fitting or simulations can cause slow execution.
- Massive amounts of data are stored in databases. Extraction into traditional computing environments requires massive data movement, which results in very poor performance.
- Many algorithms require multiple passes of the data for training models. This exacerbates the performance problem.
- Is this for you? It depends.

3

3



Why Do We Care in This Class?

- Two of the modeling nodes that we are going to use are high performance.
 - High-Performance Support Vector Machine
 - High-Performance Forest
- There are not non-high-performance versions of these modeling nodes.

4

4

Some Terminology

Appliance	A grid of computers that runs database software.
Access Engine	Enables communication between the SAS server and the appliance. It is the mechanism by which data is read from and written to databases.
Threaded Kernel (TK) Extensions	The mechanisms by which distributed computing is done.
HPDM nodes	High-Performance Data Mining nodes. The High-Performance nodes in Enterprise Miner.

5

5

What Makes HPDM Nodes High Performance?

Enterprise Miner HPDM nodes use SAS High-Performance Analytics procedures.

- SAS High-Performance Analytics procedures
 - are written using threaded kernel extensions
 - are designed to run in a distributed environment
 - are multi-threaded, to take advantage of multiple cores in single-machine mode
 - use algorithm choices that consider data movement and replication
 - use in-memory analytics to minimize data movement and substantially reduce required computing time, especially for algorithms that require multiple passes of the data.

6

6

Distributed Computing Means Knowing Faster

SAS processing in tandem with data



Apache Hadoop on Commodity Hardware

7 ...

7

SAS Solution to the Performance Challenge

- Single-machine mode – multiple cores on a single computer (also called *symmetric multiprocessing* or *SMP mode*)
- Distributed mode – many architectures and configurations to take advantage of massively parallel processing (MPP)

8

8



In This Class

- The HPDM nodes are run in single-machine (that is, SMP, non-distributed) mode.
- Some of the actions shown in the demonstrations that follow will not work in a distributed computing environment.

9

9



HP SVM in SAS Enterprise Miner

- Experimental SVM node (non-high performance) prior to Enterprise Miner 13.1
- PROC HPSVM
- PROC SVMSCORE

10

10



HP SVM Target Measurement Level

- Only binary targets supported
- No multi-class targets supported, nominal or ordinal
- No interval targets supported

11

11



Available Optimization Methods

Two methods are available:

- Interior point
- Active set

- **Note:** High-performance modes of operation:
 - Interior point runs in either single-machine or distributed mode.
 - Active set runs only in single-machine mode.

12

12



Business Problem and Data Description

- PMAD_PVA data
- This data set was used in an earlier class
 - Lapsing donor
 - Target variable is Target_B
 - There is a bunch of Input variables

13

13



SAS EM Details

- Add the PMAD_PVA data set with all default settings to your project
- Import the diagram “SVM”
- Look at settings of metadata
 - Target variable needs to be binary for SVM in SAS EM
- Look at properties panel of 4 SVM nodes (HPDM Tab)
- Run and look at results from the Model Comparison node and a few of the SVM models
- Use View > Model > Model Information to understand high performance mode of operation for any SVM in SAS

14

14



Base SAS for SVM

Syntax: HPSVM Procedure

The following statements are available in the HPSVM procedure:

- **PROC HPSVM DATA=SAS-data-set** <options>;
- **CODE** FILE=filename;
- **ID** variables;
- **INPUT** variables-list / LEVEL=level <option>;
- **KERNEL** kernel-type / kernel-parameters;
- **OUTPUT** <output-options>;
- **PARTITION** <partition-options>;
- **PENALTY** C=constant-list;
- **PERFORMANCE** <performance-options>;
- **SELECT** <select-options>;
- **TARGET** target / <option>;

15

15



These are Some Self-study Slides

Research is being done in the area of SVM for regression purposes (that is, building SVM models on interval targets), but Enterprise Miner currently does not have this capability. The following information about SVM for regression purposes is provided simply for student information as a self-study section.

16

16