

Intro to Machine Learning

Ask the Expert

Melodie Rush

Customer Success Principal Data Scientist

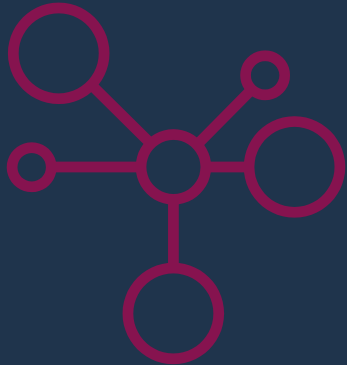
Connect with me:

LinkedIn: <https://www.linkedin.com/in/melodierush>

Twitter: @Melodie_Rush



How does SAS support Machine Learning?



Agenda

- What is Machine Learning?
- Terminology and key characteristics
- Introduction to Decision Trees, Random Forest, Gradient Boosting, Neural Networks, and k-means Clustering
- How you can use machine learning in SAS
- Examples in SAS 9.x and SAS Viya

Machine Learning



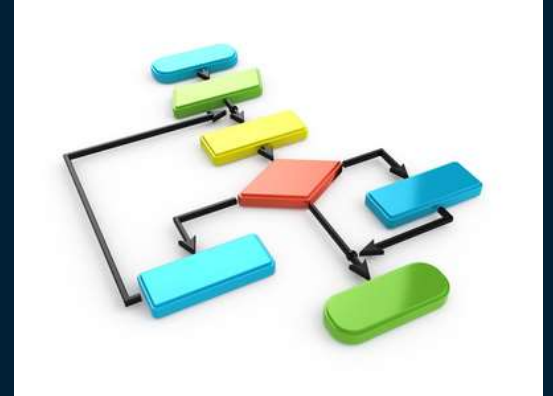
Why is it so important now?



Data



**Computing
Power**



Algorithms

Terminology

Terminology

Machine learning terms versus inferential statistics terms

What are all these archaic, outmoded and confusing terms?

What are all these new fangled and confusing terms?

- Feature
- Input
- Target
- Object



- 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.



Does Machine Learning Work?

Distinguish apple from orange



How Does Machine Learning Work?

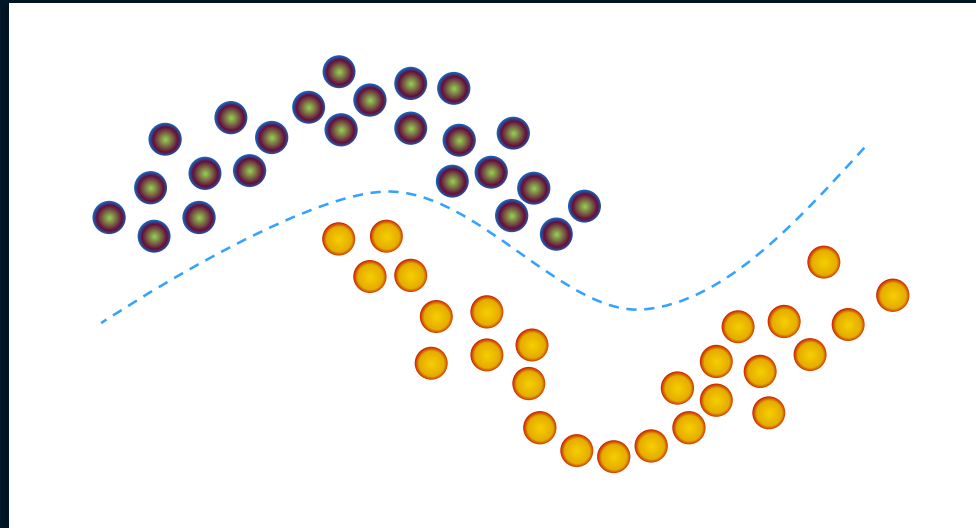
Distinguish Granny Smith apple from Fuji apple



How Does Machine Learning Work?

Supervised Learning

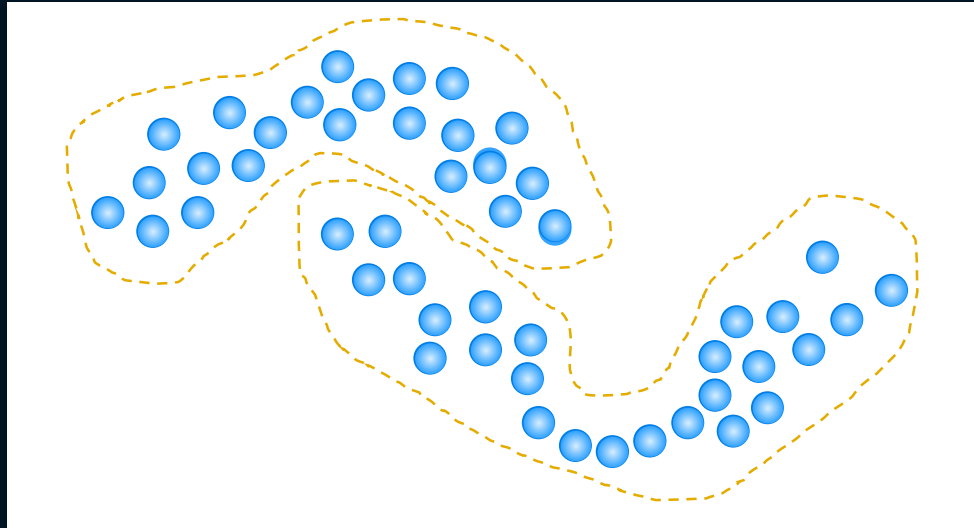
Trained on labeled examples



How Does Machine Learning Work?

Unsupervised Learning

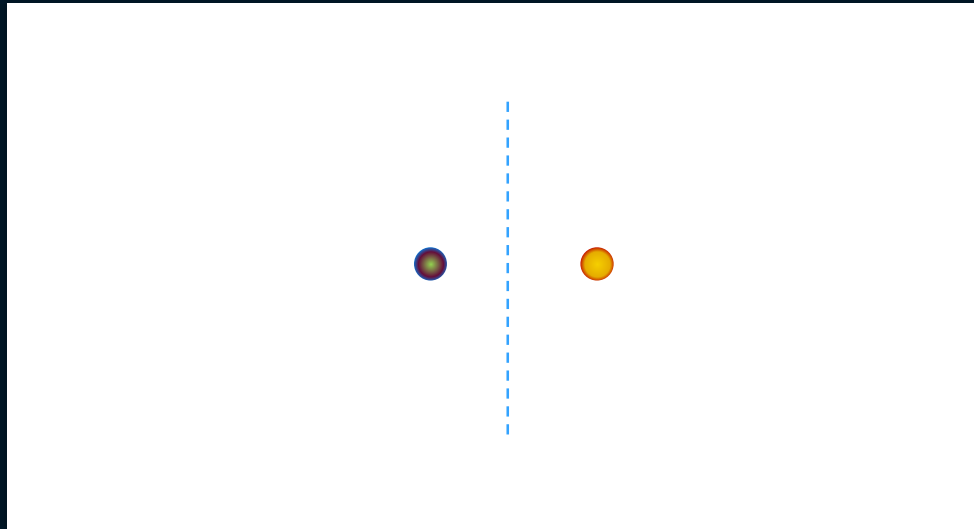
Trained on unlabeled examples



How Does Machine Learning Work?

Semi-Supervised Learning

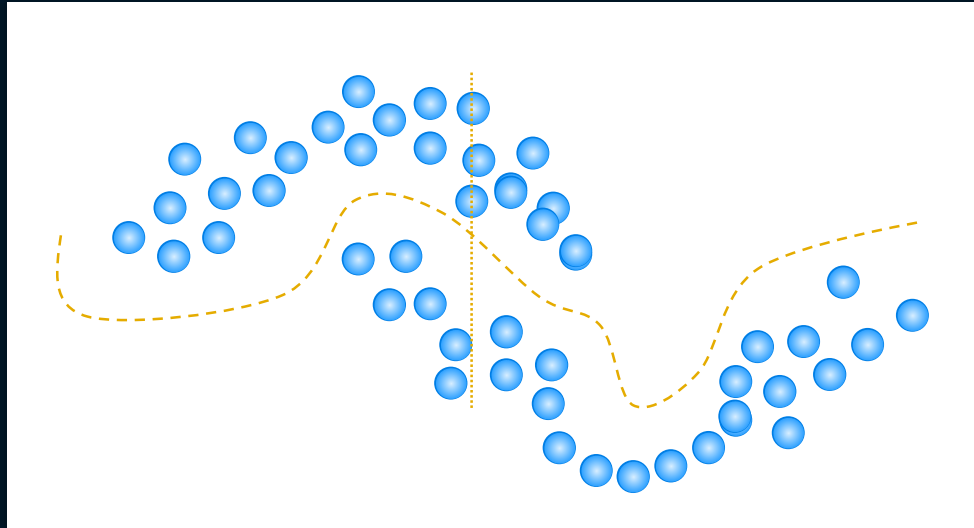
Use labeled and unlabeled observations



How Does Machine Learning Work?

Semi-Supervised Learning

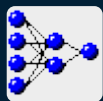
Use labeled and unlabeled observations



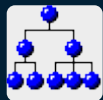
How Does Machine Learning Work?

Not New for SAS

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



Neural Networks



Decision Trees



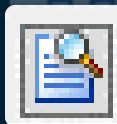
Random Forests



Clustering



Gradient Boosting



Text Analytics

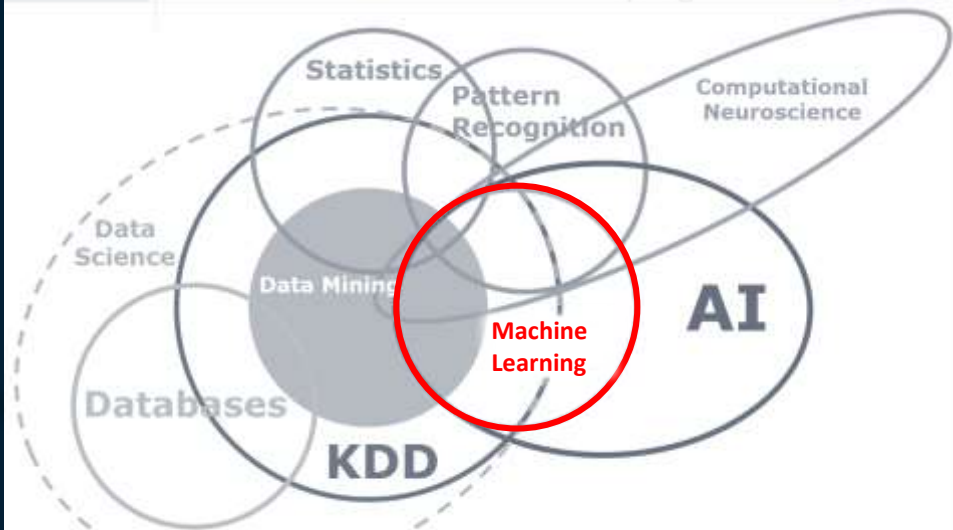


Regression

PROC DISCRIM (K-nearest-neighbor discriminant analysis)

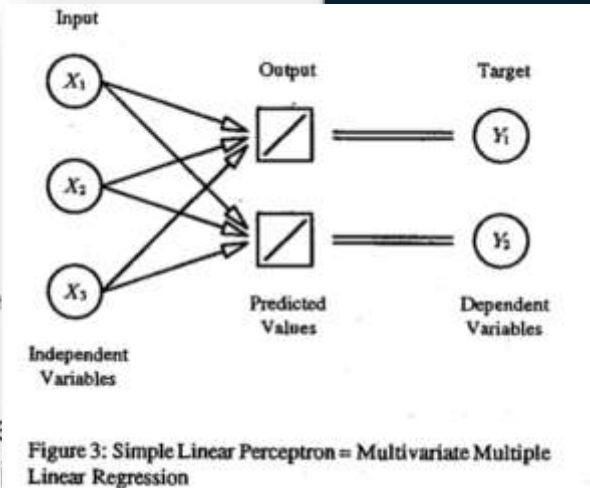
– Dr. James Goodnight, SAS founder and CEO, **1979**

- 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

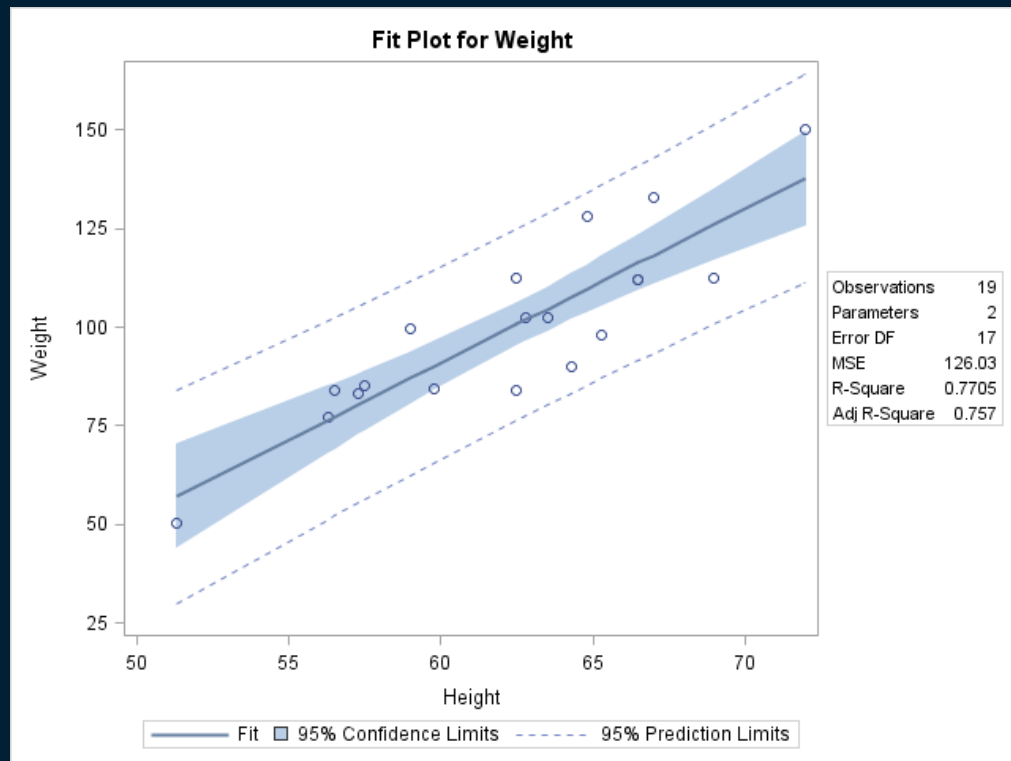
- Neural networks
- Decision trees
- Random forests
- Associations and sequence discovery
- Gradient boosting and bagging
- Support vector machines
- Nearest-neighbor mapping
- k-means clustering
- Self-organizing maps
- Local search optimization techniques such as Genetic algorithms
- Regression
- Expectation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

Machine Learning Algorithms

Regression

What Is It?

- Used to identify the relationship between a dependent variable and one or more independent variables
- Many types – linear, logistic, quantile, polynomial, stepwise, ridge, lasso, ElasticNet, etc...
- Oldie but goodie



Decision Trees

What Is It?

- Linear separation of data using “if then else” logic
- Separation is performed via an exhaustive search of splitting points for each variable.
- Many different architectural variations based on the above architecture
- Users might refer to them as
 - CHAID Trees
 - CART Trees
 - C4.5 Trees
 - C5.0 Trees.
 - Each of the above is simply a variation on

Decision Tree

Statistics	Train	Validation
1:	2.00%	2.00%
0:	98.00%	98.00%
Count:	13549	5808

Saving Balance

<2500.0000 or Missing

Statistics	Train	Validation
1:	1.46%	1.45%
0:	98.54%	98.55%
Count:	12083	5117

>=2500.0000

Statistics	Train	Validation
1:	6.43%	6.09%
0:	93.57%	93.91%
Count:	1466	691

Money Market Balance

Saving Balance

<7000.0000 or Missing

>=7000.0000

Statistics	Train	Validation
1:	1.18%	1.18%
0:	98.82%	98.82%
Count:	11165	4689

Statistics	Train	Validation
1:	4.89%	4.39%
0:	95.11%	95.61%
Count:	918	428

<8700.0000 or Missing

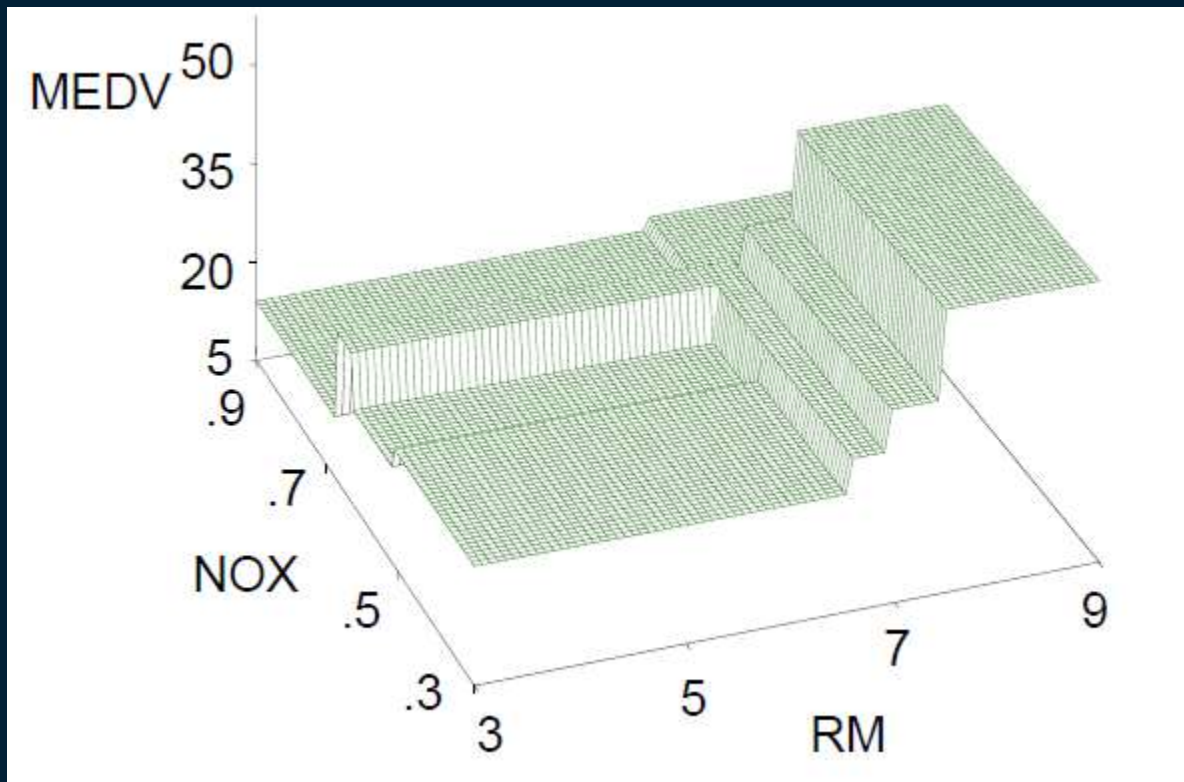
>=8700.0000

Statistics	Train	Validation
1:	4.82%	4.61%
0:	95.18%	95.39%
Count:	972	427

Statistics	Train	Validation
1:	9.60%	8.50%
0:	90.40%	91.50%
Count:	494	264

Decision Trees

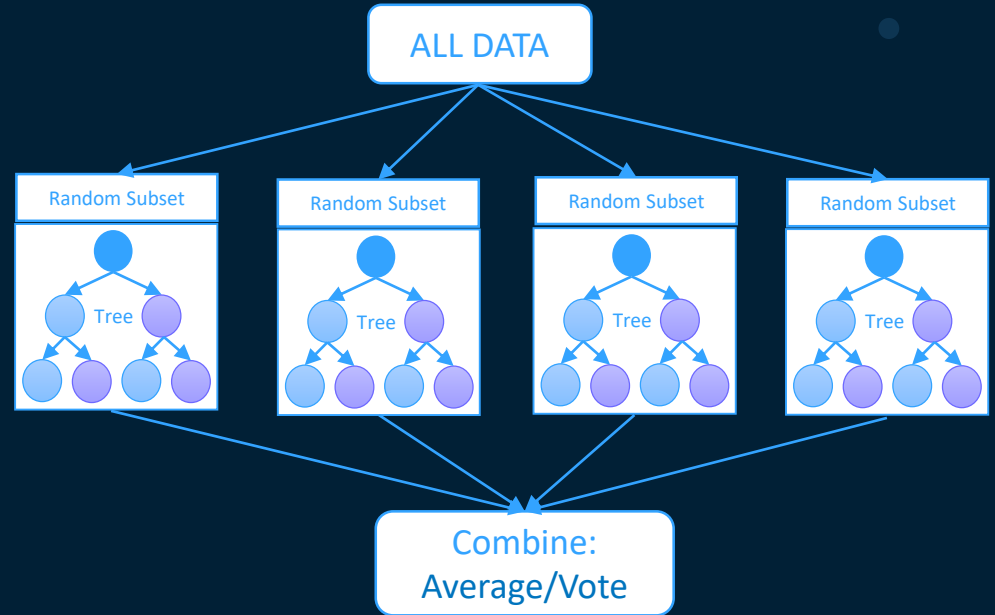
Multivariate Step Function



Random Forest

What Is It?

- A combination of several “decision trees.”
- A random forest consists of a forest of fully trained decision trees.
- The random forest averages the output of all the decision trees in the “forest.”



Random Forest

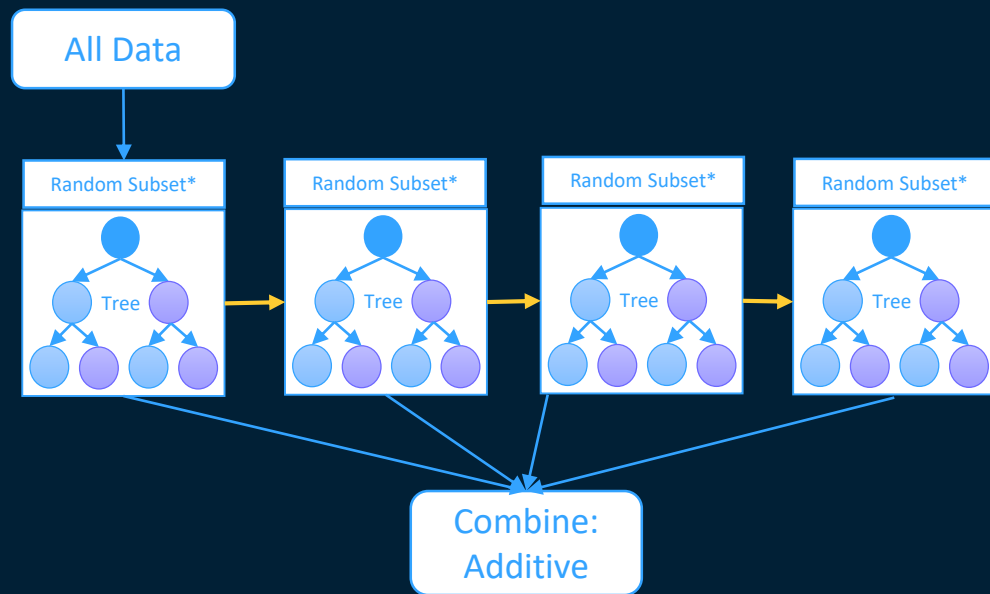
Algorithm

- Select a number of trees in the random forest.
- For each tree in the forest, use the following split algorithm:
 - Select a random sample of data.
 - Select a random subset of variables.
 - Determine the best split from the sample of data and the sample of variables.
 - Keep selecting random data and random subsets of variables until the maximum number of trees is trained.
- When all the trees are built, the prediction is the average of all trees.

Gradient Boosting

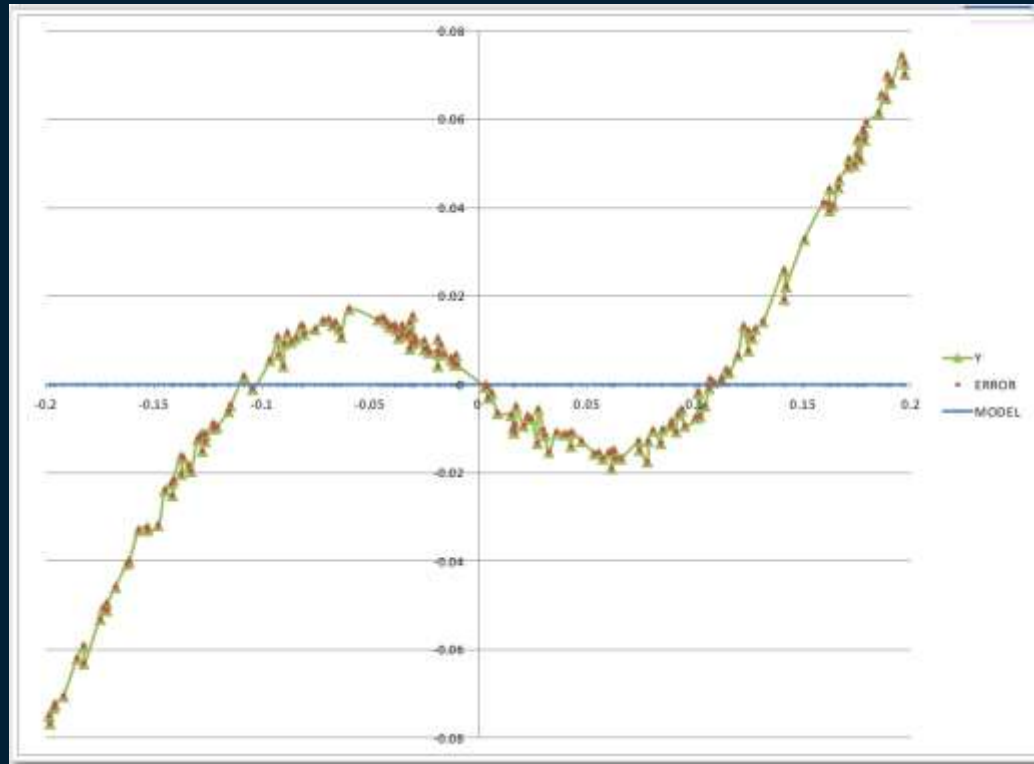
What Is It?

- A combination of several “decision trees.”
- Gradient boosting consists of a **forest** of **small** decision trees (“**shrubs**”, “stumps”).
- Each **shrub** is poor at predicting target, but each subsequent shrub tries to fit the remaining error.
- Eventually converges to good solution.



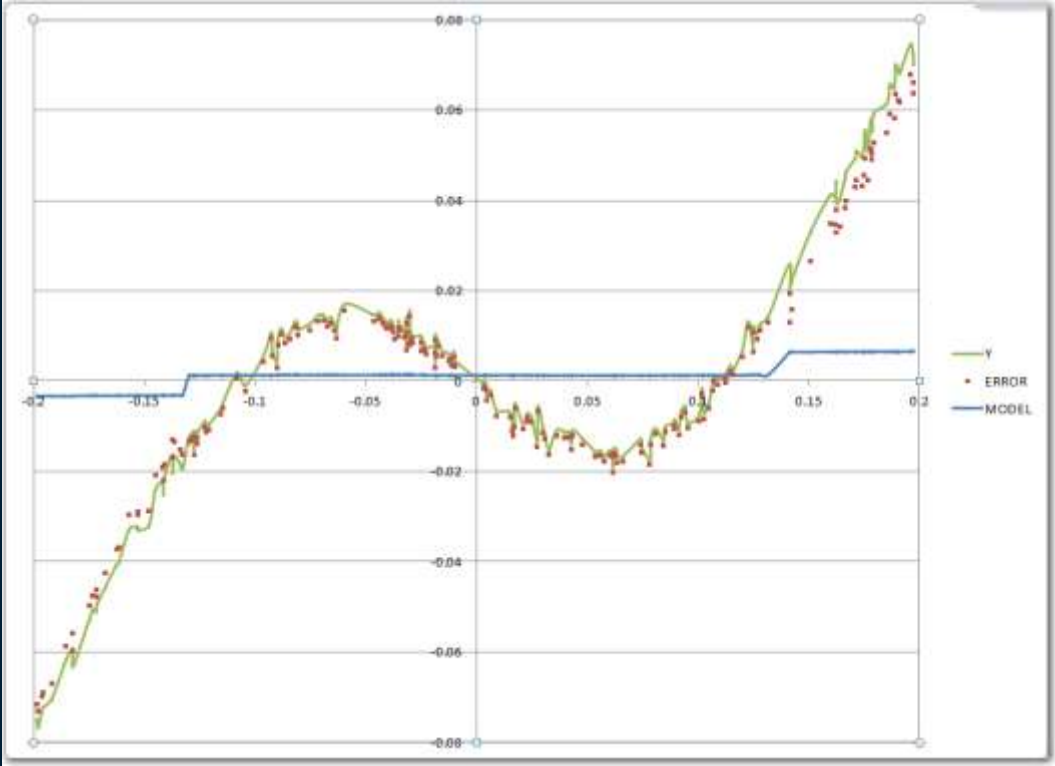
Gradient Boosting

Example: Iterations=0



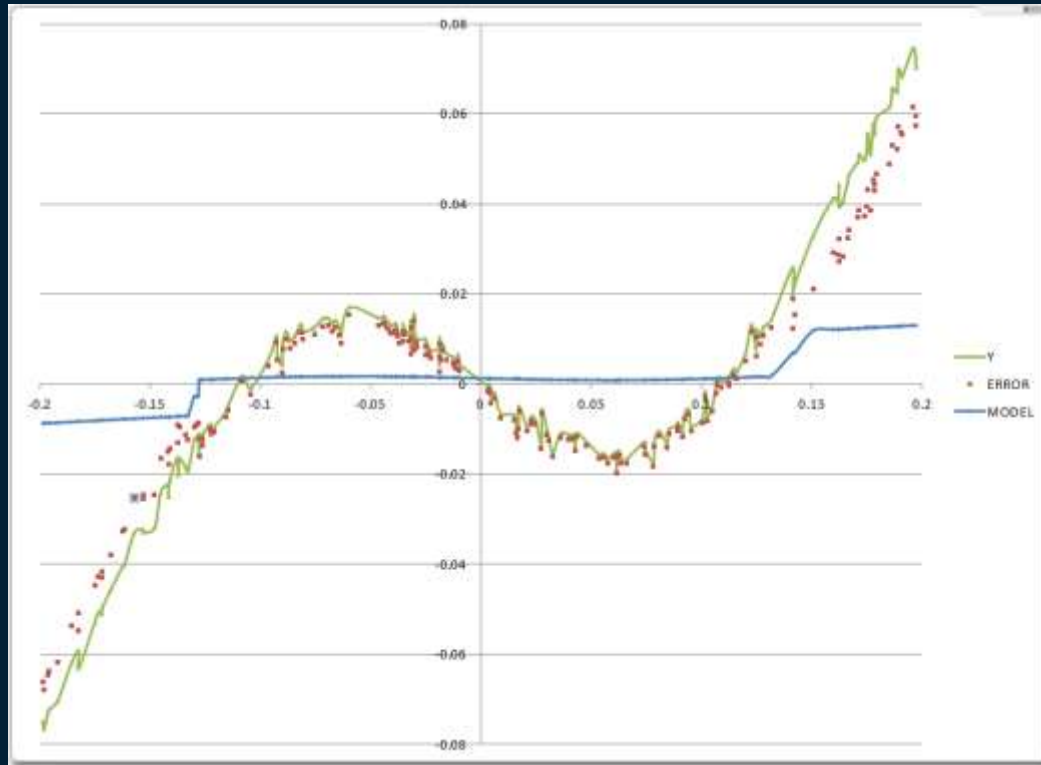
Gradient Boosting

Example: Iterations=1



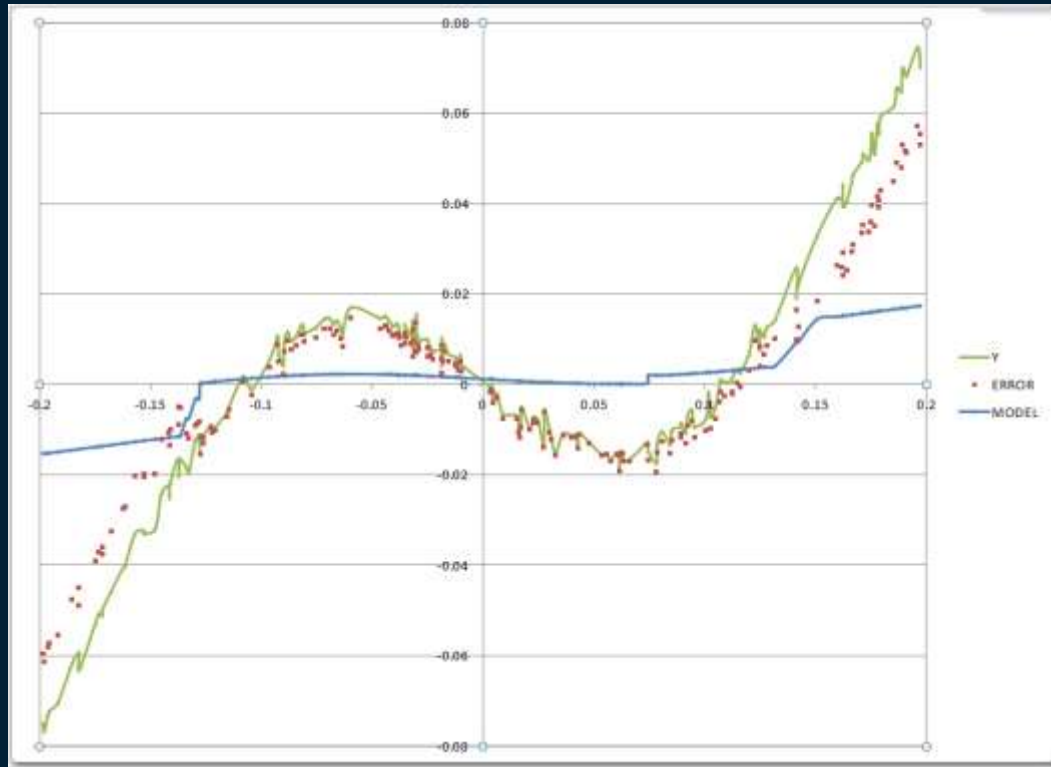
Gradient Boosting

Example: Iterations=10



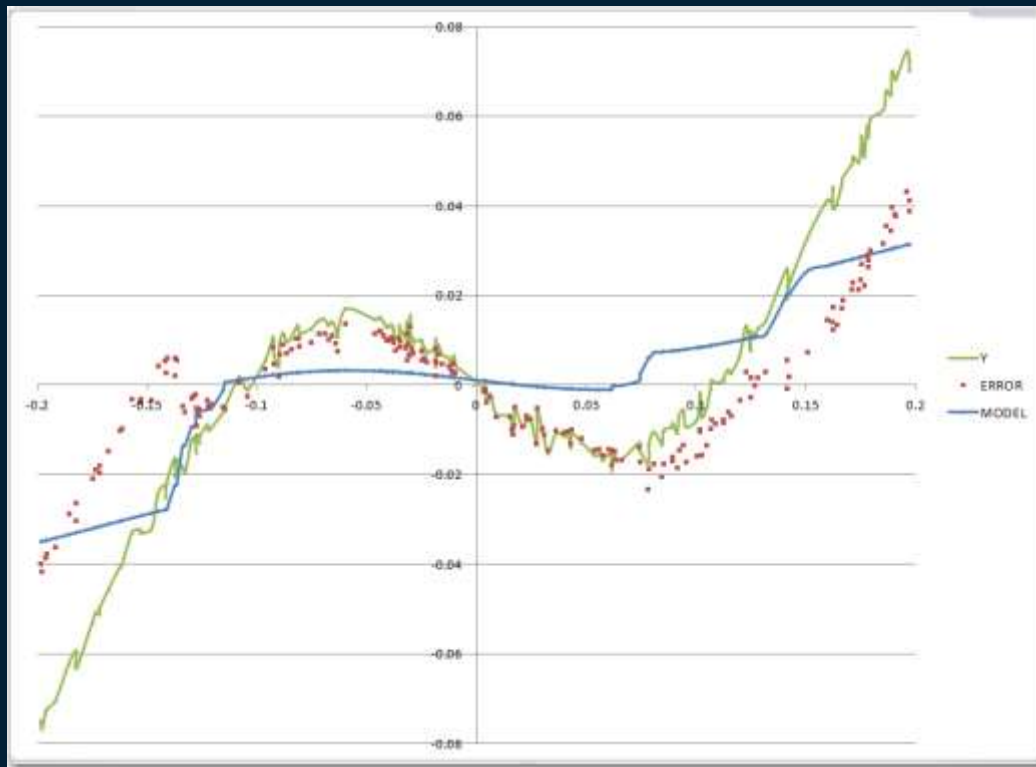
Gradient Boosting

Example: Iterations=25



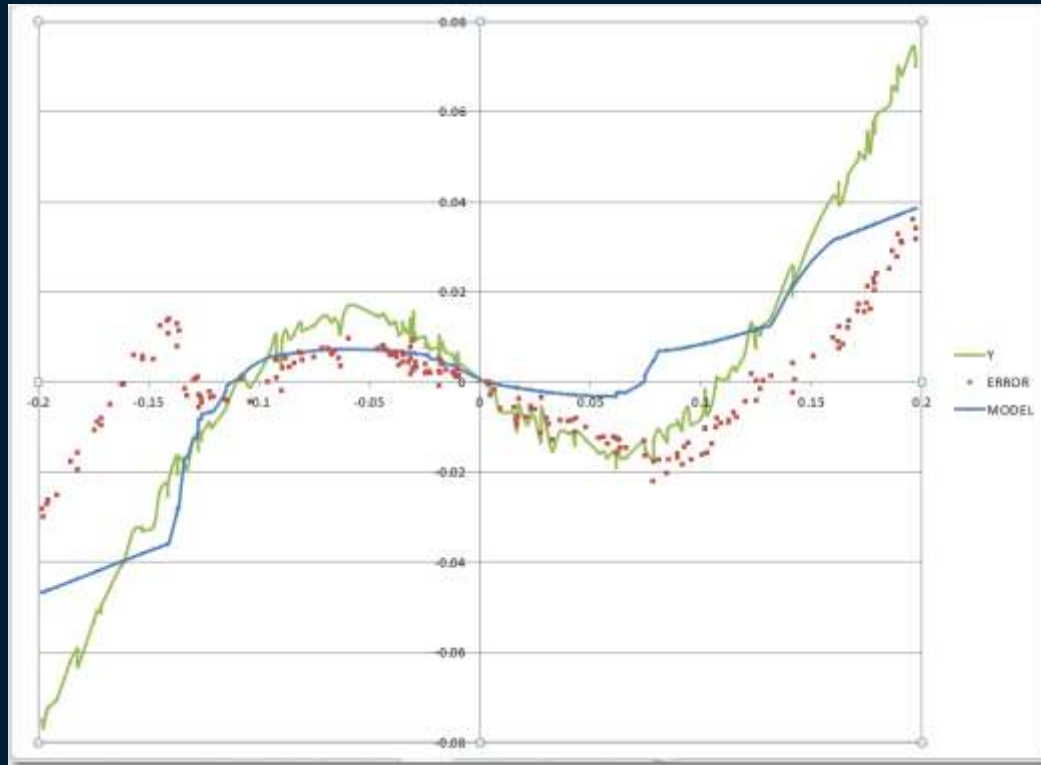
Gradient Boosting

Example: Iterations=50



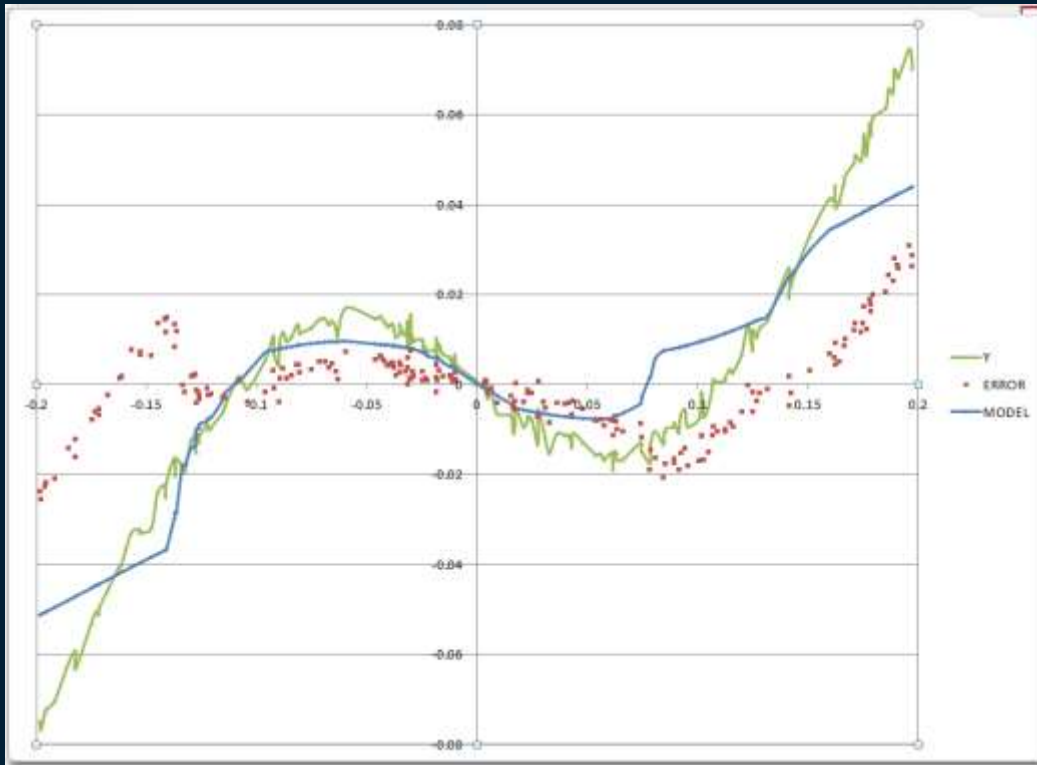
Gradient Boosting

Example: Iterations=75



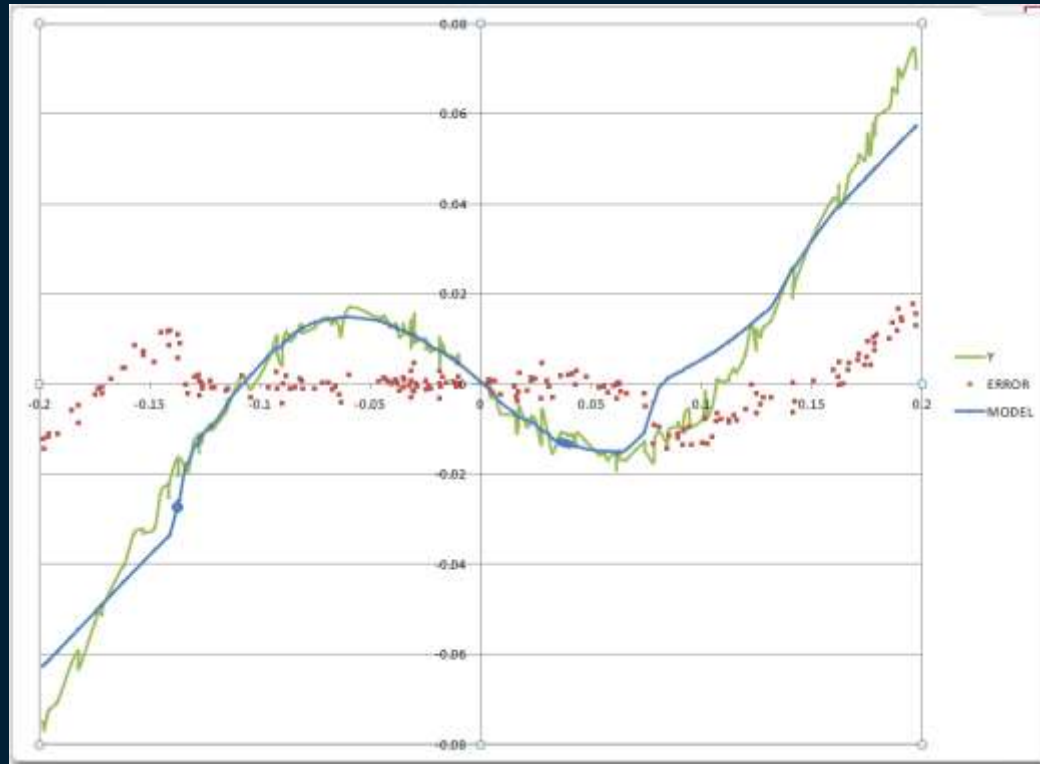
Gradient Boosting

Example: Iterations=100



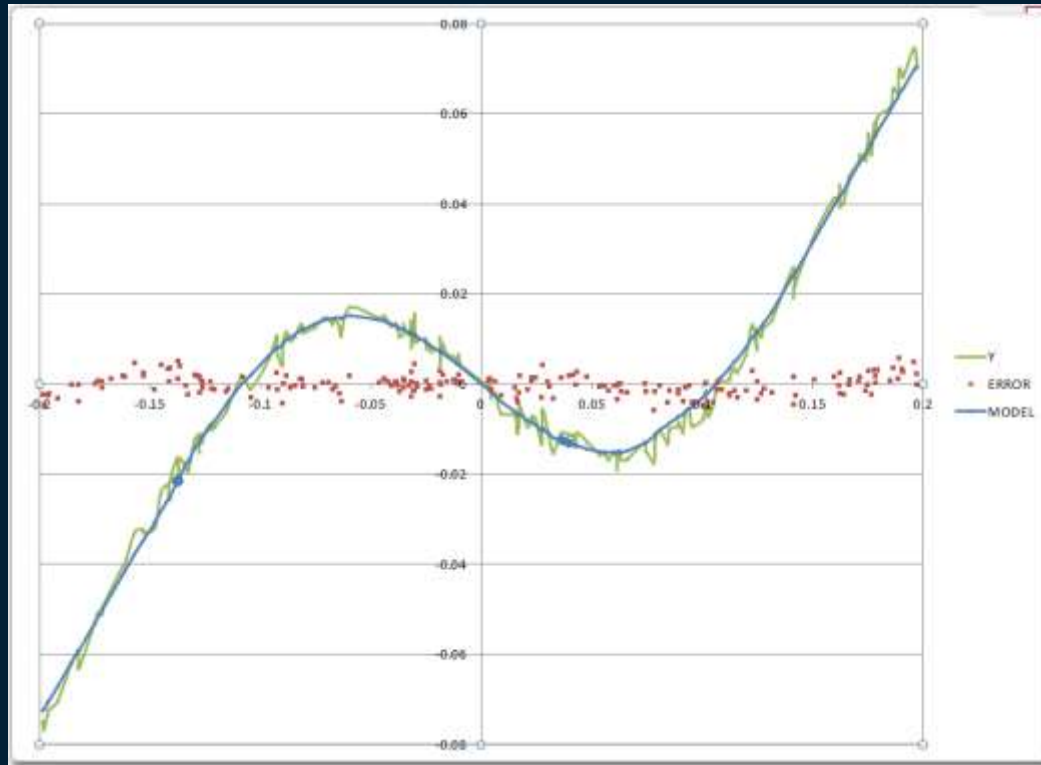
Gradient Boosting

Example: Iterations=200

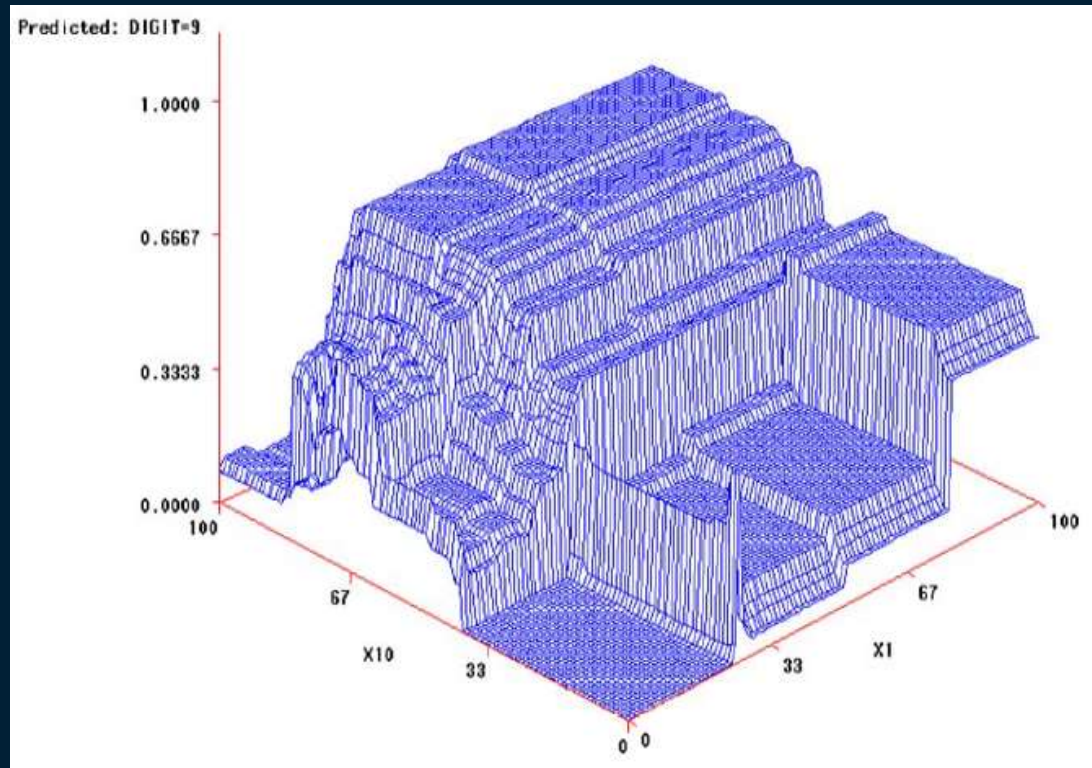


Gradient Boosting

Example: Iterations=300

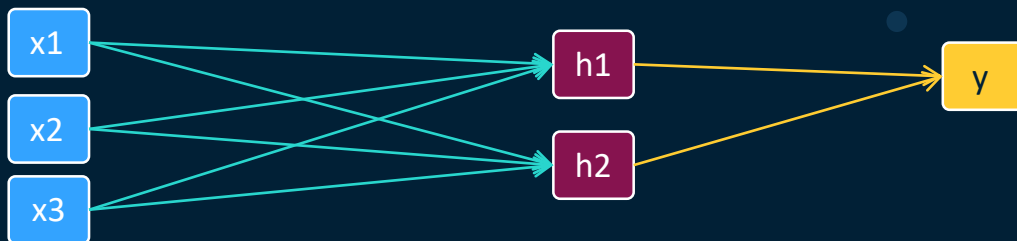


Gradient Boosting



Neural Network

What Is It?



- Non-linear relationship between inputs and output
- Prediction more important than ease of explaining model
- Requires a lot of training data
- Users can specify the number of hidden layers, the number of hidden neurons, and associated activation functions for each layer
- Users can configure Input and Target Standardizations, Target Error, and Activation Functions

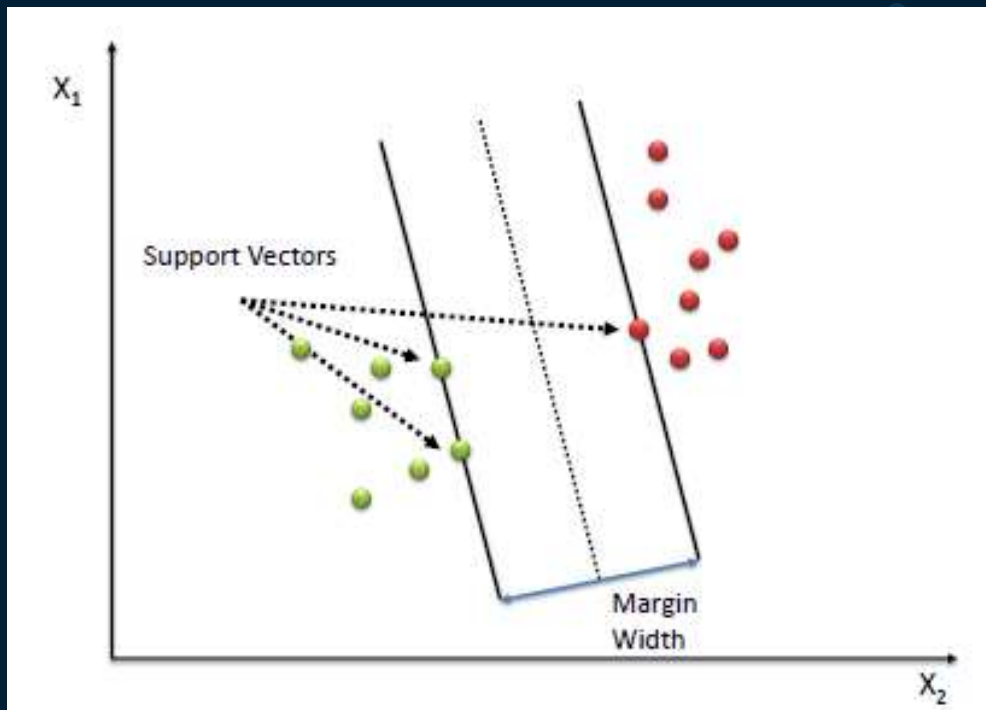
Many types...

- Feedforward Neural Network
- Radial Basis Function Neural Network
- Multilayer Perceptron
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN)
- Modular Neural Network.
- Sequence-To-Sequence Models

Support Vector Machines

What Is It?

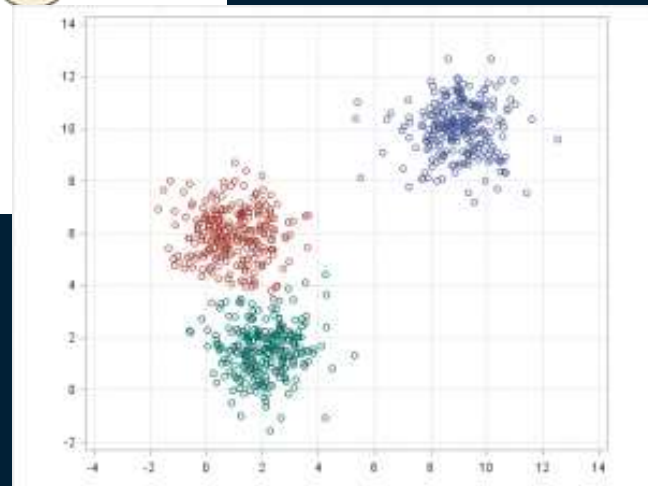
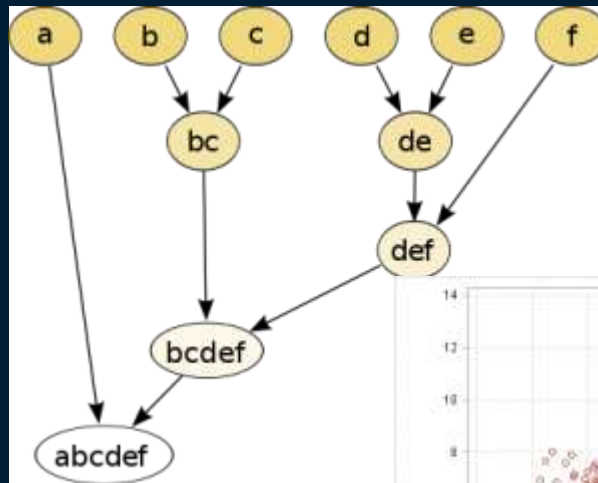
- Enables the creation of linear and nonlinear support vector machine models
- Constructs separating hyperplanes that maximize the margin between two classes
- The vectors (cases) that define the hyperplane are the support vectors
- Enables use of a variety of kernels: linear, polynomial, radial basis function, and sigmoid function. The node also provides interior point and active set optimization methods.



Clustering

What Is It?

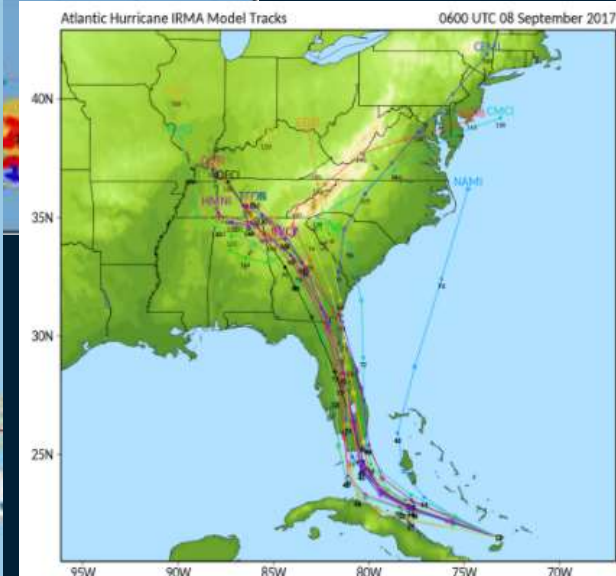
- Goal: The goal of clustering is to partition data into groups so that the observations within a group are as similar as possible to each other, and as dissimilar as possible to the observations in other groups.
- Many types - Hierarchical, k-means, SOM, etc..



Ensemble Modeling

What Is It?

- **Two or more** predictive models **combined** to create a potentially more accurate model
- Works better when model predictions are uncorrelated
- Creates new models by combining the posterior probabilities (for class targets) or the predicted values (for interval targets) from multiple predecessor models.
- 3 Methods
 - Average
 - Maximum
 - Voting



SAS 9.x

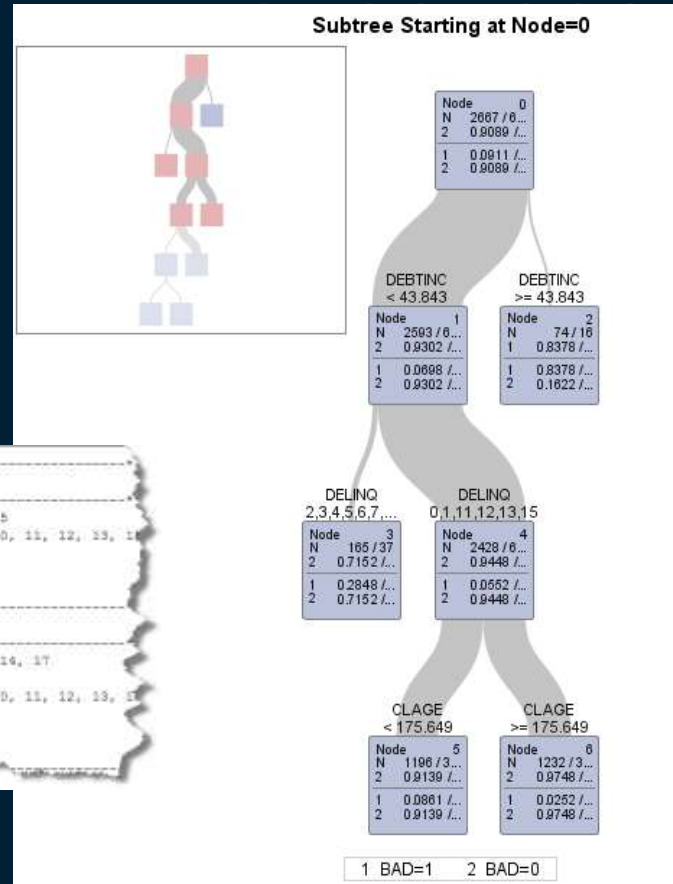
SAS/STAT and SAS Enterprise Miner

Did you know?

HPSplit Procedure

```
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;
```

```
-----  
NODE = 2  
-----  
DELINQ IS ONE OF 5, 6, 7, 8, 10, 11, 12, 13, 15  
AND DELINQ IS ONE OF 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15  
PREDICTED VALUE IS 1  
PREDICTED 1 = 0.9942( 71/76)  
PREDICTED 0 = 0.04579( 5/76)  
-----  
NODE = 4  
-----  
NINQ IS ONE OF 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 17  
AND DELINQ IS ONE OF MISSING, 1, 2, 3, 4  
AND DELINQ IS ONE OF 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15  
PREDICTED VALUE IS 1  
PREDICTED 1 = 0.8714( 61/70)  
PREDICTED 0 = 0.12861( 9/70)  
-----
```

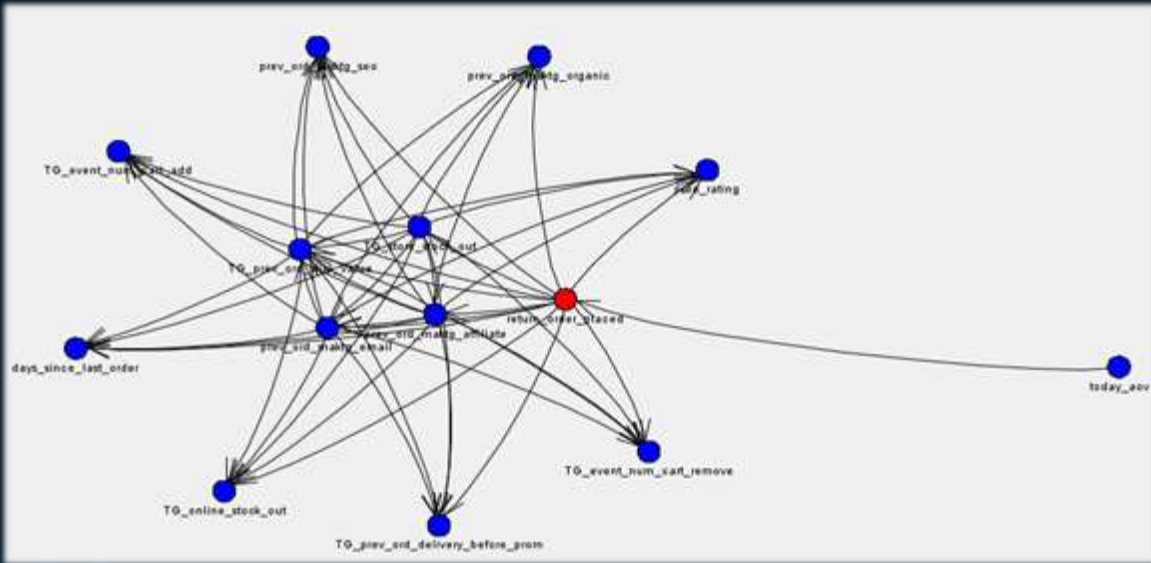


[HPSPLIT Procedure Documentation](#)

SAS Enterprise Miner

- Algorithms – basic and advanced

- 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

#	Variable	Type	Len	Label
1	BAD	Num	8	Default or seriously delinquent
10	CLAGE	Num	8	Age of oldest credit line in months
12	CLNO	Num	8	No. of trade credit lines
13	DEBTINC	Num	8	Debt to income ratio
9	DELINQ	Num	8	No. of delinquent credit lines
8	DEROG	Num	8	No. of major derogatory reports
6	JOB	Char	7	Prof/Exec/Office/Self/Other
2	LOAN	Num	8	Amount of current loan request
3	MORTDUE	Num	8	Amount due on existing mortgage
11	NINQ	Num	8	No. of recent credit inquiries
5	REASON	Char	7	Home improvement or Debt Consolidation
4	VALUE	Num	8	Value of current property
7	YOJ	Num	8	Years on current job

SAS 9.4 Machine Learning Demo

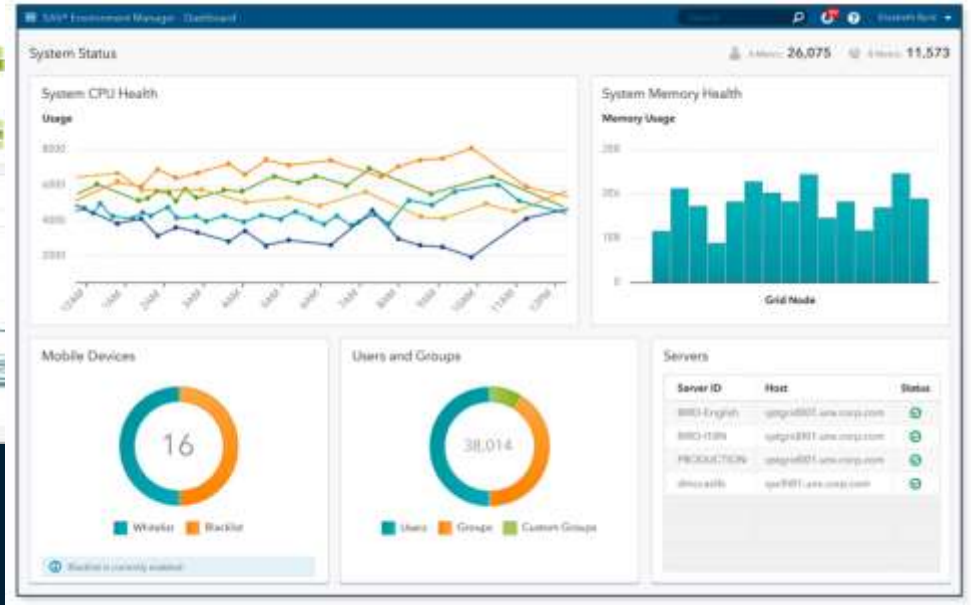
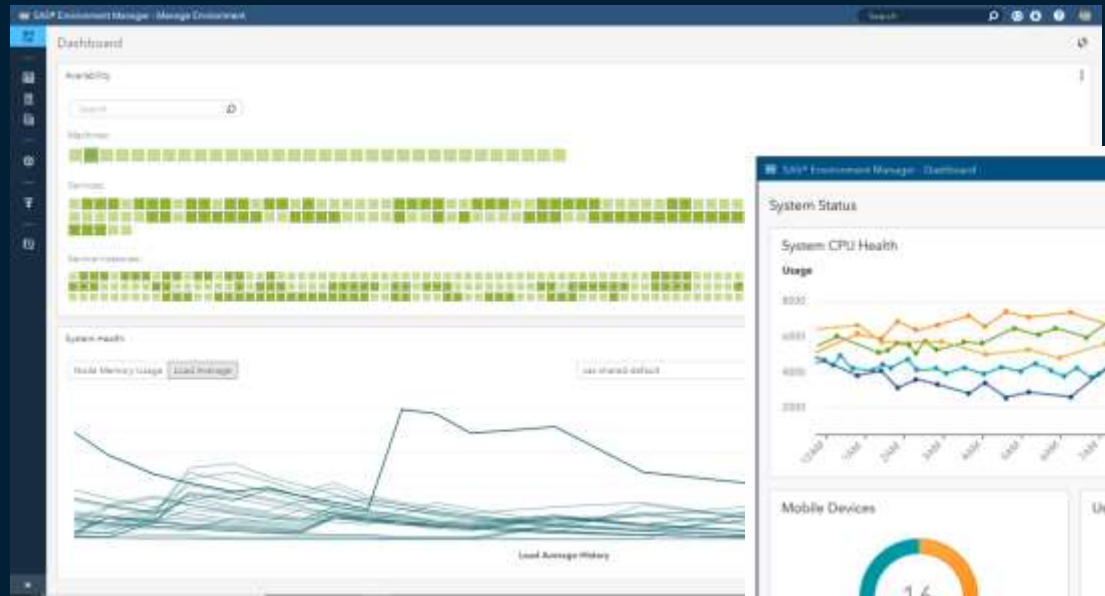
SAS Enterprise Miner

SAS Viya

SAS Visual Statistics and
SAS Visual Data Mining and Machine Learning

What is SAS Viya?

Viya is a cloud-enabled, in-memory analytics engine that provides quick, accurate and reliable analytical insights.



SAS Viya Products

SAS Viya takes 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 Analytics
- SAS Optimization
- SAS Econometrics
- SAS Model Manager
- SAS Data Preparation
- SAS Visual Investigator
- SAS Business Analytics
- SAS Intelligent Decisioning
- SAS Cybersecurity
- SAS Detection and Investigation
- SAS Event Stream Processing
- And more...

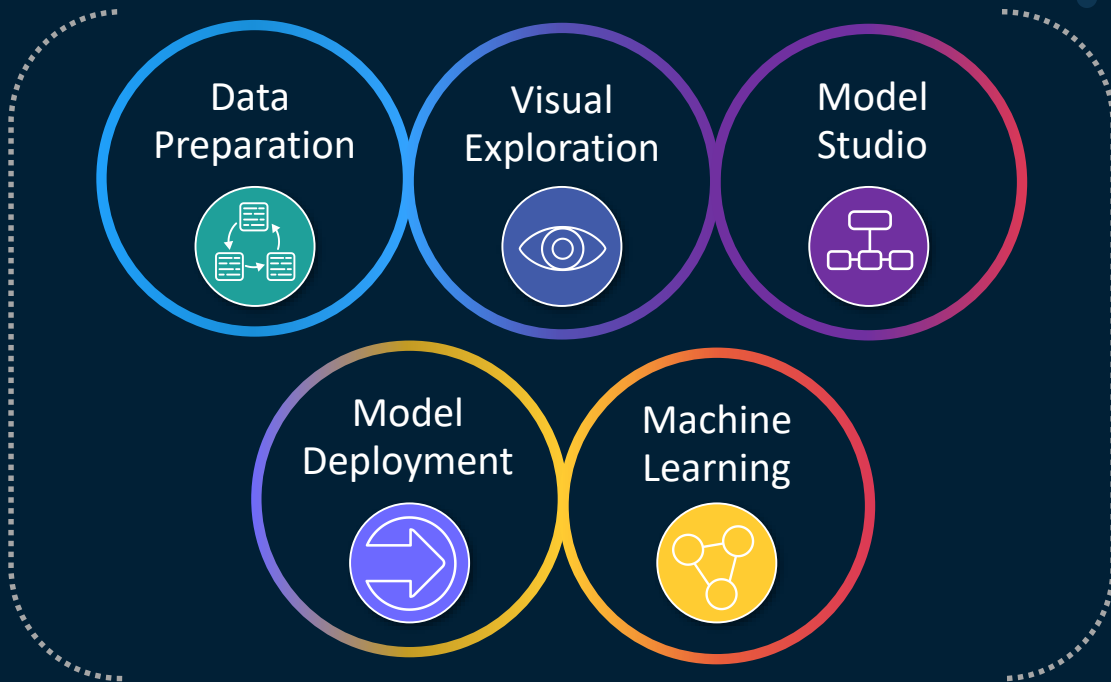


SAS® Visual Data Mining and Machine Learning

Visual "drag & drop"
Interface

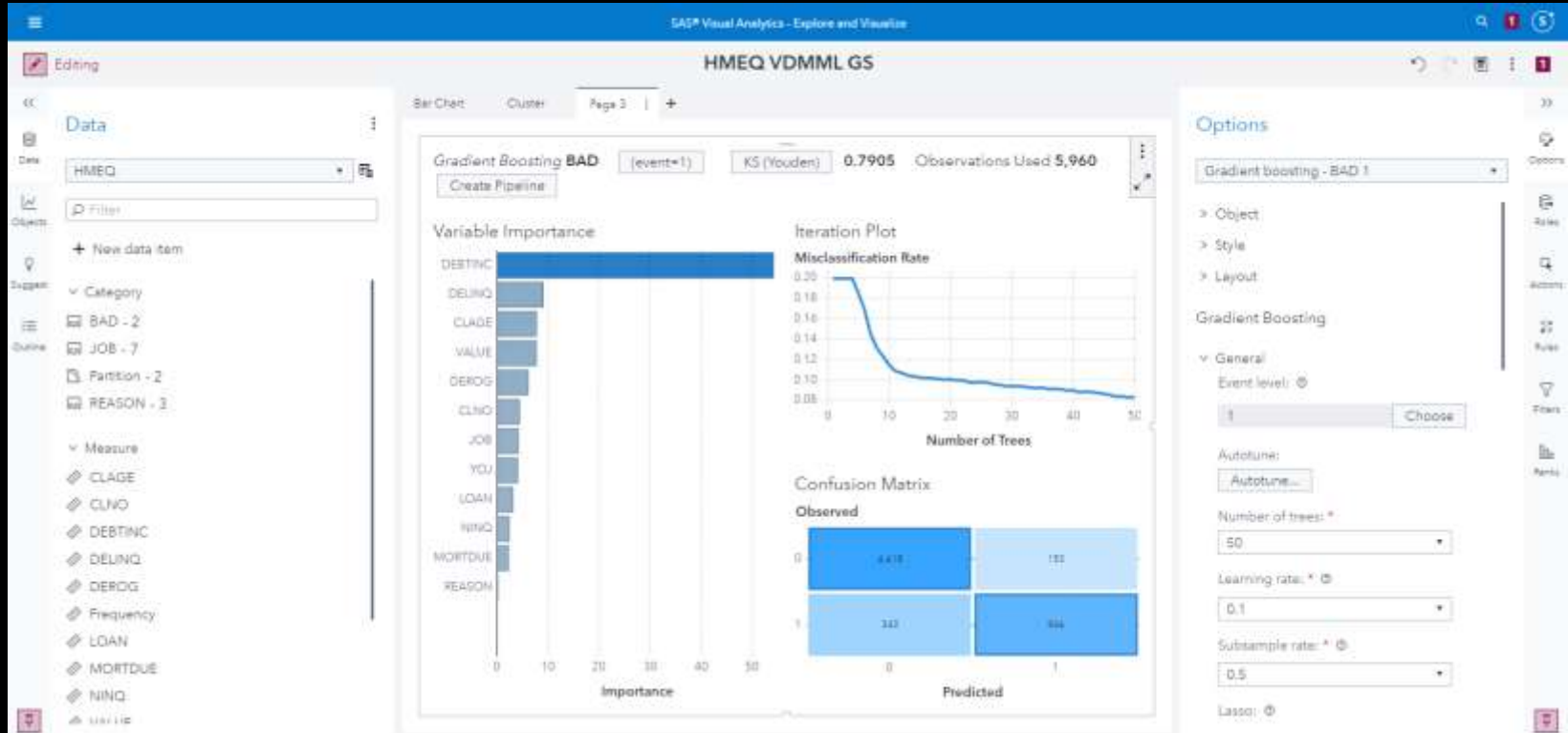


Programming
Interface



Interfaces

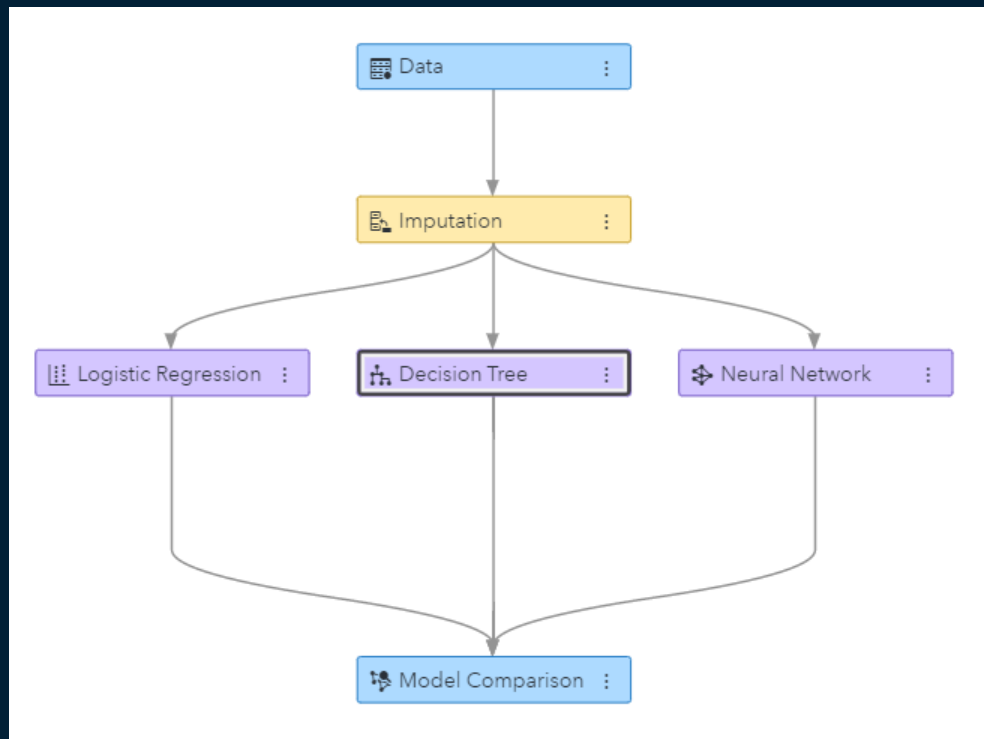
Building a Model from Scratch in the Visual Reporting Interface



Interfaces

Build Models Using Pipelines in Model Studio

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



[Example Code for Pipeline](#)

Interfaces

Building a Model Using SAS Studio Tasks

The screenshot displays the SAS Studio interface for configuring a Gradient Boosting task. The left sidebar shows the 'Tasks' pane with 'Gradient Boosting' selected under 'SAS Viya Machine Learning'. The main workspace is divided into three panes: 'DATA', 'ROLES', and 'Code'.

DATA Pane:

- Dataset: PUBLIC.HMEQ
- Partition Data: Input data contains training data. Includes:
 - Validation data
 - Test data
- Identify partitions: Specify a sample proportion (dropdown menu)
- Proportion of validation cases: 0.30
- Random number seed

ROLES Pane:

- Target:
 - Use a nominal target
 - Use an interval target
- Nominal target: BAD

Code Pane:

```
1 /*
2 *
3 * Task code generated by SAS® Studio 5.2
4 *
5 * Generated on '2/9/20, 1:41 PM'
6 * Generated by 'sasdemo'
7 * Generated on server 'sasserver'
8 * Generated on SAS platform 'Linux LIN X64 3.10.0-957.27.1.el7.x86_64'
9 * Generated on SAS version 'V.01.05PBP110619'
10 * Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit
11 * Generated on web client 'http://10.96.17.31/SASStudioW/main/locale=en_US'
12 */
13
14 ods noproctitle;
15
16 proc gradboost data=PUBLIC.HMEQ;
17   partition fraction(validate=0.3);
18   target BAD / level=nominal;
19   input LOAN MONTHLY VALUE Y01 DEROG DELINQ CLAGE NIMQ CLMO DEBTINC /
20     level=interval;
21   input REASON JOB / level=nominal;
22 run;
```

Interfaces

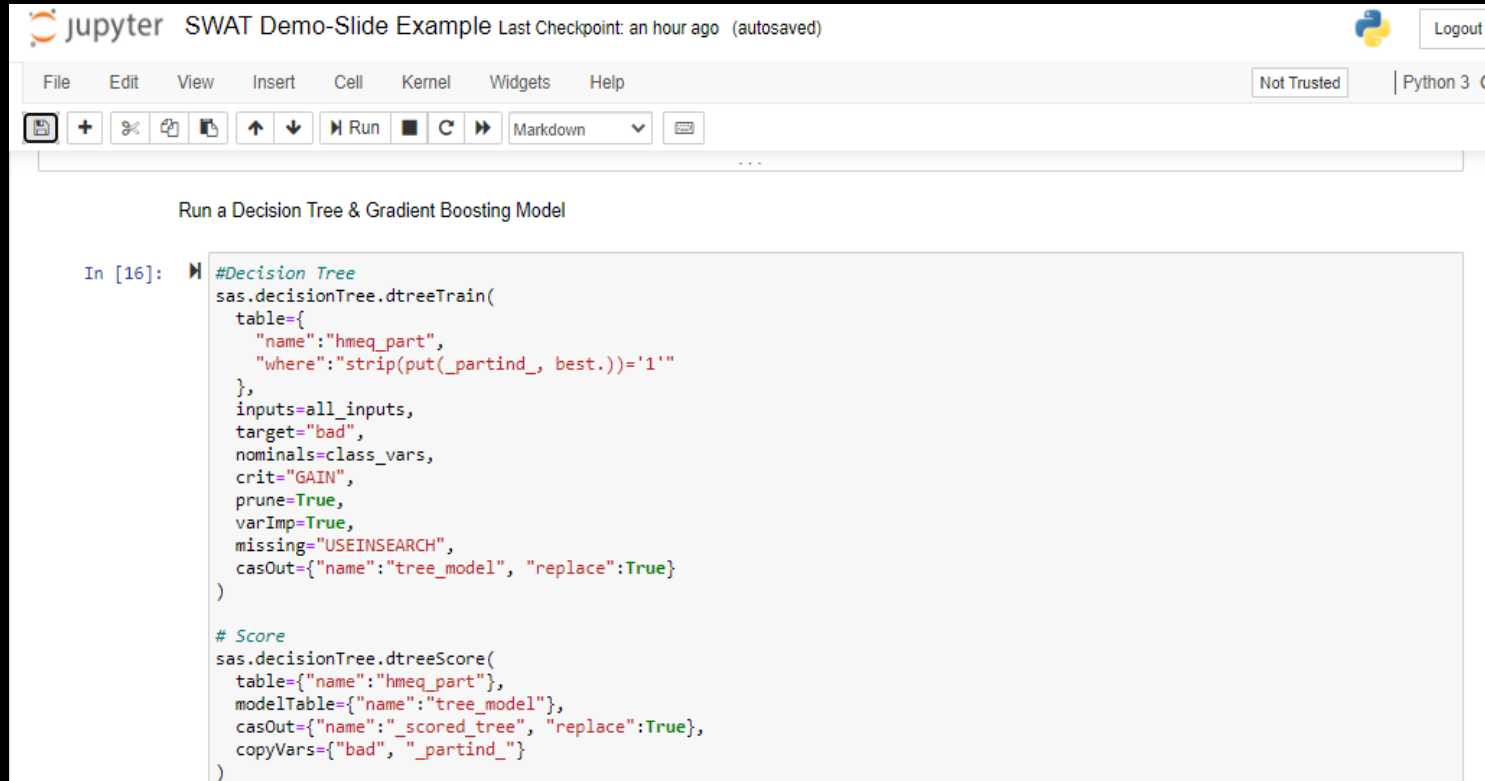
Building a Model Using SAS Studio Snippets

The screenshot displays the SAS Studio interface. On the left, the 'Snippets' pane shows a tree view of available snippets, with 'Supervised Learning' selected. The main editor shows the following SAS code:

```
115     by _NAME_;
116 run;
117
118 /* Variance explained by iteration plot */
119 proc sgplot data=out_iter_trans;
120     title "Variance Explained by Iteration";
121     yaxis label="Variance Explained";
122     vbar iteration / response=CGI group=_NAME_;
123 run;
124
125 /*-----*/
126 /* Build a predictive model using Random Forest */
127 /*-----*/
128 proc forest data=&caslibname._prepped ntrees=50 numbin=20 minleafsize=5;
129     input &interval_inputs. / level = interval;
130     input &class_inputs. / level = nominal;
131     target &target / level = nominal;
132     partition rolevar=_partind (train='1' validate='0');
133     code file="&outdir.\\forest.sas";
134     ods output FitStatistics=fitstots;
135 run;
136
137 /*-----*/
138 /* Score the data using the generated model */
139 /*-----*/
140 data &caslibname._scored_forest;
141     set &caslibname._prepped;
142     %include "&outdir.\\forest.sas";
143 run;
144
```

Interfaces

Building a Model Using Open Source



The screenshot shows a JupyterLab interface with the title "SWAT Demo-Slide Example" and a "Last Checkpoint: an hour ago (autosaved)" status. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a code editor. The code in the editor is as follows:

```
In [16]: #Decision Tree
sas.decisionTree.dtreeTrain(
  table={
    "name":"hmeq_part",
    "where":"strip(put(_partind_, best.))='1'"
  },
  inputs=all_inputs,
  target="bad",
  nominals=class_vars,
  crit="GAIN",
  prune=True,
  varImp=True,
  missing="USEINSEARCH",
  casOut={"name":"tree_model", "replace":True}
)

# Score
sas.decisionTree.dtreeScore(
  table={"name":"hmeq_part"},
  modelTable={"name":"tree_model"},
  casOut={"name": "_scored_tree", "replace":True},
  copyVars={"bad", "_partind_"})
```



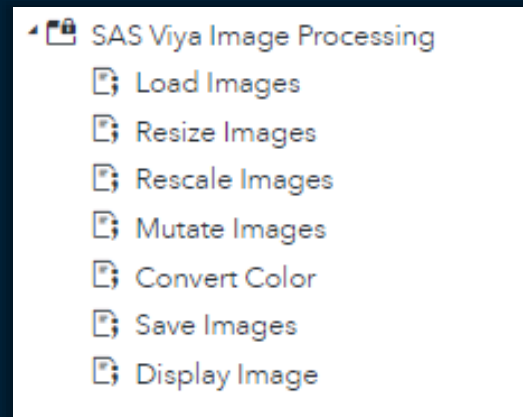
SAS Viya Machine Learning Demo

Visual Interface, Pipelines and Open Source

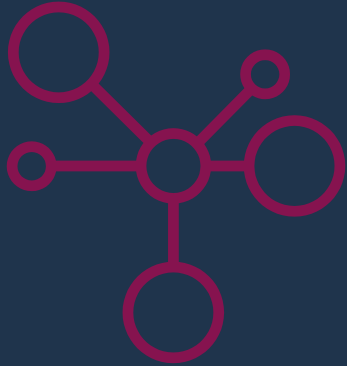
Other Features in SAS VDMML

Additional Analytical Algorithms and Options

- Tensor Factorization
- Neural Network Autoencoders
- Clustering mixed variables
- Deep Learning algorithms - Deep forward neural networks (DNNs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs)
- Bayesian Network
- Market Basket Analysis
- Image Processing (CAS Actions)
 - Load images recursively & at random
 - Retrieve Image labels across all folders when importing
 - Convert image table action (wide format)
 - Support image processing with Deep Learning



How does SAS support Machine Learning?



Review

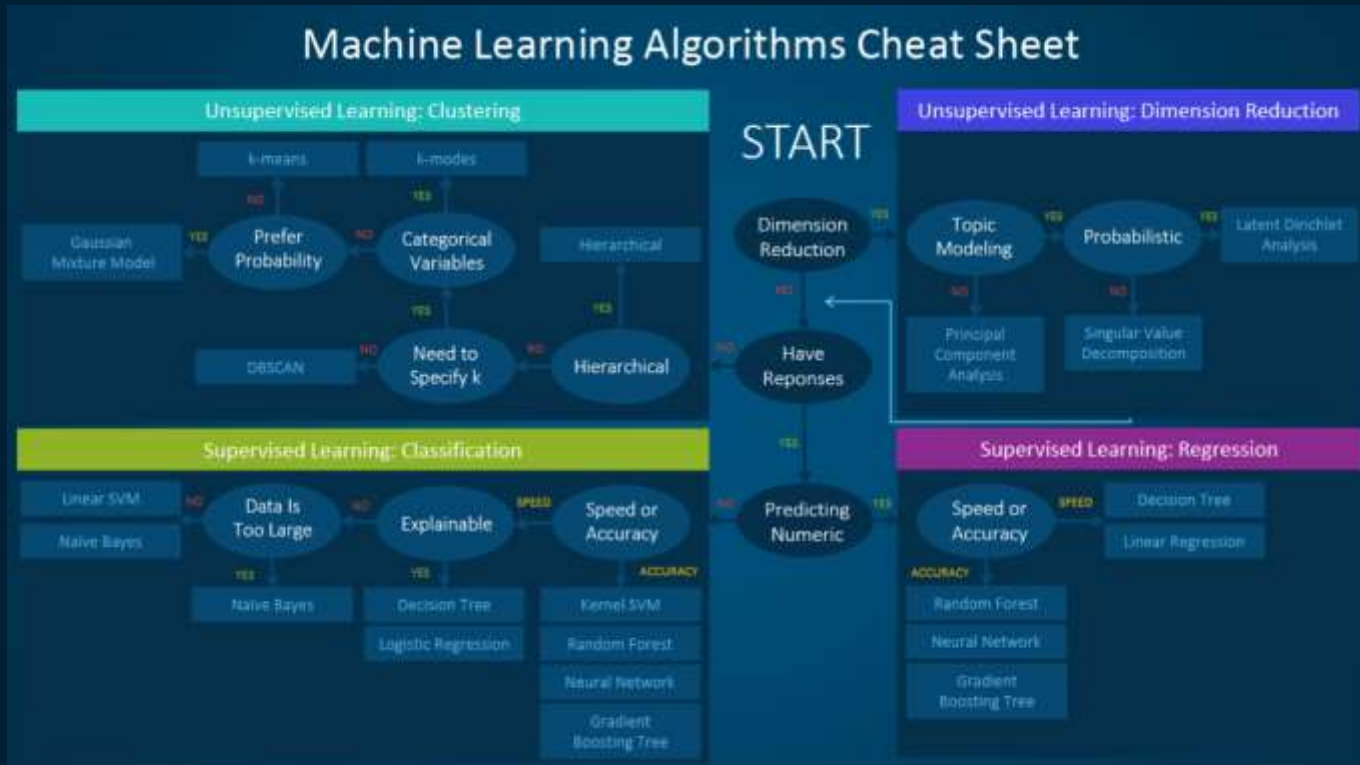
- What is Machine Learning?
- Terminology and key characteristics
- Introduction to Decision Trees, Random Forest, Gradient Boosting, Neural Networks, and k-means Clustering
- How you can use machine learning in SAS
- Examples in SAS 9.x and SAS Viya

Resources

Where to learn more

Model Selection

Machine Learning Algorithms Cheat Sheet



[Access Here](#)

Recommended Resources

An Overview of SAS® Visual Data Mining

<https://support.sas.com/resources/papers/proceedings17/SAS1492-2017.pdf>

Video - Automated Machine Learning at Scale

http://www.sas.com/en_us/webinars/automated-machine-learning-scale.html

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

http://www.sas.com/en_us/insights/analytics/machine-learning.html

Live web and classroom training - Big Data, Data Mining, and Machine Learning

[Big Data course](#)

SAS Tutorial

Videos

How to Choose a Machine Learning Algorithm

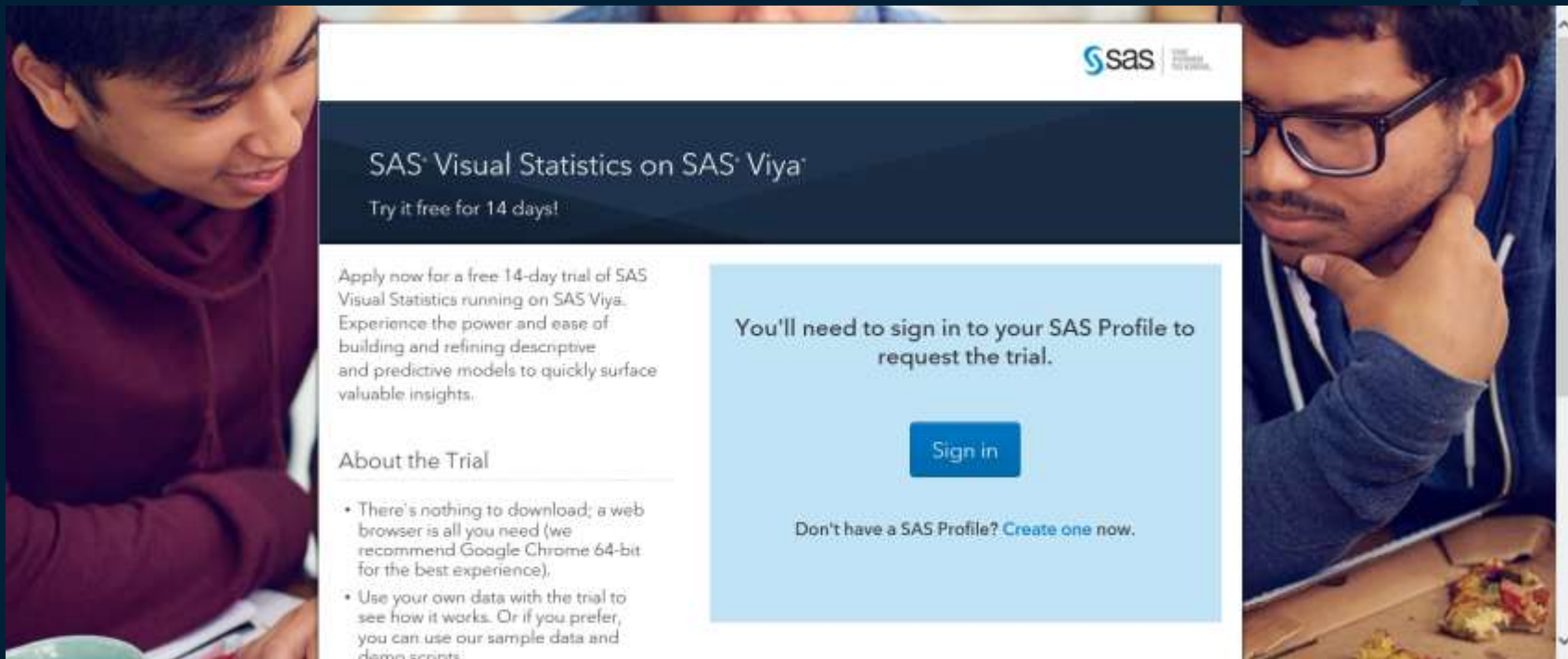
<https://youtu.be/-oZcf0QEzYM>


Transforming variables in SAS

<https://communities.sas.com/t5/SAS-Data-Mining-and-Machine/New-video-Transforming-Variables-in-SAS/m-p/710687#M8553>

SAS® Visual Statistics

[Try it before you buy!](#)



 SAS Institute Inc.

SAS® Visual Statistics on SAS® Viya®

Try it free for 14 days!

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 demo reports.

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

[Overview](#)

[Support](#)

[Free Trial](#)

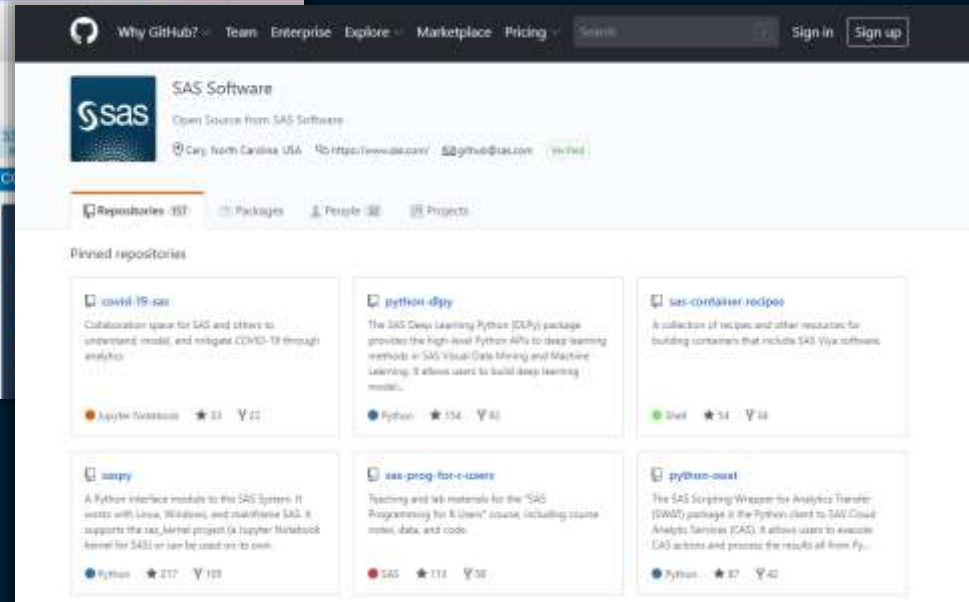
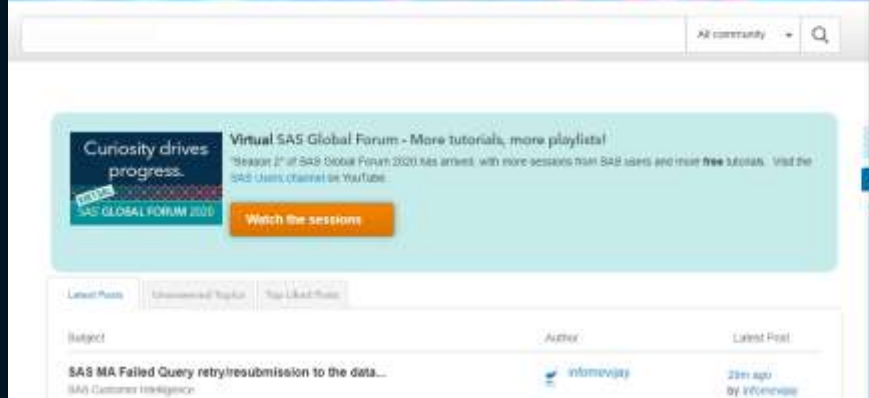
SAS® VISUAL DATA MINING AND MACHINE LEARNING

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

[Try it for free](#)



Communities



[Communities.sas.com](https://communities.sas.com)
[Github.com/sassoftware](https://github.com/sassoftware)
[Developer.sas.com](https://developer.sas.com)



Questions?

Thank you for your time and attention!

Connect with me:

LinkedIn: <https://www.linkedin.com/in/melodierush>

Twitter: @Melodie_Rush

sas.com



GLOBAL VIRTUAL EVENT

Travel to Faster, Trusted Decisions in the Cloud

February 11 | [Learn More](#)



Bernard Marr
Futurist, Influencer for Business
and Technology



Shalini Kantayya
Director, Coded Bias



Ritu Jyoti
Vice President, Artificial
Intelligence Research, IDC

The logo for sas viya, featuring a stylized 'S' icon followed by the text 'sas viya' in a lowercase, sans-serif font.

When curious minds compete, the world wins

2021

HACKATHON

#HackinSAS. Much more than a competition!

[Register Now](#)



In partnership with



Microsoft intel.

```
/*retrieve service endpoint*/
%let BASE_URI=%sysfunc(getoption(servicesbaseurl));

/* create filenames to hold responses*/
filename rcontent temp;

/* Make request */
proc http
  url="&BASE_URI/reports/reports/char?char=visual%20element"
  method="GET"
  headers
  application/vnd.sas.collection+json";
run;

/*retrieve service endpoint*/
%let BASE_URI=%sysfunc(getoption(servicesbaseurl));

/* create filenames to hold responses*/
filename rcontent temp;

/* Make request */
```

Explore Helpful Resources

[Ask the Expert](#)

View other user webinars that provide insights into using SAS products to make your job easier.

[FREE Training](#)

Learn from home – free for 30 days. Get software labs to practice and online support if needed.

[SAS Support Communities](#)

Ask questions, get answers and share insights with SAS users.

[SAS Analytics Explorers](#)

An exclusive platform to collaborate, learn and share your expertise. Gain access to a diverse network to advance your career. Special rewards and recognition exclusively for SAS users.

[SAS Users YouTube Channel](#)

A plethora of videos on hundreds of topics, just for SAS users.

[Newsletters](#)

Get the latest SAS news plus tips, tricks and more.

[Users Groups](#)

Meet local SAS users, network and exchange ideas – virtually.

[SAS Profile](#)

If you haven't already done so, create your SAS Profile to access free training, SAS Support Communities, technical support, software downloads, newsletters and more.

Did you know?

SAS is the only vendor named a Leader in the top 4 analyst assessments focused on the AI and Analytics market in 2020.



Learn more at sas.com/leader

