Machine Learning:
Ready to Go Solutions or Complex Scientific Projects?

Alex Glushkovsky, BMO Financial Group

Agenda

- Scientific Learning
- Objectives
- Stability
- Feature Engineering
- Randomization
- Regularization
- Tuning
- Outliers

Machine Learning Keywords

Supervised Learning
Recurrent Neural Networks
Optimization
Feature Engineering
Hyperparameters
Support Vector Machine
ReLU
Outliers
Data Science
Unstructured Data
Big Data
GPU
Robustness
Backpropagation
Iterative Process
Activation Function
Stability
Network Layer
Overfitting
Interactions
Epochs
Partial Derivative
Cost Function
Learning Rate
Autotuning
Convergence
Stochastic Gradient Descent
Validation
Tanh
Regularization
LSTM
XOR
Sigmoid Function
Autoencoder
Accuracy
Gradient Boosting
Randomization
Random Forest
Convolutional Neural Networks
Performance Monitoring
Deep Learning
Tuning
Model Deployment
Dropout Rate
Machine Learning
Unsupervised Learning

Why Machine Learning In Business?

- Ability to address a wide spectrum of business objectives from specific experiences to profitability
- Suitable for "Big Data": thousands of features and millions of observations
- Discovers sound business insights
  - Predictive Modeling
  - Supervised/Unsupervised Segmentation
  - Patterns Recognition
  - Dimensionality Reduction
  - Outlier Detections
- Provides accurate yet robust outcomes
- Compatible with integrative and dynamic solutions

Scientific Learning Approaches

Theories

Fundamental knowledge based on assumed laws

For example: based on conservation of energy, momentum, or mass

Design of Experiments

Scientific empirical research by active changes of factors' levels

- Rich library of DOE matrixes
- Requires ability to change factor levels according to DOE matrix
- Backed by statistical analysis (hypothesis testing, ANOVA, regression models)

Machine Learning

Collected data exploration constrained by historical variability and interactivity

DOE is still playing a role in tuning and validation

Learning from facts / data

Mutually supportive

Validation and ongoing monitoring

Scientific Theories And Solutions

Fundamental Navier-Stokes equation describes fluid dynamics and it is based on Conservation Laws:

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} = \mu \cdot \nabla^2 \vec{v} - \frac{1}{\rho} \nabla P + \vec{g}$$

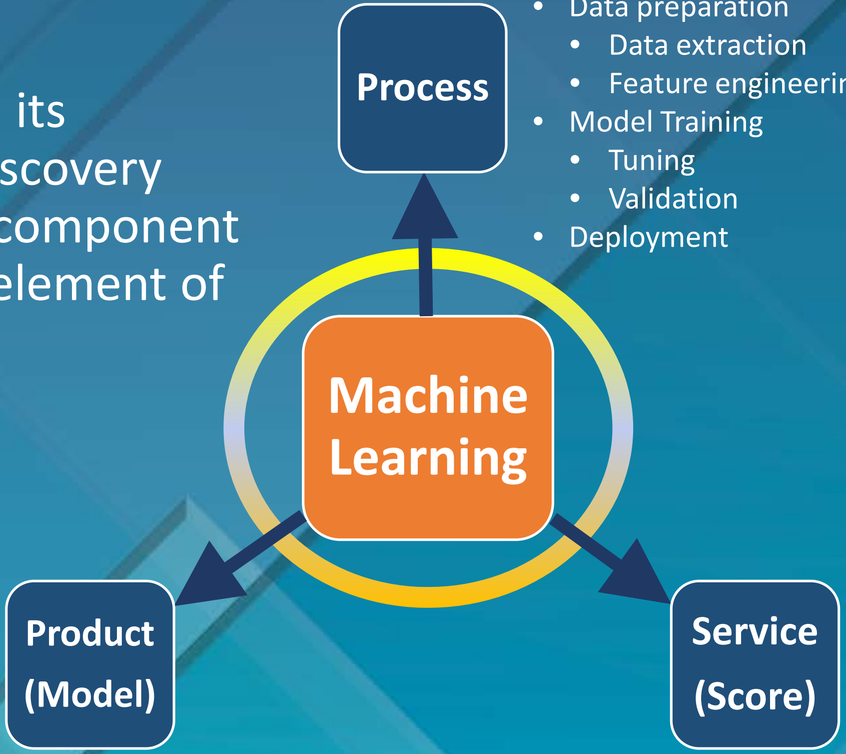
Ref: https://en.wikipedia.org/wiki/Navier-Stokes_equations

Turbulent flow? Millennium Problem by Clay Mathematics Institute

Machine Learning has many theoretically undecidable issues

Machine Learning Dimensions

Machine learning has broadened its capabilities to being an insight discovery mechanism of objects, a plug-in component of high-tech devices, and a core element of classification or rating services



- Data preparation
  - Data extraction
  - Feature engineering
- Model Training
  - Tuning
  - Validation
- Deployment

Machine Learning Is Based On "Learning From Experiences"

Experiences = Observations

- Results can be distributed, cloned, and teleported
- It is still subject to performance monitoring and further improvements

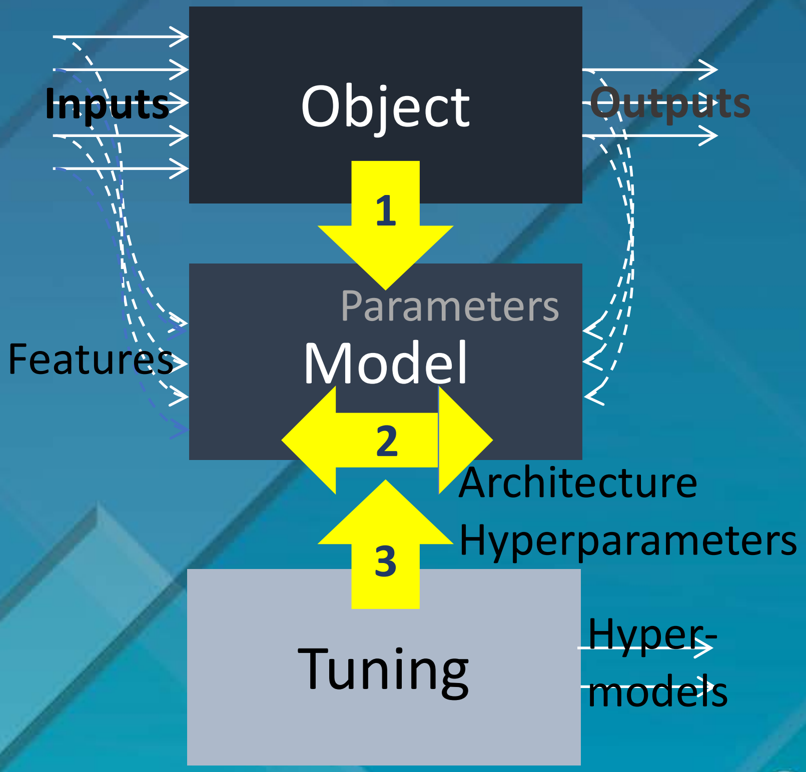
Three Learnings Of Machine Learning

1. Learning about an object

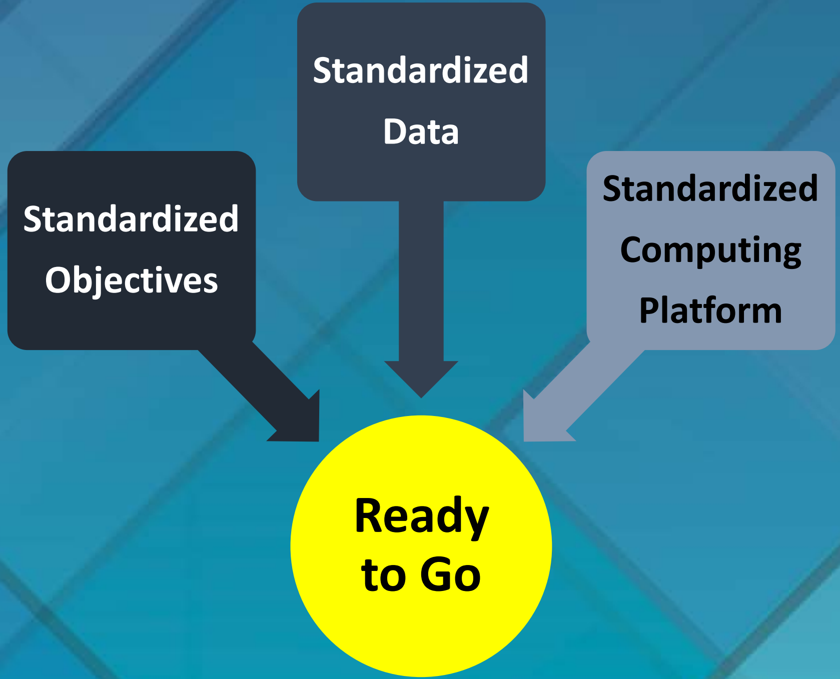
   What inputs are important, how they interact and impact outputs (if any)

2. Learning model parameters during fitting and tuning
3. Learning about the machine learning process

   Study by tuning: Hypermodels



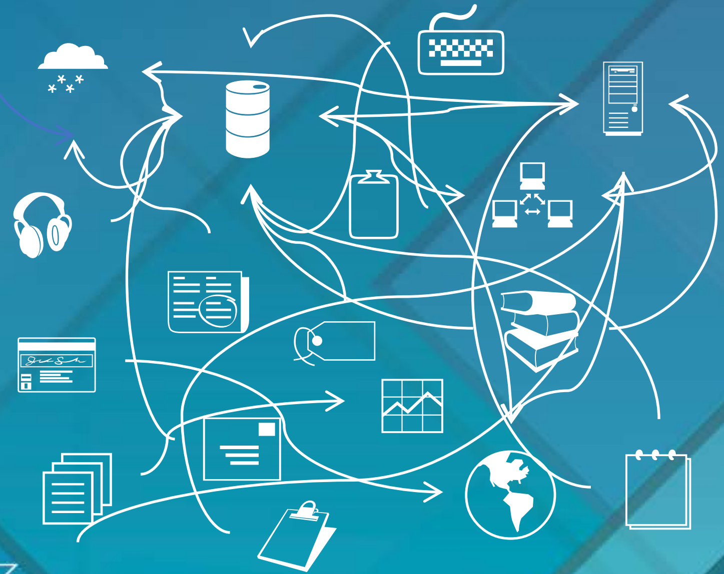
3S Ready To Go Conditions

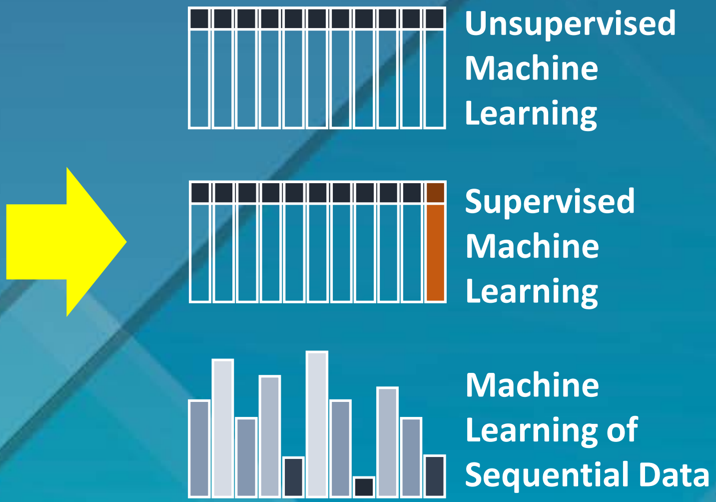


- Some examples:
  - OCR
  - Image Recognition
  - Image-to-Text
  - Voice-to-Text
  - Text-to-Sentiment Rating, . . .
- All applications above have conclusive objectives
- Success in those applications depends on machine learning methodologies/algorithms where data inputs are standardized

Data Sources And Machine Learning

- Business applications, especially focusing on microeconomic results, have multiple data sources of different formats
- Data preparation of input features is a very challenging issue





Machine Learning Ready To Go Solutions

Levels of readiness:

1. Total plug-in solution
   - All three conditions are met
2. Recalibration parameters of the existing model
   - Common variances in features but not in data structure
3. Fitting including tuning within predefined model type and architecture
   - Changes in data structure or specific variances in features
4. Building from scratch

Objectives

Machine Learning Is Explicitly Driven By Its Objectives

Unsupervised ML

Unsupervised Machine Learning may have the same data but different objectives

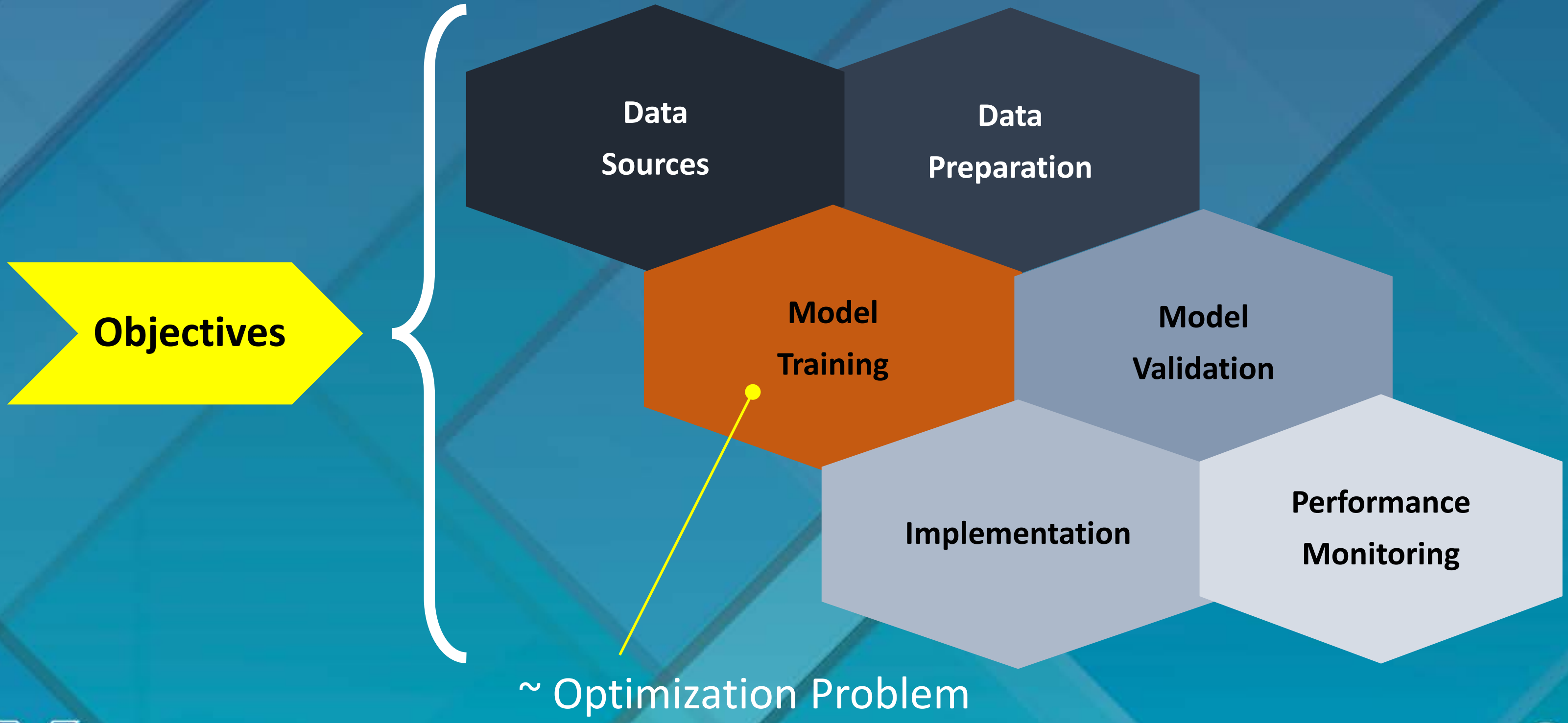
- Features reduction/selection
- Autoencoding
- Outlier detection

Different objectives but the same data structure or even the same data require full trainings with tuning cycles.

Examples:

- Prediction of attrition versus defaults can be done using the same data but different targets
- Identification of the top decile versus discrimination scoring means the same target but different model selection statistics

Objectives Guide Machine Learning Elements At All Stages



Machine Learning Objectives ➔ Robust Model Fitting Criteria

- Is the model performance sound and does it cover its objectives?

  What type of model performance measurement should be selected?

- How stable is the model performance?

  Is the model overfitted? How do you deal with the variance of the performance measure validating the model?

Quality Function Deployment Relationship Matrix

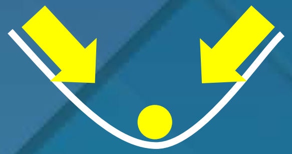
	How?		
What?	Criteria		
Model Objectives	Misclassification rate	K-S at first decile	Gini
➔ Classification of objects	●	△	○
Discrimination (scoring) of population	○	△	●
Lift of the top decile	△	●	○

Relationships
● Strong
○ Medium
△ Weak

Source: Glushkovsky, A. 2018. “Robust Tuning for Machine Learning”, *Proceedings SAS Institute Inc.*, Paper 1868-2018

Stability

Stabilities Of Machine Learning



- Convergence of training

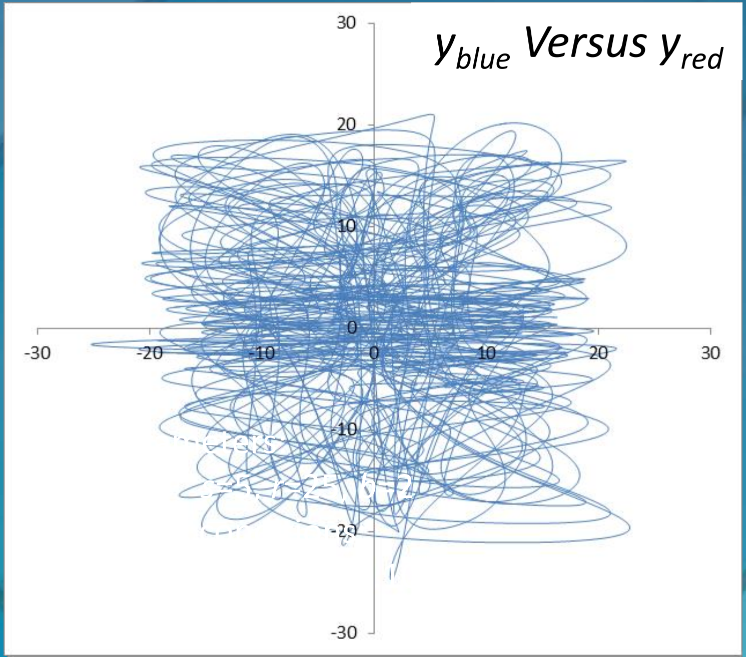
  $$\frac{\partial W_i}{\partial t} = -\alpha \frac{\partial C}{\partial W_i}$$

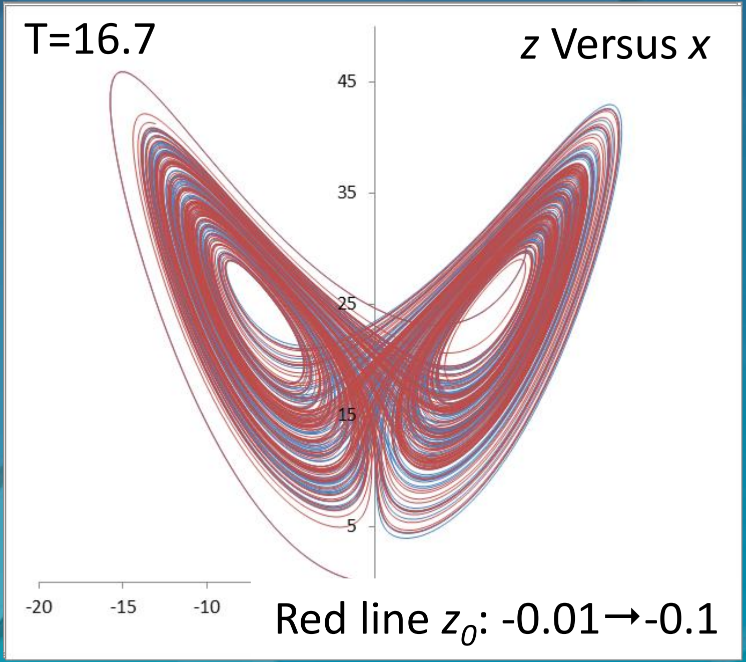
  Where W_i are weights of the model, C is a cost function and α is a learning rate

- Performance of the trained model
  - Out-of-Sample validation (~Static)
  - Out-of-Time validation for long-lasting models (~Dynamic)
    - Macroeconomic cycles
    - Market competiveness
    - Regulatory requirements

Machine Learning Means Convergence Of Dynamic Solutions

Illustration of Stability Issues of Non-linear Dynamic System: Lorenz Attractors

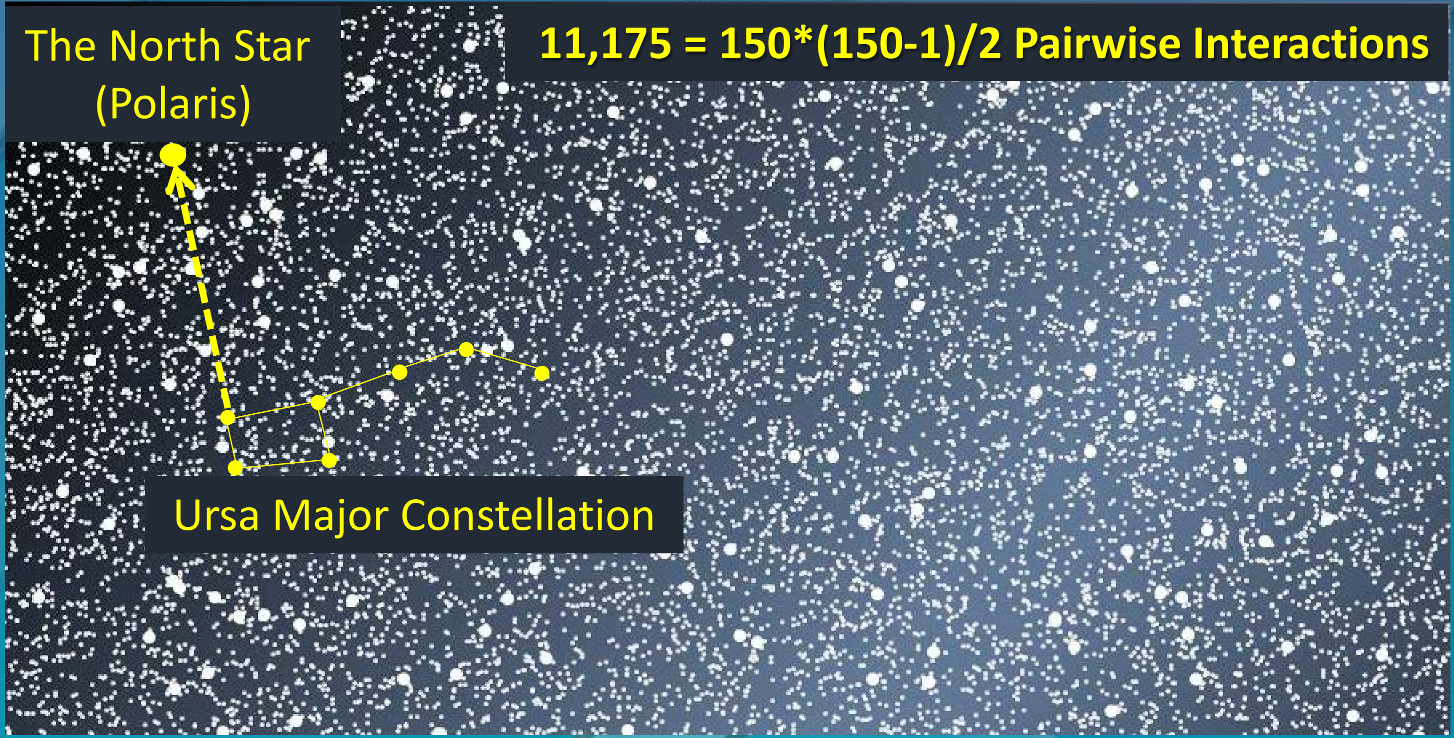




Ref: https://en.wikipedia.org/wiki/Lorenz_system

Feature Engineering

Features: Skyview Illustration

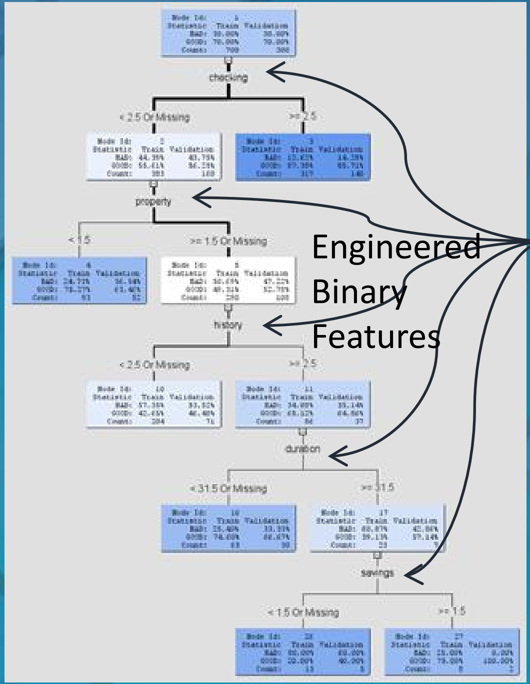


Feature Engineering

- Manual / Supervised engineering
  - Transformation
  - Aggregation
  - Transposing
  - Dummy coding

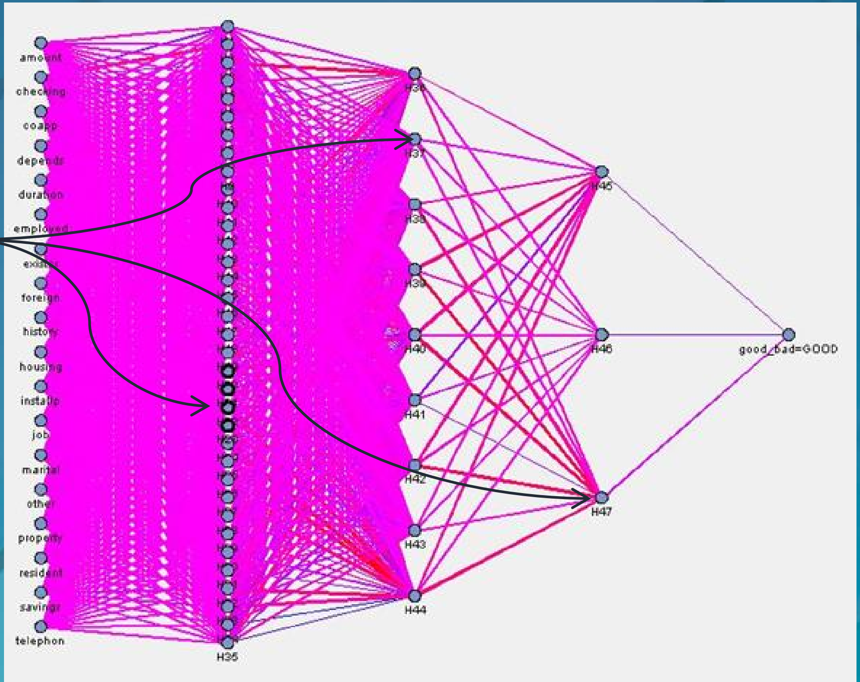
  SAS® Enterprise Guide® tools, coding, macros, Enterprise Miner™ "Modify" nodes are helpful for this
- Binning/Interactive Grouping in SAS® Enterprise Miner™ ("Credit Scoring" node) to address non-linear and monotonic effects
- Autoencoders, Clustering, and Principal Component Analysis to minimize dimensionality ("Multivariate" tasks in SAS® Enterprise Guide® or Enterprise Miner™ "Explore" and "Modify" nodes)
- Creates virtually endless opportunities

Engineering Of Interactive "Derivative" Features



Engineered Binary Features

"Derivative" Features



Input Features 1st Layer 2nd Layer 3rd Layer

Dataset "German Credit Data Set" is from the sample datasets for SAS® Enterprise Miner™

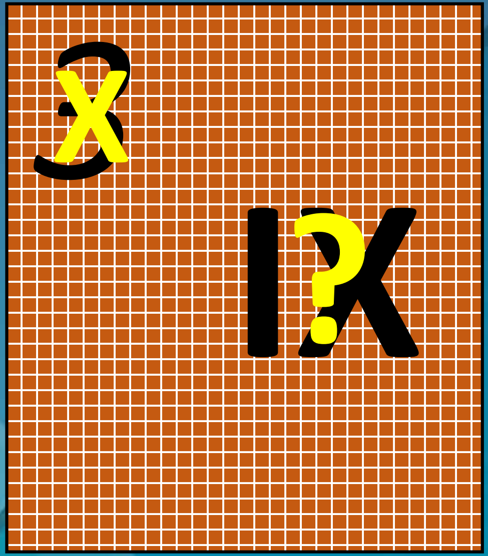
Some Issues Of Input Features

- The most important feature(s) may be not included or has no variance in the collected data
- Some applications have conditions that are driven by business sense such as monotonity (SAS® Enterprise Miner™ "Interactive Grouping" node) and defined direction of the impact of features
- Features such as ID may be mistakenly included. It usually produces overfitted results with no practical sense
- Pre-screening of variables may reject the important features related to outliers
- Masking effect when there are too many features. Dropping one of the most important features, it may have almost no effect on the model performance results but huge impact on the model description

Screening Of Input Features

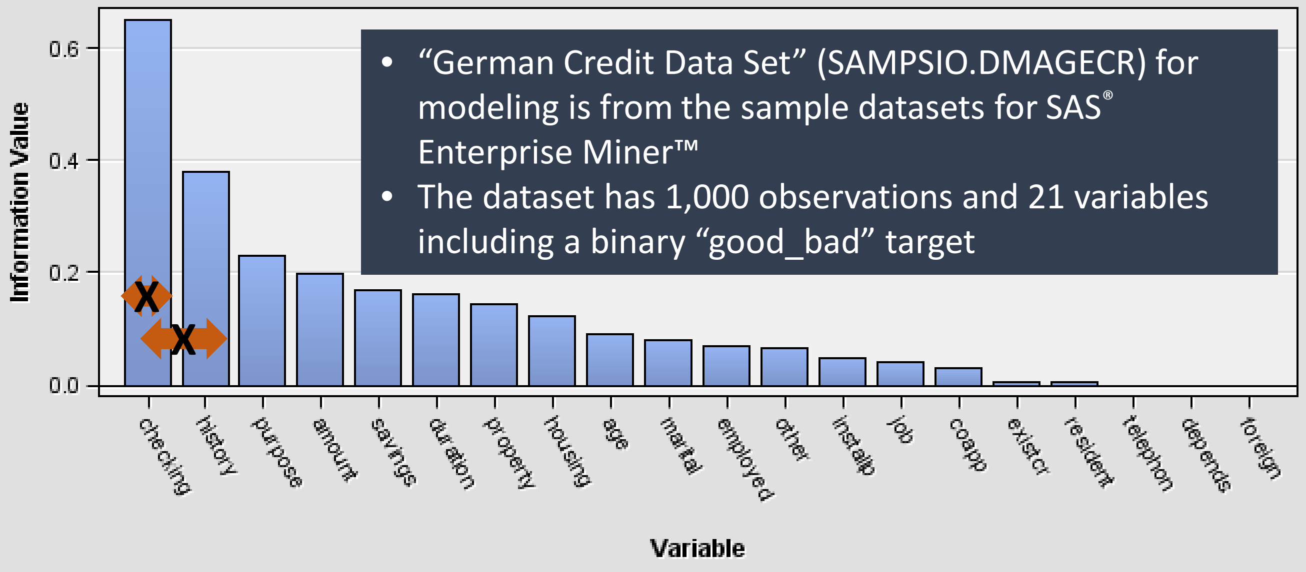
Features of red area will be commonly rejected

Outlier and Biased Examples



Illustrative Example

SAS® Enterprise Miner™ example with different HP Bayesian Network Classifier models rejecting some inputs



- "German Credit Data Set" (SAMPSIO.DMAGECR) for modeling is from the sample datasets for SAS® Enterprise Miner™
- The dataset has 1,000 observations and 21 variables including a binary "good_bad" target

Illustrative Example (cont’d)

SAS® Enterprise Miner™ HP Bayesian Network Classifiers

Model Description	HPBNC	HPBNC2	HPBNC3
Misclassification Rate	0.267	0.317	0.331

Ref: SAS® Enterprise Miner™ Reference Help

Randomization

Randomization In Machine Learning

Random shaking accelerates processes and it is a root of robustness

- Initialization of iterative process, such as initial weight values in Neural Networks
- Dropout Layers
- Stochastic Gradient Descent

Random short movements better fit pockets

- Random Forest
- Stochastic Gradient Boosting

Randomization Plays A Crucial Role In Machine Learning

- It brings confidence in model validation by splitting between training and validation datasets
- It allows fast Stochastic Gradient Descent by reasonable subsampling
- It works as a stochastic regularization mechanism by applying dropout layers in Neural Networks that prevent overfitting by reducing number of nodes
- It is a core element in Random Forest selecting observations and features of sub-models
- It initializes values of weights of Neural Networks in iterative optimization processes
- It is applied in tuning
  - Stochastic Design of Experiments
  - Accelerated Tuning

Regularization

Regularization In Machine Learning

- Supports model robustness and stability
- Penalization by model complexity
- Built-in mechanism to prevent overfitting
- Random Forest, Stochastic Gradient Boosting, and Dropout layers of Neural Networks are examples where regularization meets randomization
- Statistical approaches: degree of freedom, regression model selection (such as the backward method)

Tuning

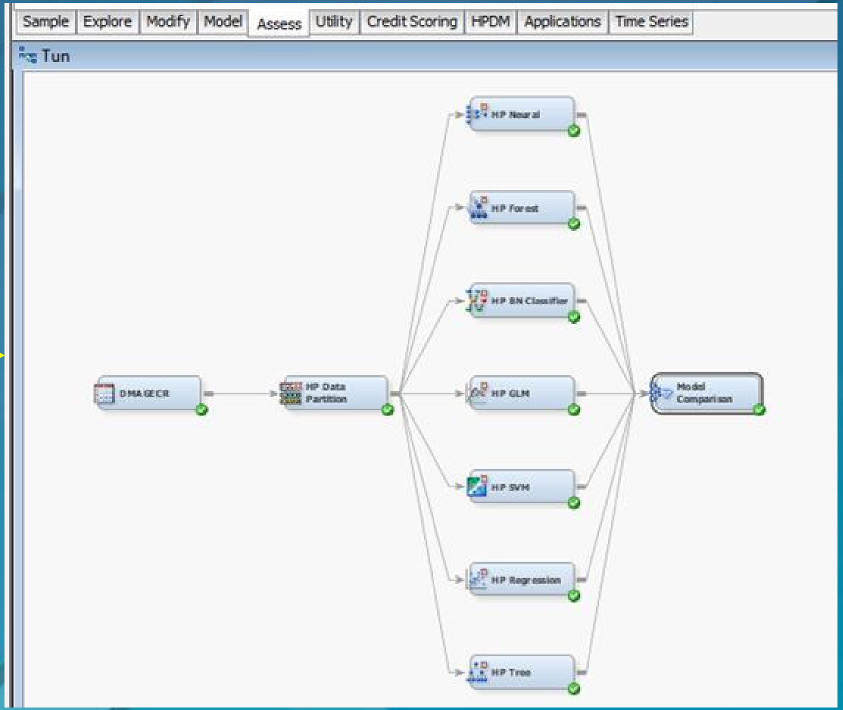
Tuning Is Part Of Our Everyday Activity

It is a learning mechanism as well

Source: Glushkovsky, A. 2018. "Robust Tuning for Machine Learning", *Proceedings SAS Institute Inc.*, Paper 1868-2018

Illustrative Example

Dataset "German Credit Data Set" for modeling is from the sample datasets for SAS® Enterprise Miner™ The dataset has 1,000 observations and 21 variables including a binary "good_bad" target



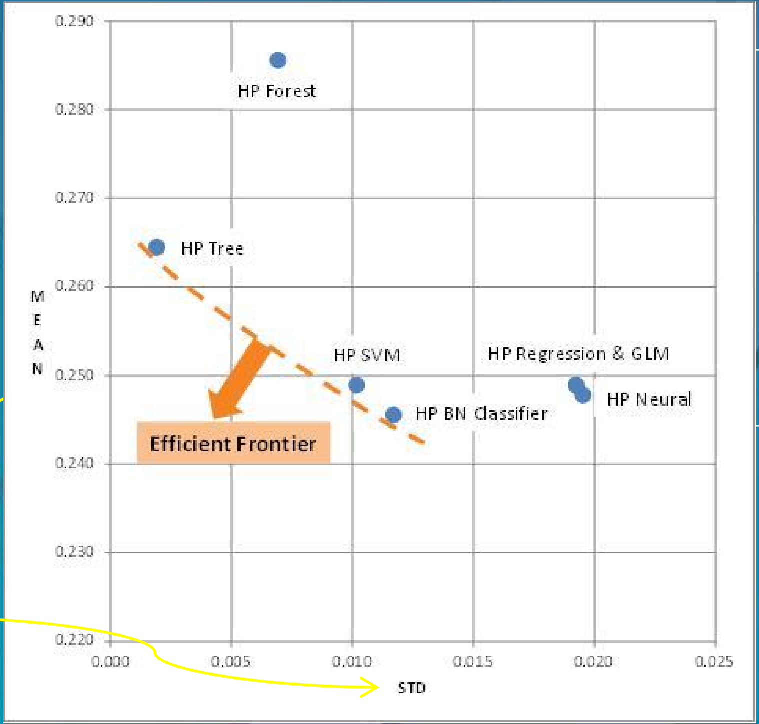
Source: Glushkovsky, A. 2018. "Robust Tuning for Machine Learning", *Proceedings SAS Institute Inc.*, Paper 1868-2018

Validation Misclassification Rates For Seven SAS® Enterprise Miner™ HP Models At Default Setups

	Validation: Misclassification Rate				
	Partition Random Seed			Dual Response	
Model Description	12345	11223	54321	MEAN	STD
HP BN Classifier	0.233	0.257	0.247	0.246	0.012
HP Forest	0.280	0.293	0.283	0.286	0.007
HP GLM	0.227	0.260	0.260	0.249	0.019
HP Neural	0.233	0.270	0.240	0.248	0.020
HP Regression	0.227	0.260	0.260	0.249	0.019
HP SVM	0.240	0.260	0.247	0.249	0.010
HP Tree	0.263	0.267	0.263	0.264	0.002

Source: Glushkovsky, A. 2018. “Robust Tuning for Machine Learning”, *Proceedings SAS Institute Inc.*, Paper 1868-2018

Dual Response Scatterplot



Seven SAS® Enterprise Miner™ HP Models at Default Setups

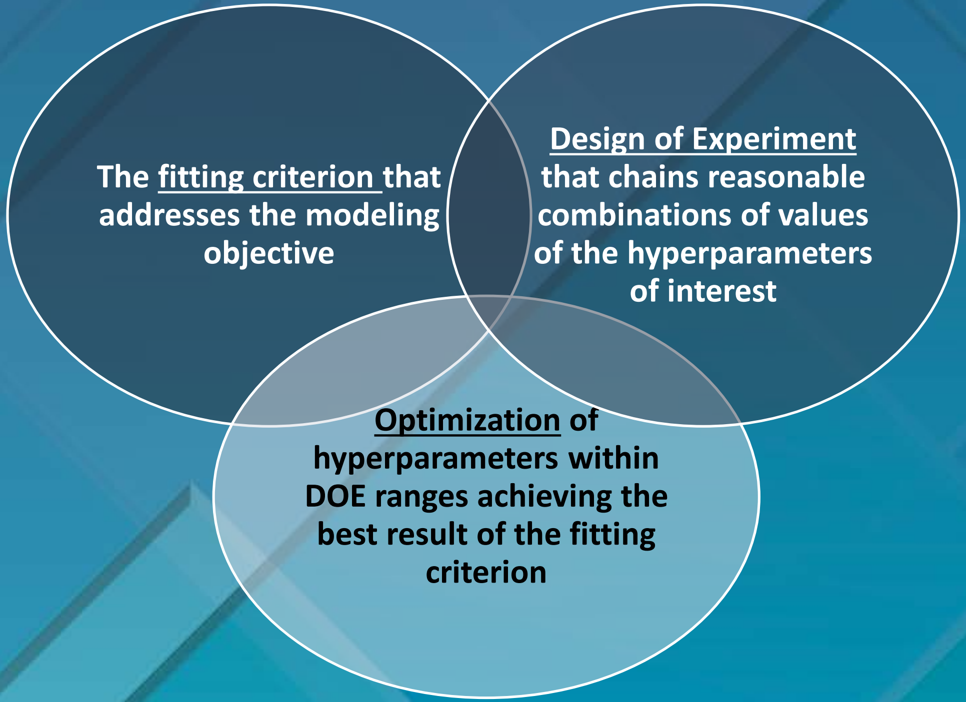
Mean of Validation Misclassification Rate

Standard Deviation of Validation Misclassification Rate

Source: Glushkovsky, A. 2018. “Robust Tuning for Machine Learning”, *Proceedings SAS Institute Inc.*, Paper 1868-2018

"Ingredients" Of Machine Learning Tuning

To succeed in the tuning process, three right "ingredients" should be mixed well



Source: Glushkovsky, A. 2018. "Robust Tuning for Machine Learning", *Proceedings SAS Institute Inc.*, Paper 1868-2018

Machine Learning Tuning

- Tuning adjusts hyperparameters both numeric (such as learning rate) and architectural (such as number of Neural Net layers and their types)
- It may apply Design of Experiments of different arrays: full factorial, Latin hypercubes, random search
- Autotuning
  - SAS® Visual Data Mining and Machine Learning includes autotuning features

    Ref: Koch, P., Wujek, B., Golovidov, O., and Gardner, S., 2017. "Automated Hyperparameter Tuning for Effective Machine Learning", *Proceedings SAS Institute Inc*, SAS Paper SAS514-2017
- Dual-response
  - Hypermodels discover impacts of hyperparameters on model performance
  - Practical implementation of dual response tuning can be done using SAS® Enterprise Miner™ and the "Export Path as SAS Program" feature

    Ref: Glushkovsky, A. 2018. "Robust Tuning for Machine Learning", *Proceedings SAS Institute Inc.*, Paper 1868-2018

Learning From Tuning

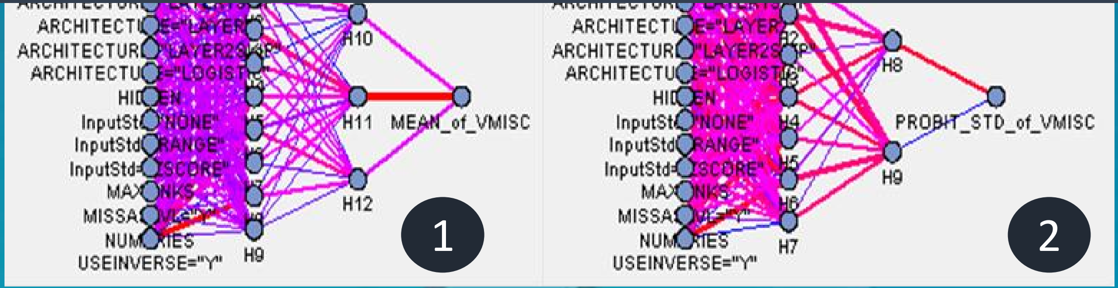
Dual Response Methodology

$$\hat{Y}_{MEAN} = fm(\overrightarrow{H_I})$$ 1

$$\hat{Y}_{STD} = fs(\overrightarrow{H_I})$$ 2

$\overrightarrow{H_I}$ is a vector of inner hyperparameters

Models (1) and (2) can be seen as "hypermodels"



Source: Glushkovsky, A. 2018. "Robust Tuning for Machine Learning", *Proceedings SAS Institute Inc.*, Paper 1868-2018

Factors And Levels Of The Simulated Design Of Experiment

	Hyperparameters						
	Inner Factors						Outer Factor
	ARCHITECTURE	HIDDEN	InputStd	MAXLINKS	NUMTRIES	USEINVERSE	RandomSeed
Levels	5	29	3	15	5	2	3
Values	LAYER1 LAYER2 LOGISTIC LAYER1SKIP LAYER2SKIP	2-30 step = 1	NONE RANGE ZSCORE	400-1100 step = 50	2-6 step = 1	Y N	11223 12345 54321

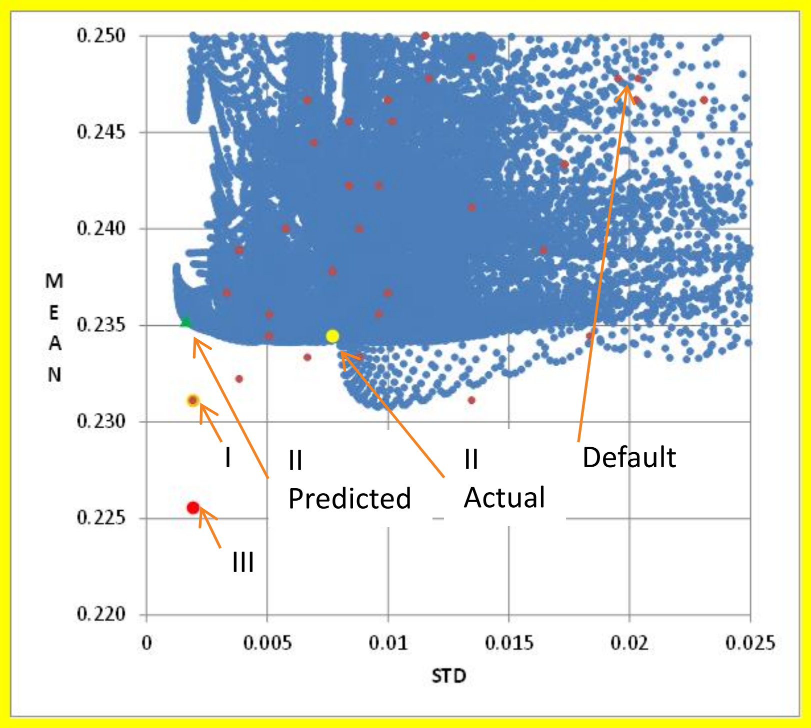
The total number of the simulated trials is 64,360

Tuning may have more runs than the original dataset!

Source: Glushkovsky, A. 2018. “Robust Tuning for Machine Learning”, *Proceedings SAS Institute Inc.*, Paper 1868-2018

Tuning Simulation

Dual Response MEAN versus STD Scatterplot

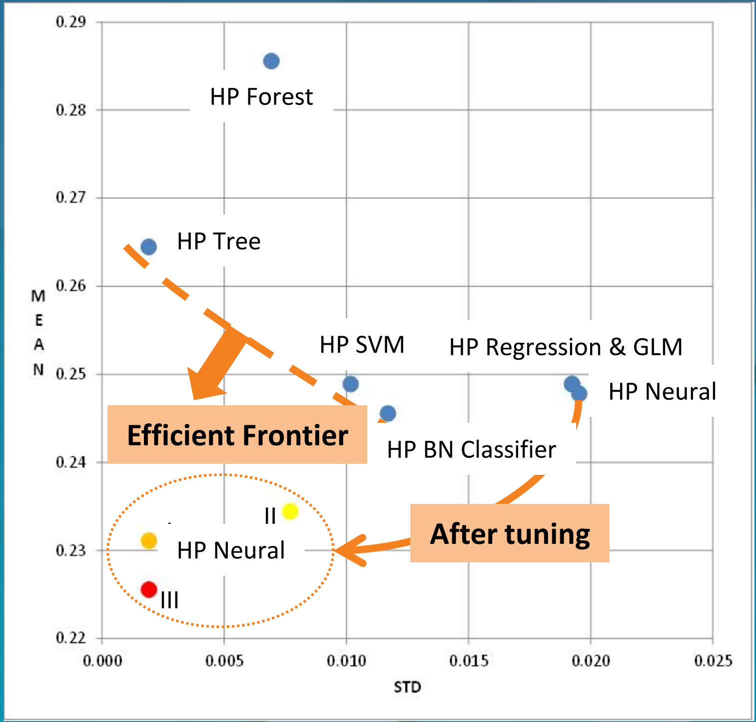


- Blue Dots – Simulated Results Based on Fitted Dual Response Models
- Red Dots – Actual DOE Results

Source: Glushkovsky, A. 2018. "Robust Tuning for Machine Learning", *Proceedings SAS Institute Inc.*, Paper 1868-2018

Illustration of Tuning Results

Dual Response Scatterplot



Source: Glushkovsky, A. 2018. “Robust Tuning for Machine Learning”, *Proceedings SAS Institute Inc.*, Paper 1868-2018

Outliers

Classification Of Outliers

Outliers = Assignable Variance
Statistical Process Control Concept

Rare, Unusual, Uncommon, Unexpected, Violated, Error

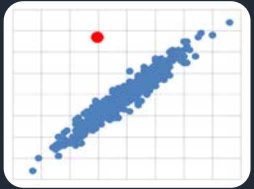
Inherent
•
Error X

Impact on Business

Correct & Prevent

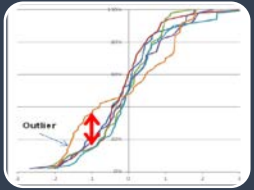
Source: Glushkovsky, A. 2017. “Outline Outliers: Adding a Business Sense”, *Proceedings SAS Institute Inc.*, Paper 0372-2017

Types Of Outliers



Univariate (Distance, p-Value, Time Series)

Multivariate (Regressions, Clusters, Autoencoders)

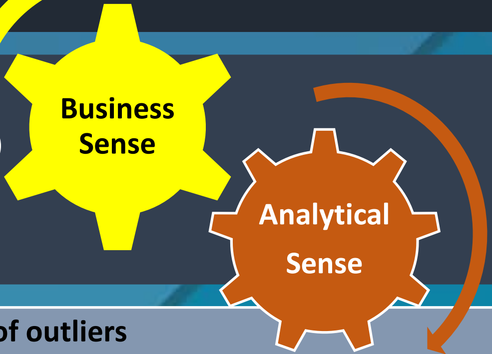


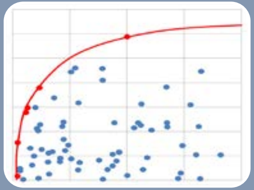
Distributions, Patterns

Combinations of items (baskets)

Sequences of events

Models





Efficient frontier, which is a set of outliers

Optimization points of the objective functions (pricing that maximizes profit)

Extraordinary features of products, services, or processes

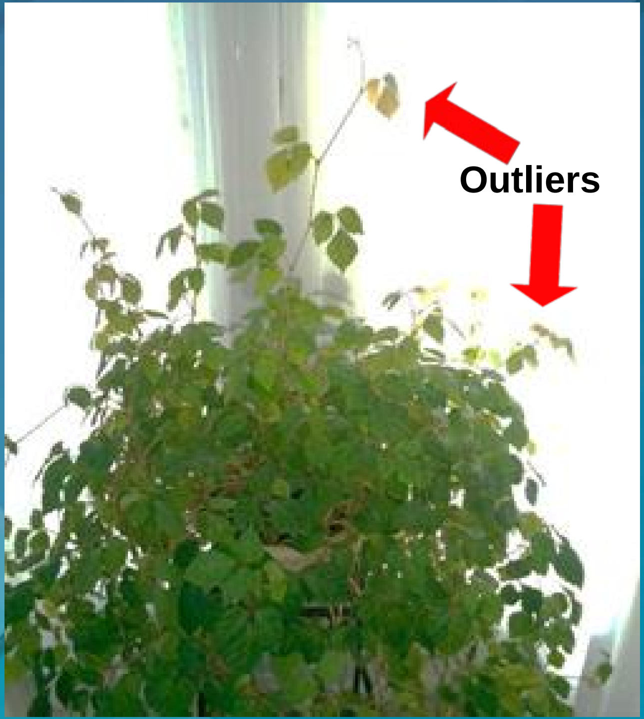
Non-dominated strategies that form the Nash equilibrium

Source: Glushkovsky, A. 2017. “Outline Outliers: Adding a Business Sense”, *Proceedings SAS Institute Inc.*, Paper 0372-2017

Outliers Support Evolution

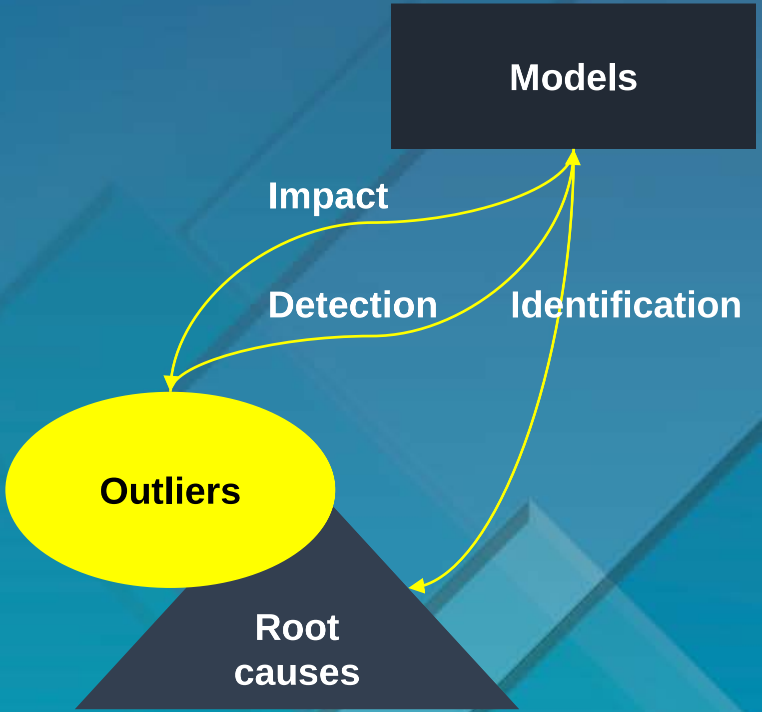
Utilization of "Outliers" Is a Natural Growing Mechanism of the Climbing Plant

- The continuous improvement strategy by spreading outliers in different directions, checking the outcomes, and selecting the most beneficial one for future growth



Source: Glushkovsky, A. 2017. "Outline Outliers: Adding a Business Sense", *Proceedings SAS Institute Inc.*, Paper 0372-2017

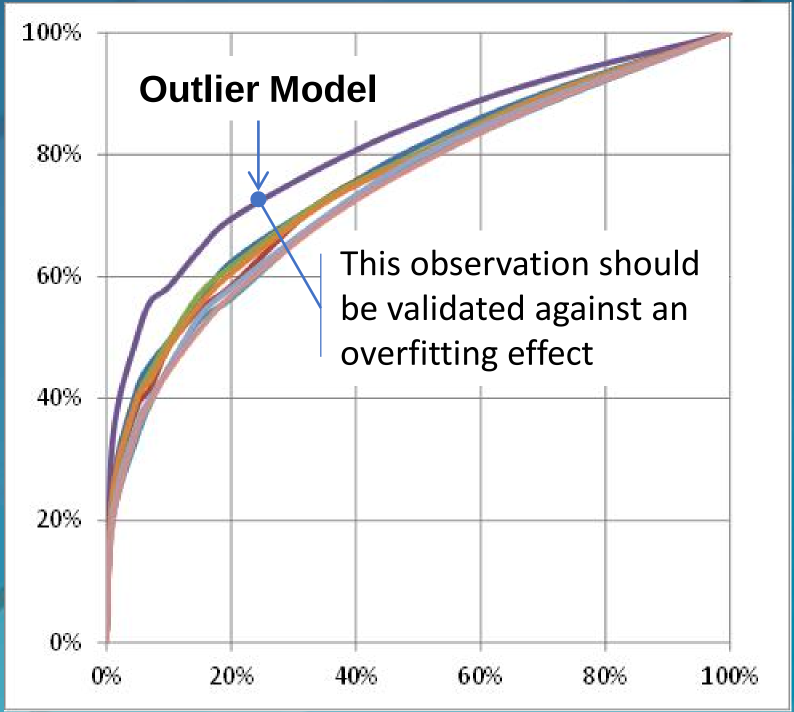
Relationships Between Outliers And Machine Learning



Source: Glushkovsky, A. 2017. “Outline Outliers: Adding a Business Sense”, *Proceedings SAS Institute Inc.*, Paper 0372-2017

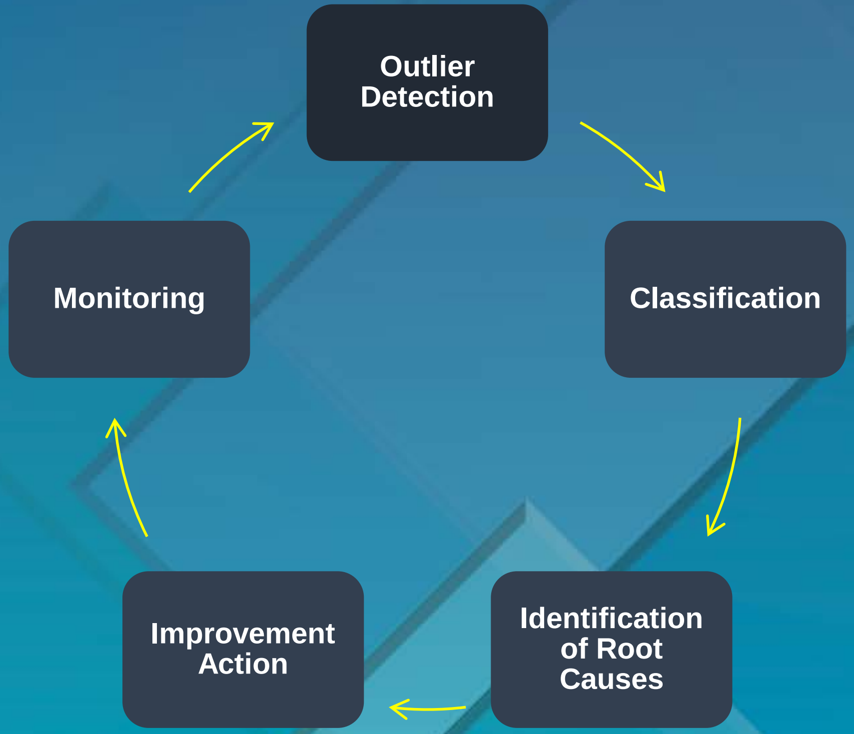
Outliers And Models

Lift Charts Comparing Different Models of the Same Problem



Source: Glushkovsky, A. 2017. “Outline Outliers: Adding a Business Sense”, *Proceedings SAS Institute Inc.*, Paper 0372-2017

Outlier Based Business Improvement Cycle



Source: Glushkovsky, A. 2017. "Outline Outliers: Adding a Business Sense", *Proceedings SAS Institute Inc.*, Paper 0372-2017

Conclusion

- Most real word business problems still require customized Machine Learning solutions
  - Various data sources, architectures, products, communications, customer populations, macroeconomic cycles, market conditions, …
  - Diverse Objectives, Features, Stability (Randomization, Regularization), Tuning, Outliers, and, of course, Machine Learning methodologies and approaches
- The starting point of Machine Learning is defining the target variable (for supervised training) and the appropriate criterion of the model performance measure that suit the **Objectives**. Do not flop on this crucial point!

Conclusion (cont'd)

- Focus on **Feature Engineering**
- Validate **Stability** of the results keeping in mind **Randomization** and **Regularization**
- Always apply **Tuning** and learn from it
- Do not ignore **Outliers**, but hunt for them! It brings discoveries and opportunities
- Even upon achieving the best results, Machine Learning models are still subject to ongoing monitoring and have virtually unlimited opportunities for improvement

Machine Learning: Ready to Go Solutions or Complex Scientific Projects?

Alex Glushkovsky, BMO Financial Group

Disclaimer

- Figures are for illustration purposes only and assume an arbitrary list of objects, relationships, and locations.
- The presentation represents the views of the author and do not necessarily reflect the views of the BMO Financial Group.

ANALYTICS EXPERIENCE