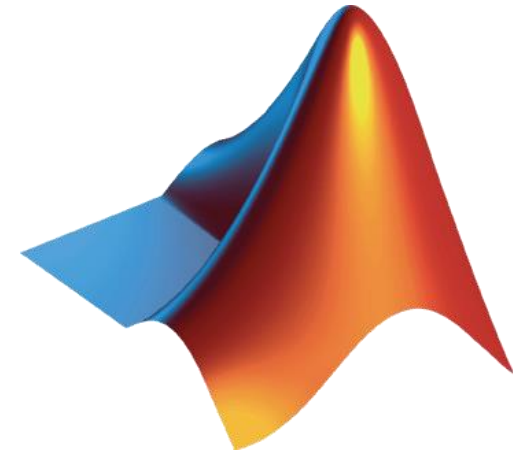


Machine Learning with MATLAB

Debbi Cohen
RPI Account Manager

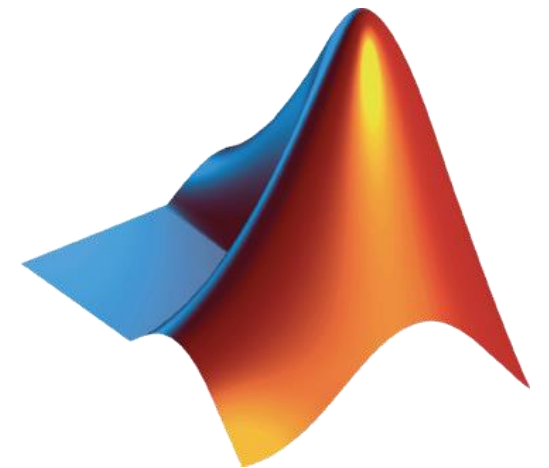
Adam Sifounakis
Application Engineer



March 30, 2016

Goals

- Overview of machine learning
- Machine learning models & techniques available in MATLAB
- Using MATLAB to streamline your machine learning workflow

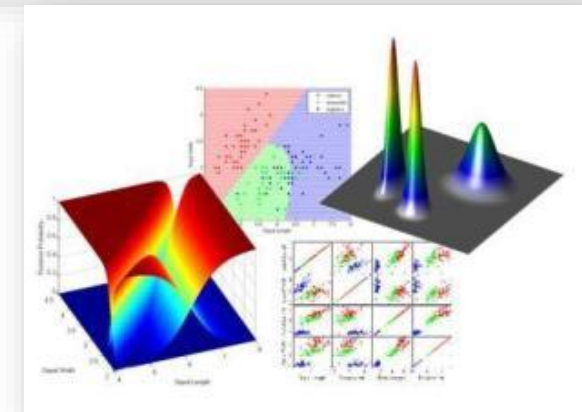
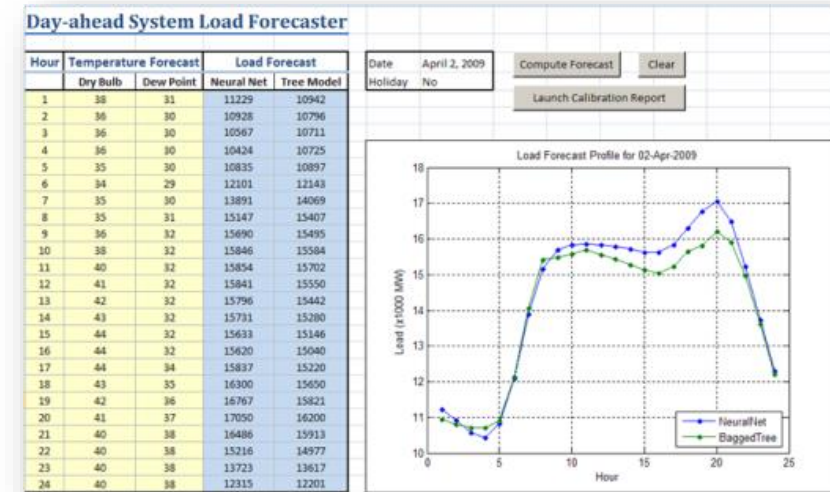


Machine Learning

Characteristics and Examples

- Characteristics
 - Lots of data (many variables)
 - System too complex to know the governing equation
(e.g., black-box modeling)

- Examples
 - Pattern recognition (*speech, images*)
 - Financial algorithms (*credit scoring, algo trading*)
 - Energy forecasting (*load, price*)
 - Biology (*tumor detection, drug discovery*)

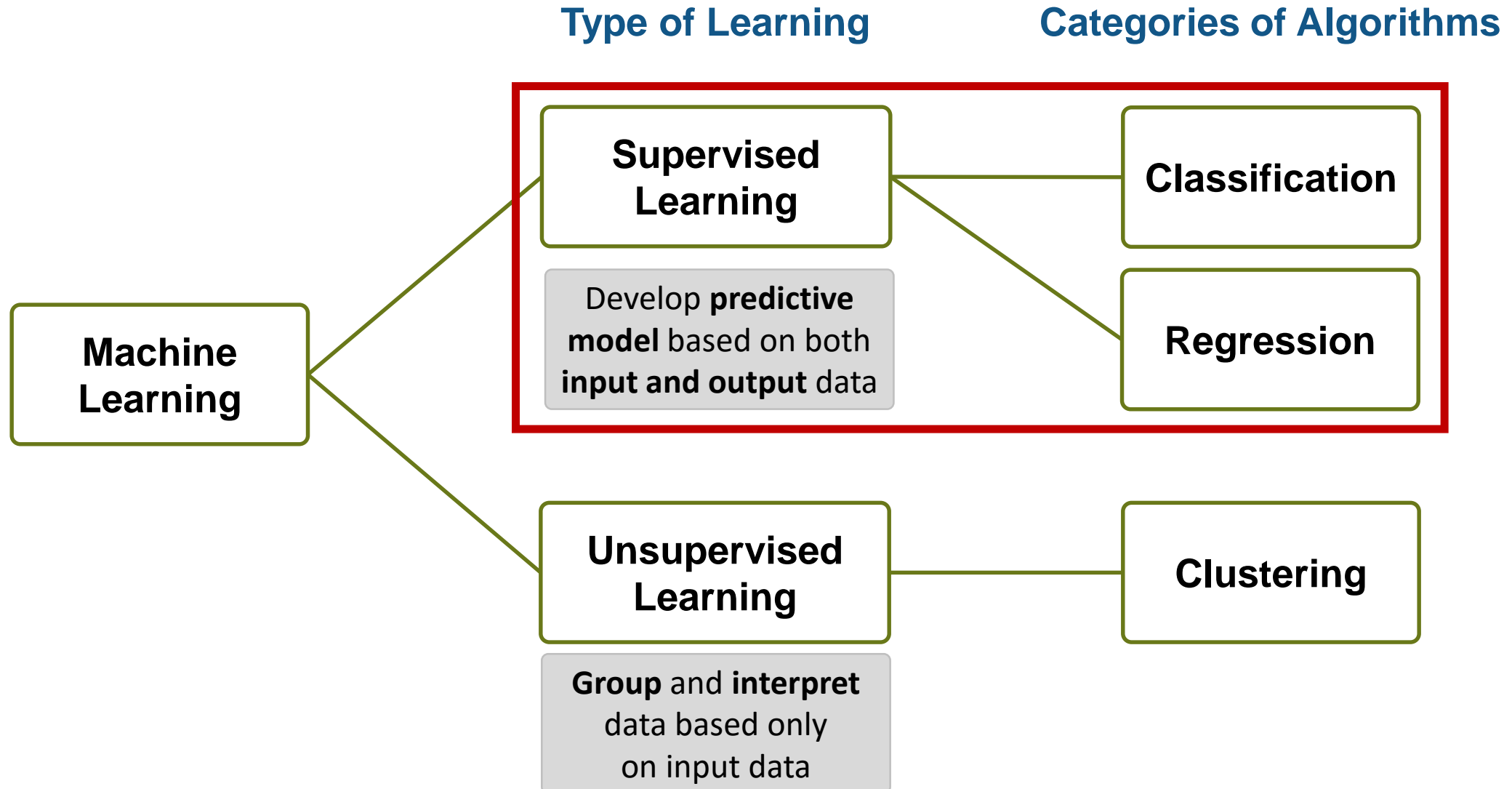


| | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|---------|
| AAA | 93.68% | 5.55% | 0.59% | 0.18% | 0.00% | 0.00% | 0.00% | 0.00% |
| AA | 2.44% | 92.60% | 4.03% | 0.73% | 0.15% | 0.00% | 0.00% | 0.06% |
| A | 0.14% | 4.18% | 91.02% | 3.90% | 0.60% | 0.08% | 0.00% | 0.08% |
| BBB | 0.03% | 0.23% | 7.49% | 87.86% | 3.78% | 0.39% | 0.06% | 0.16% |
| BB | 0.03% | 0.12% | 0.73% | 8.27% | 86.74% | 3.28% | 0.18% | 0.64% |
| B | 0.00% | 0.00% | 0.11% | 0.82% | 9.64% | 85.37% | 2.41% | 1.64% |
| CCC | 0.00% | 0.00% | 0.00% | 0.37% | 1.84% | 6.24% | 81.88% | 9.67% |
| D | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100.00% |
| | AAA | AA | A | BBB | BB | B | CCC | D |

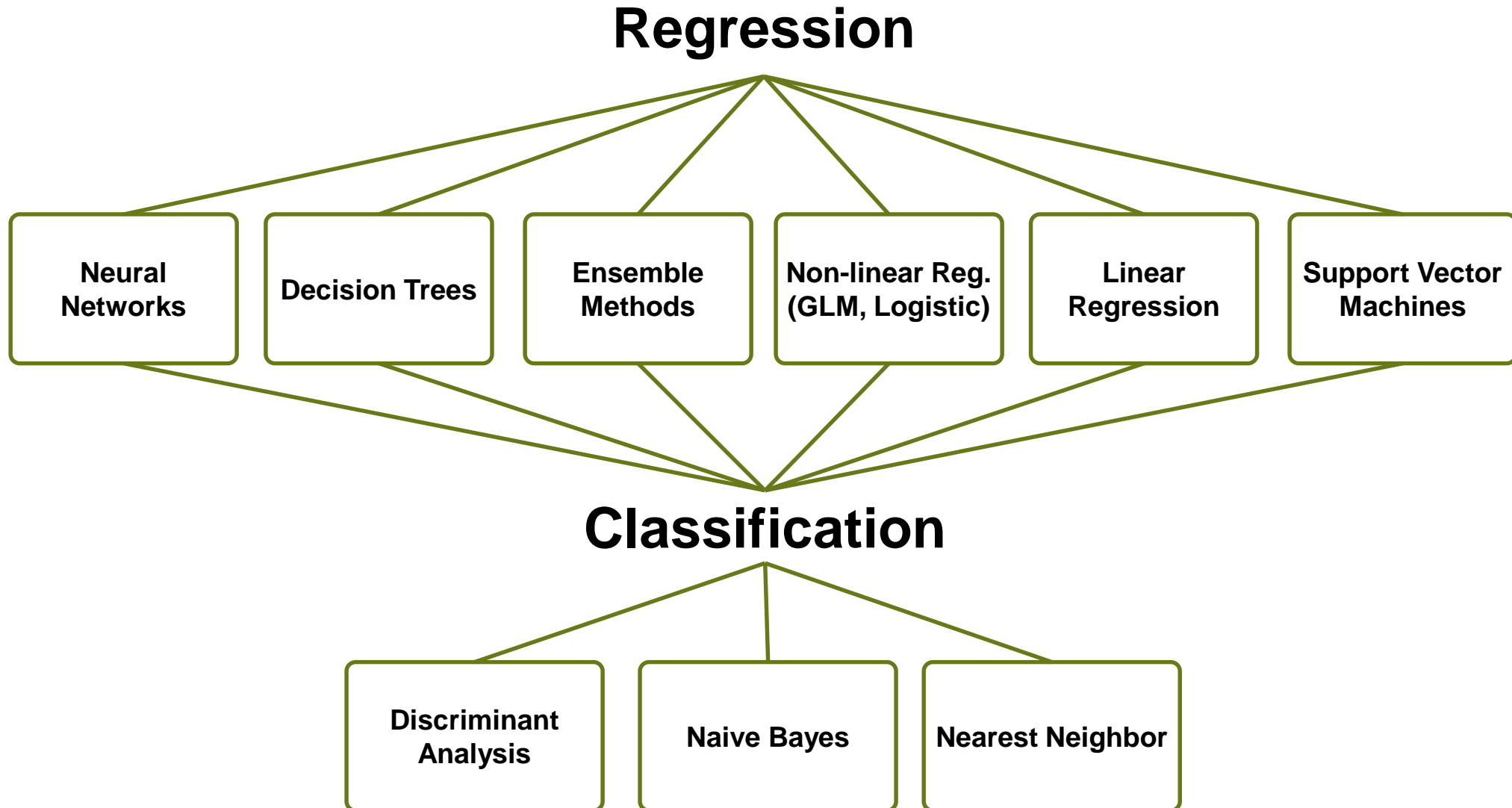
Challenges – Machine Learning

- Significant technical expertise required
- No “one size fits all” solution
- Time required to conduct the analysis
- Locked into Black Box solutions

Overview – Machine Learning

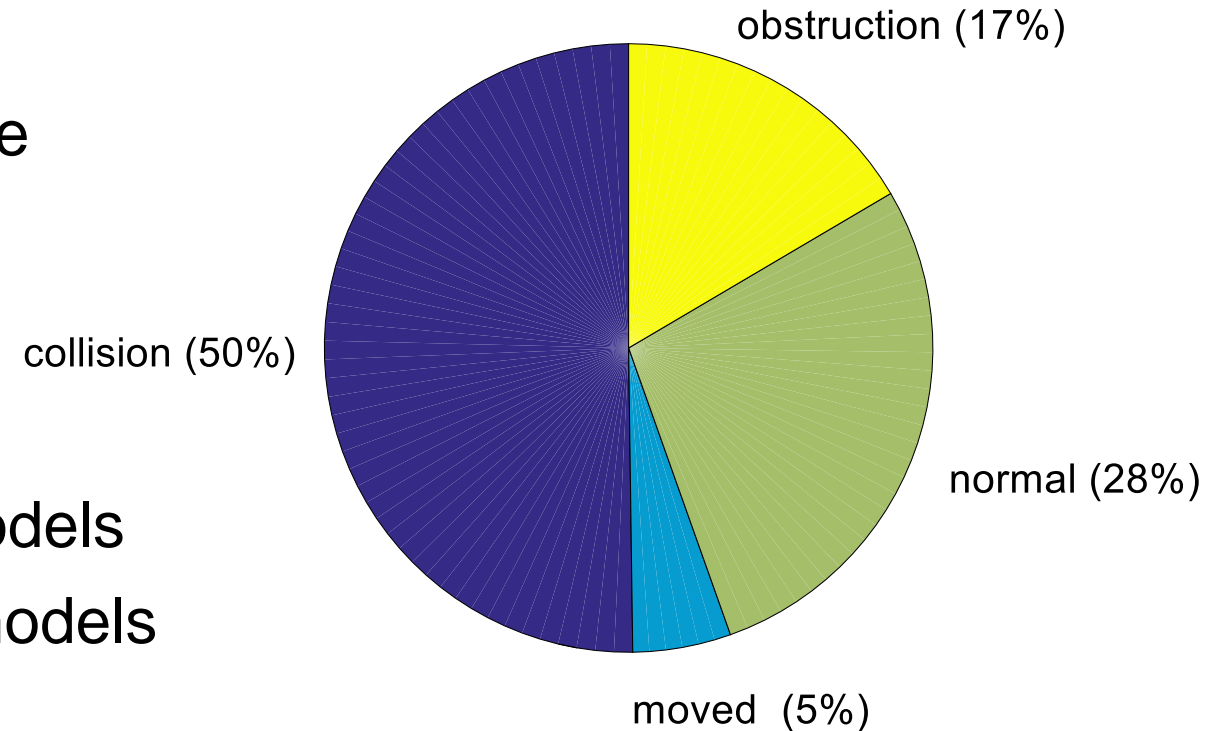


Supervised Learning



Demo: Robot Failures Identification

- Challenge:
 - Find natural groupings among large number of predictors
- Approach:
 - Train a classifier using different models
 - Measure accuracy and compare models
 - Reduce model complexity
 - Use classifier for prediction



Decision Trees – What are they?

- Branching tree structure:

OR

- Nested if statements:

Decision tree for classification

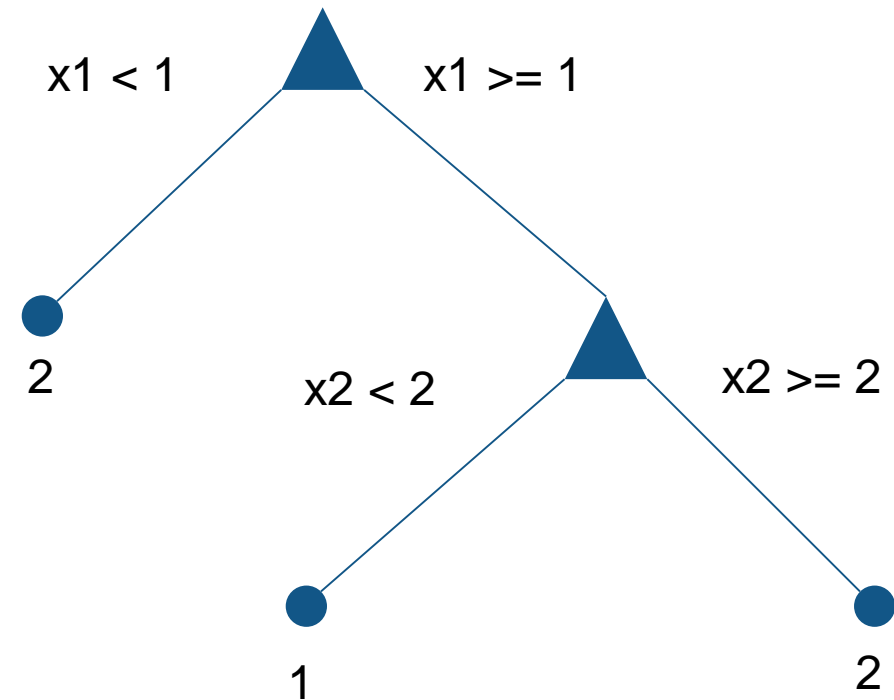
1 if $x_1 < 1$ then node 2 else node 3

2 class = 2

3 if $x_2 < 2$ then node 4 else node 5

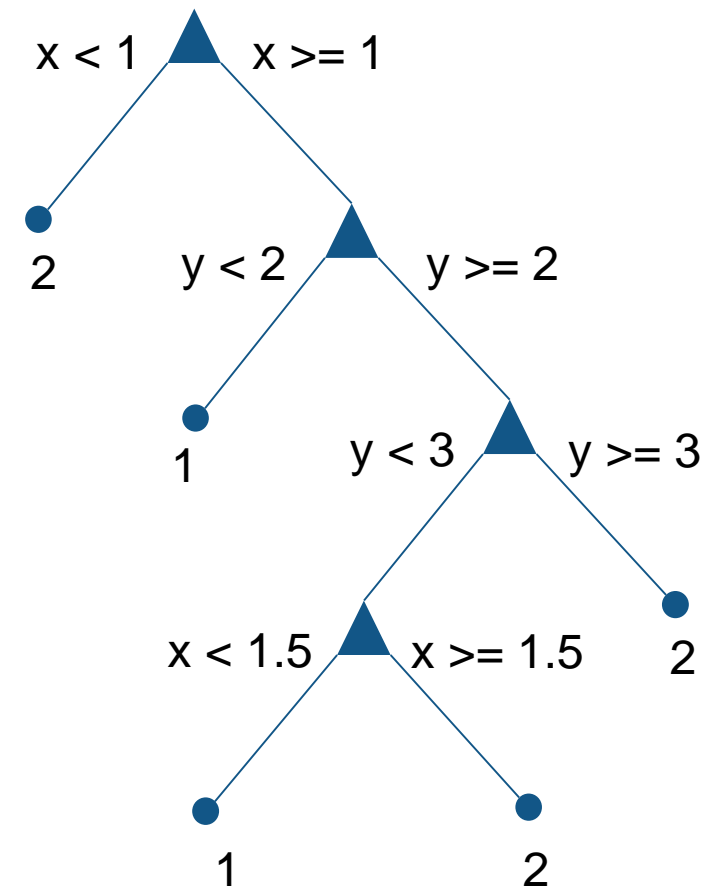
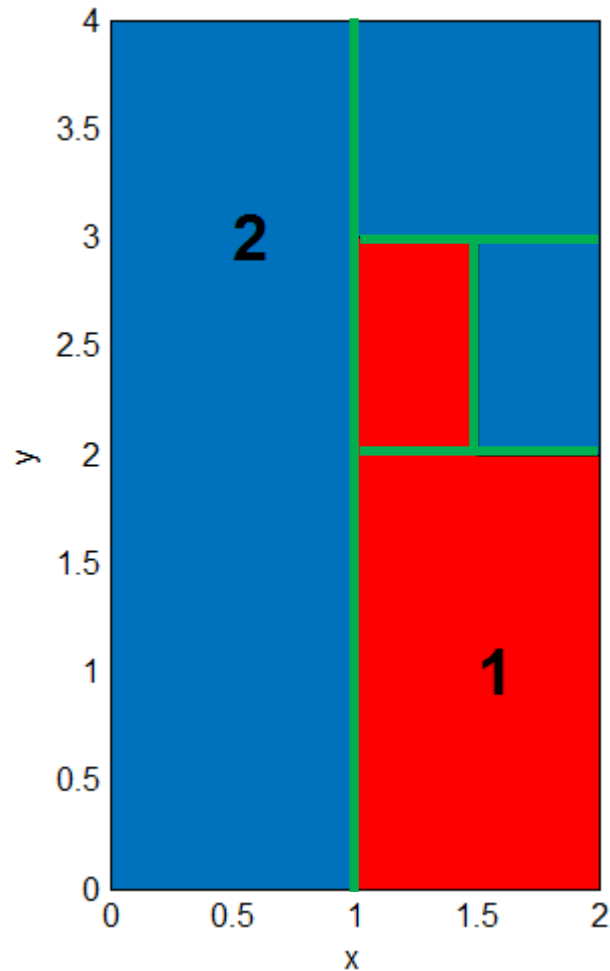
4 class = 1

5 class = 2



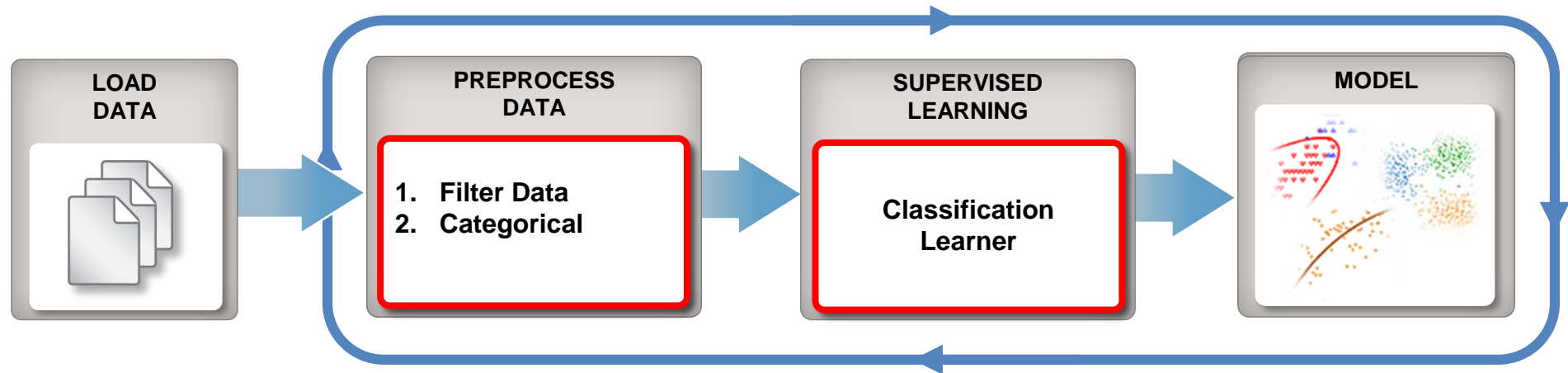
Decision Trees – How do they work?

Design a classifier for the following situation:

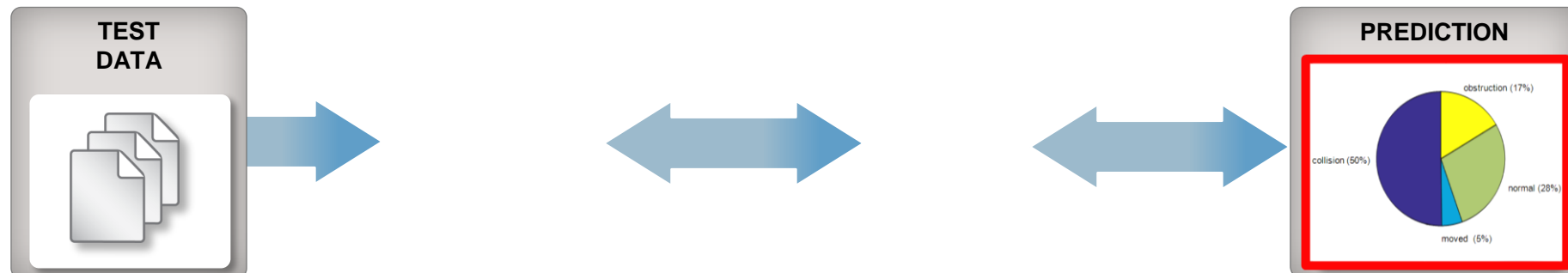


Machine Learning Workflow

Train: Iterate until you find the best model

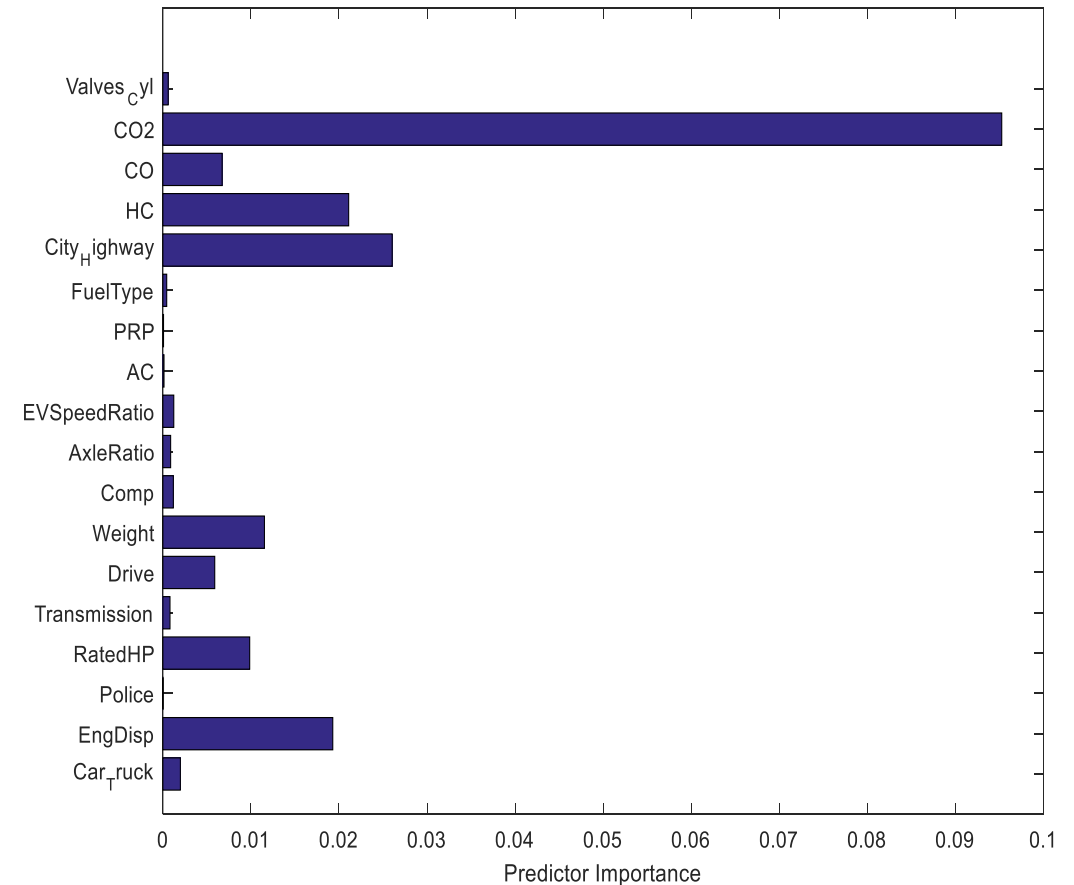


Predict: Integrate trained models into applications



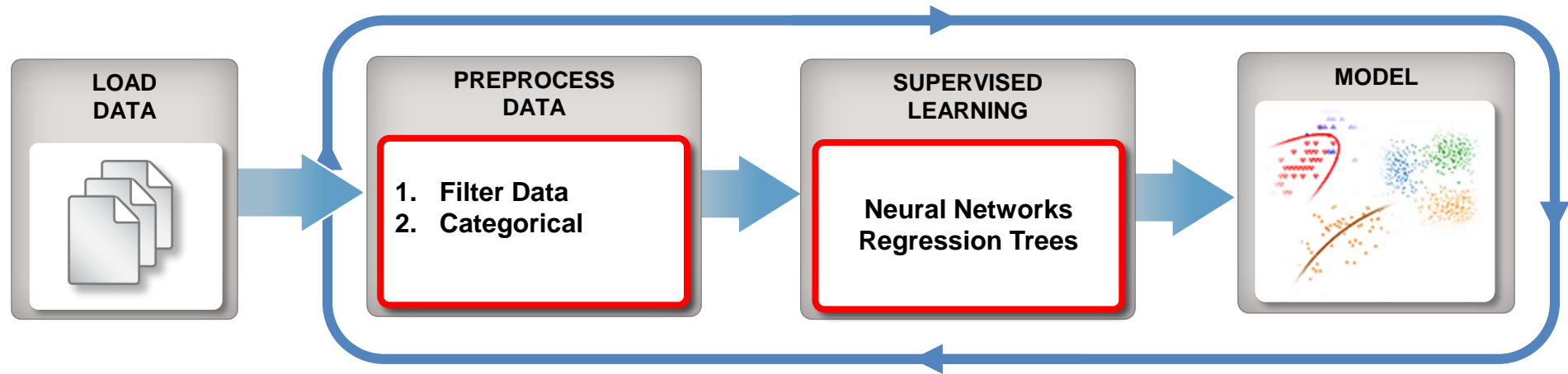
Demo: Predicting Fuel Economy

- Goal:
 - Study the relationships between fuel economy, horsepower, and type of vehicle
- Approach:
 - Access data from Excel
 - Create a predictive model
 - Improve the model accuracy

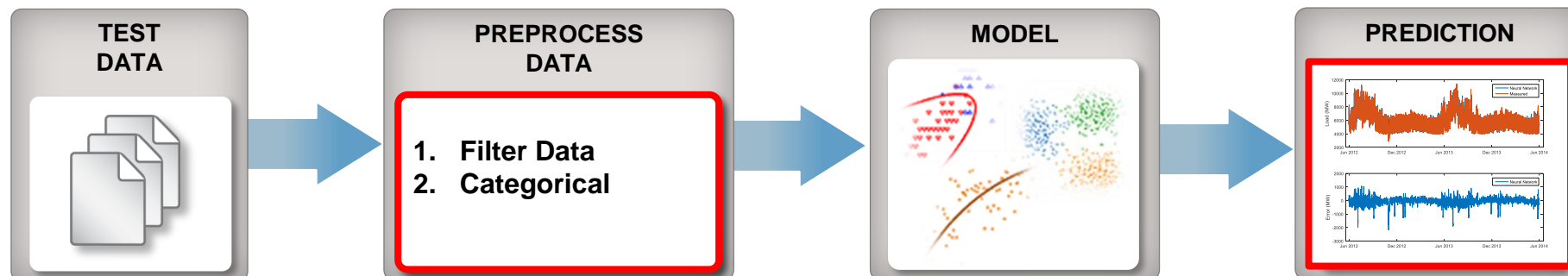


Machine Learning Workflow

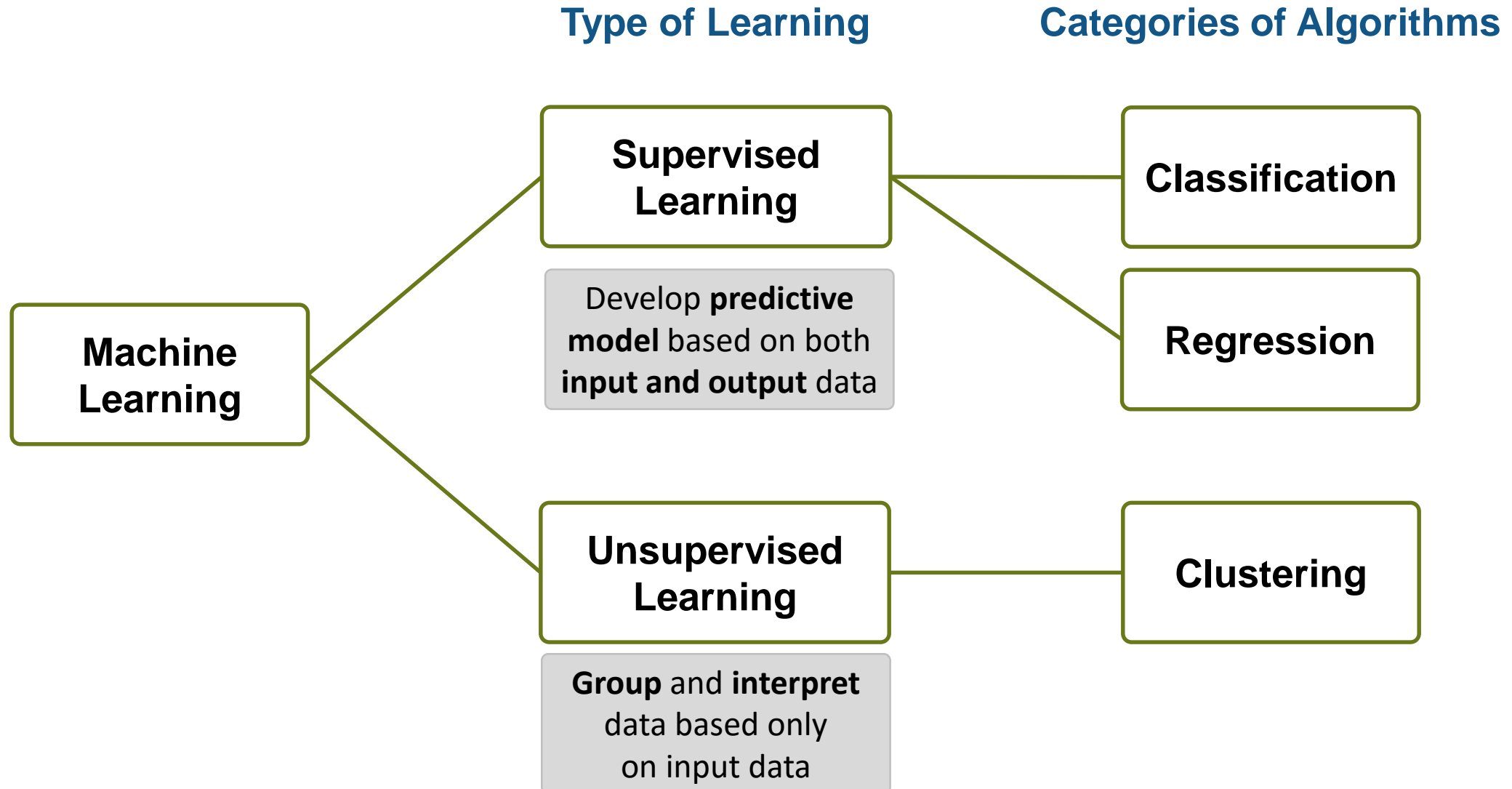
Train: Iterate until you find the best model



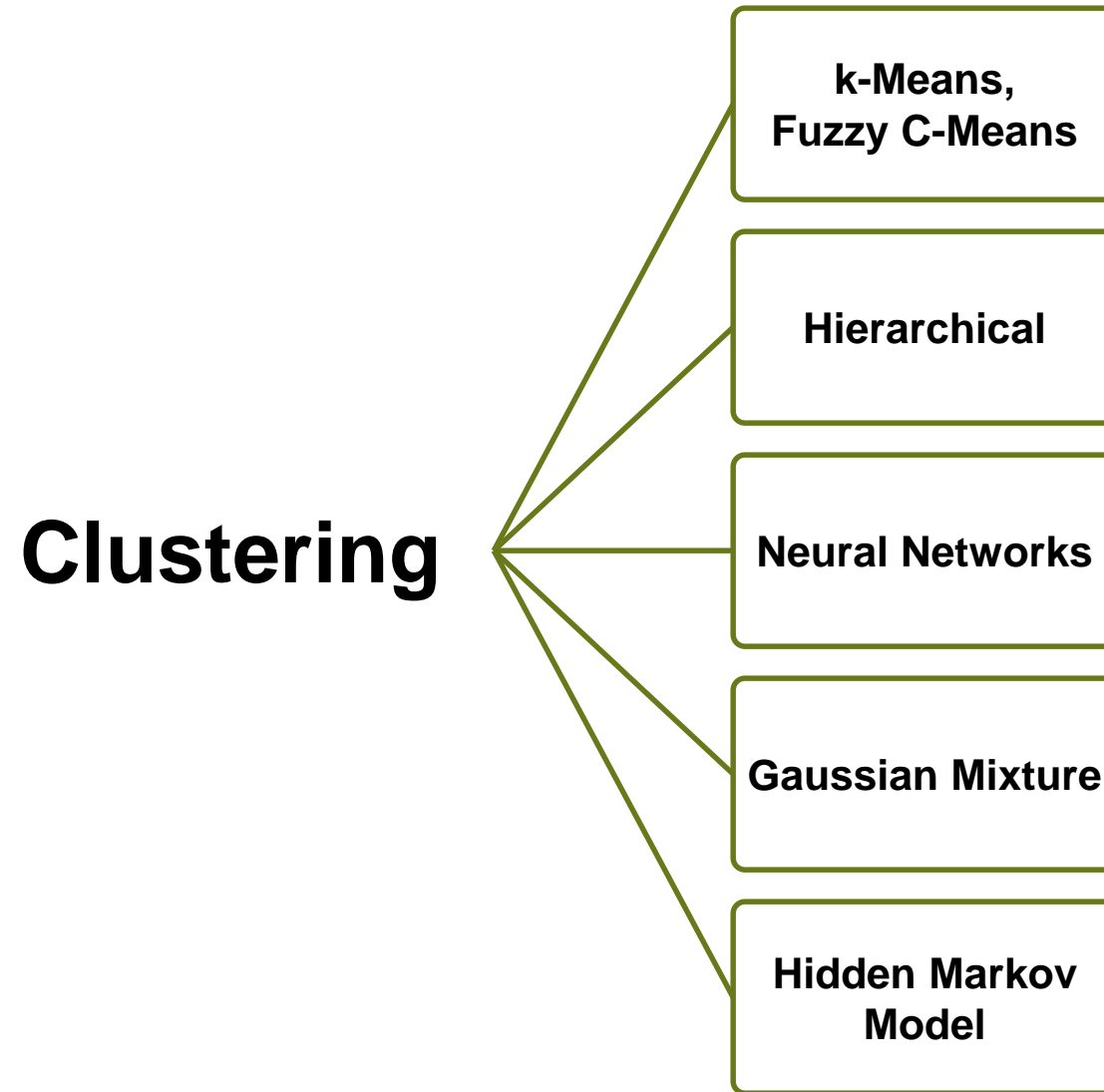
Predict: Integrate trained models into applications



Overview – Machine Learning

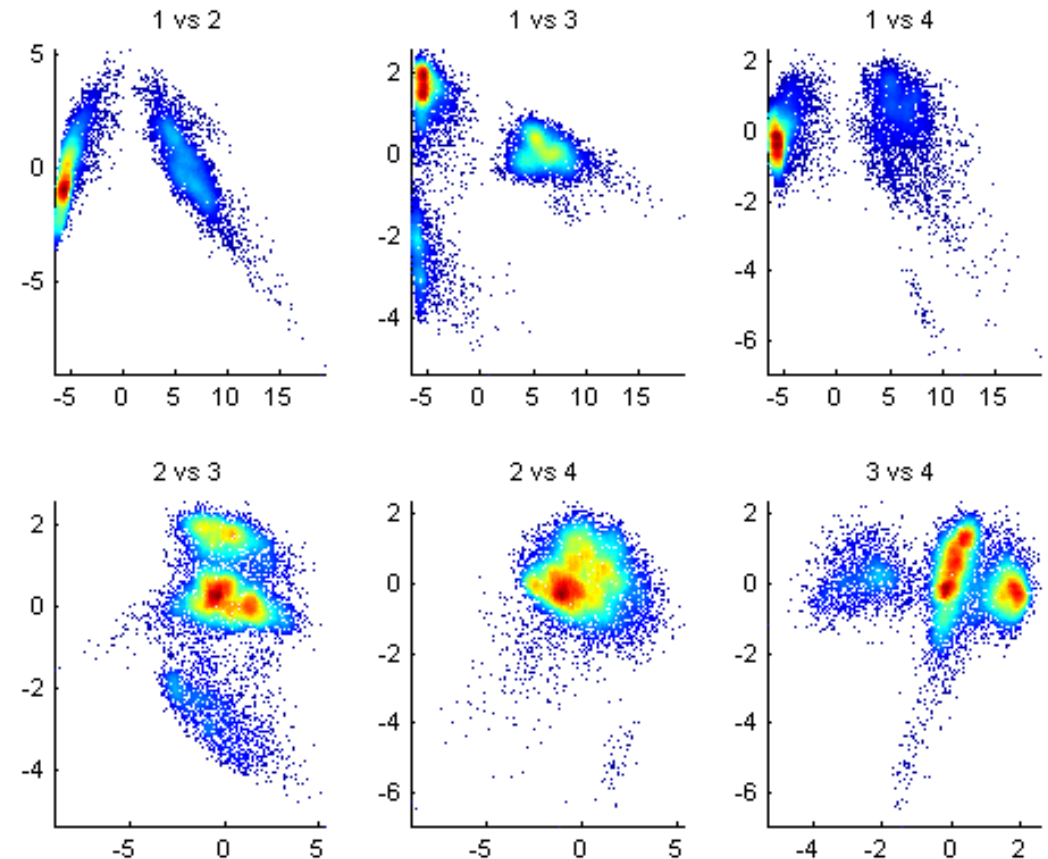


Unsupervised Learning

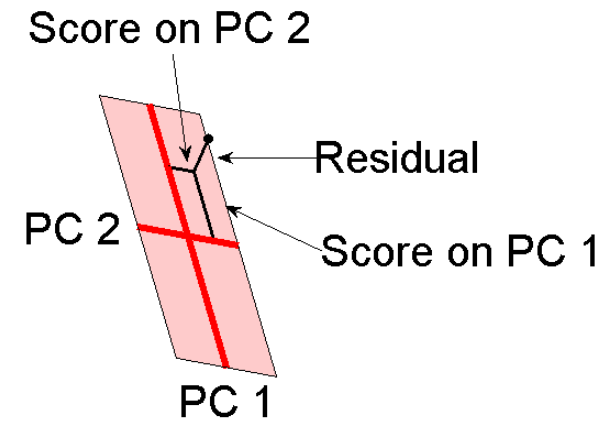
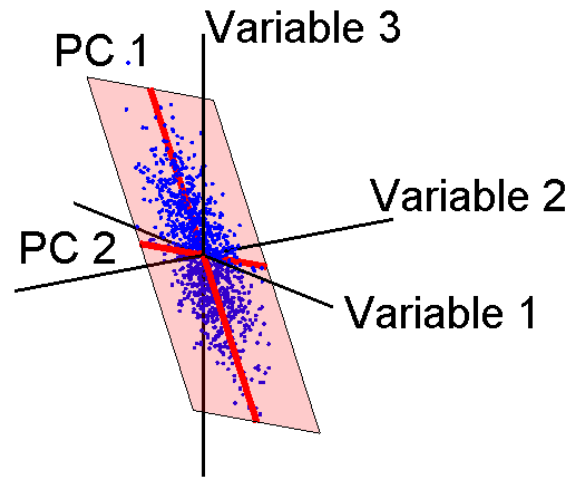
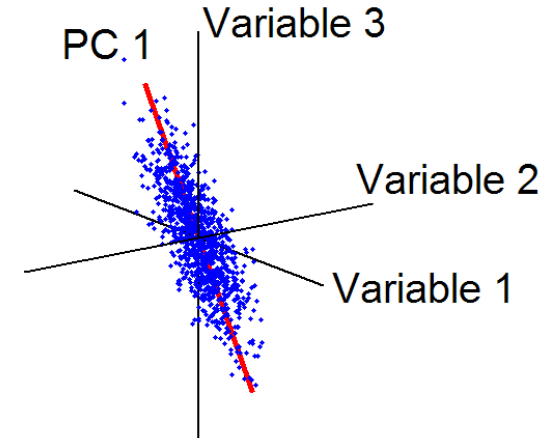
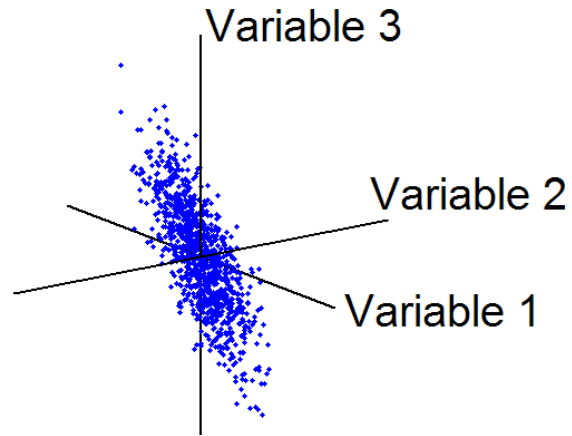


Example: Human Activity Recognition

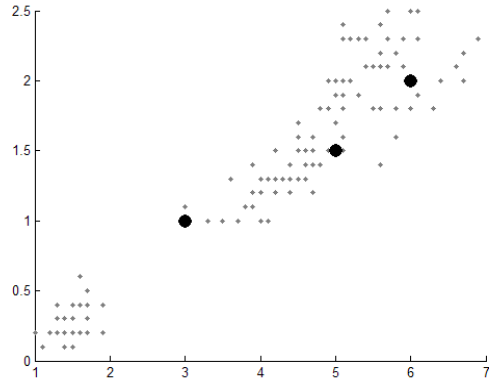
- Challenge:
 - Find natural groupings among large number of predictors
 - Build predictive model for classifying new data points
- Approach:
 - Reduce dimensionality and visualize structure of data
 - Evaluate different clustering techniques to identify groups of behaviors
 - Determine types of activities for new data points using classification techniques



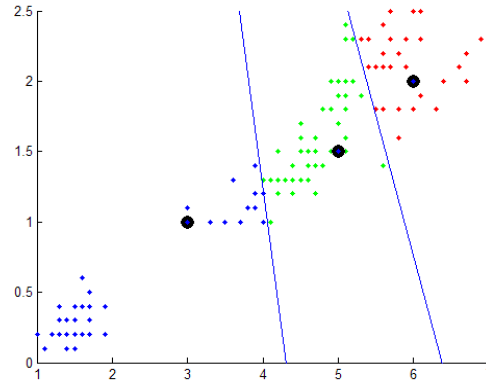
Principal Components Analysis – what is it doing?



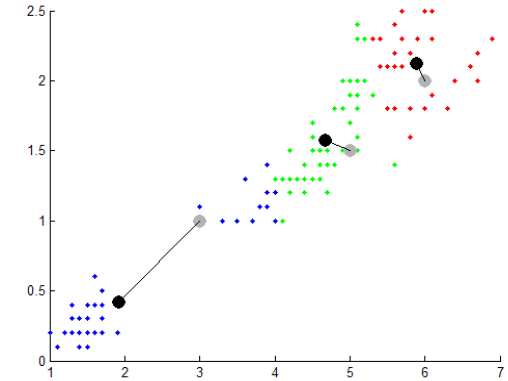
K-Means Cluster Analysis – what is it doing?



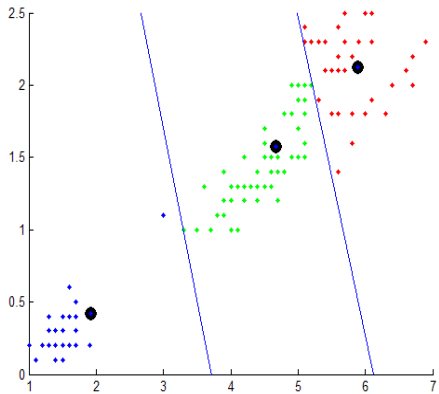
Randomly pick K cluster centroids



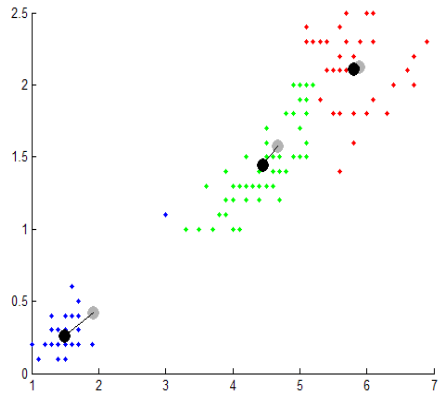
Assign points to the closest centroid



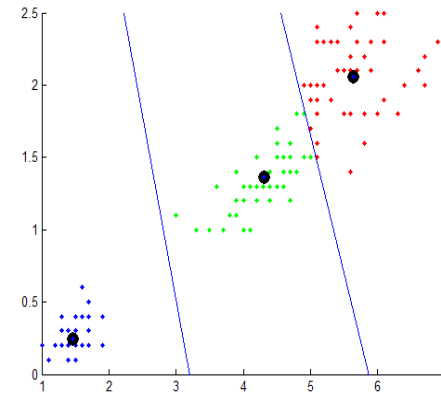
Recalculate positions of cluster centroids



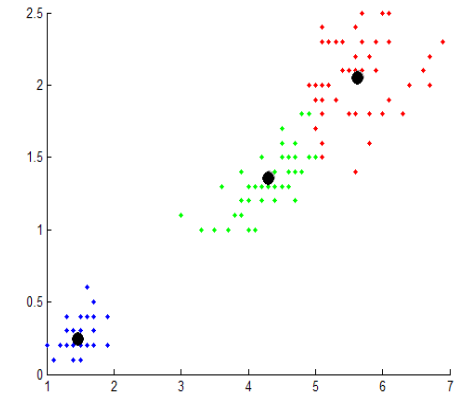
Reassign points to the closest centroid



Recalculate positions of cluster centroids

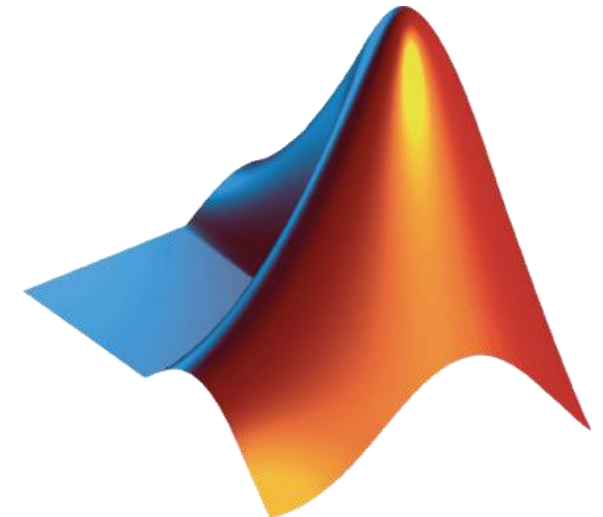


Repeat until centroid positions converge

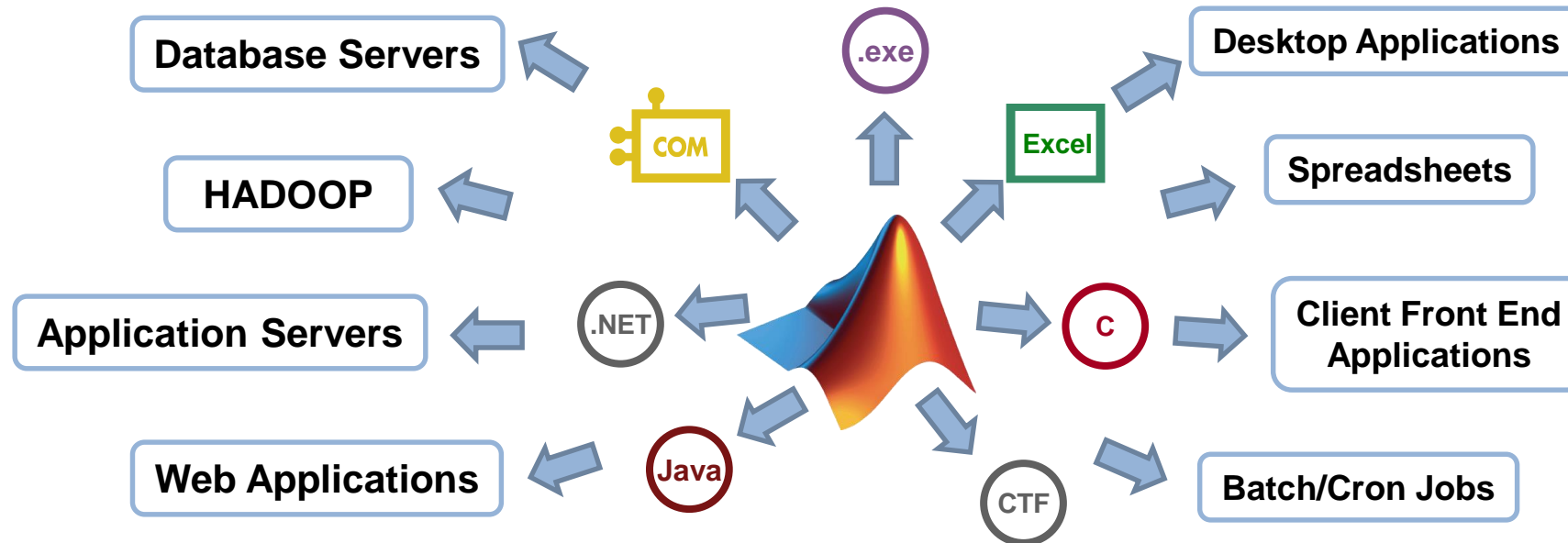


Machine Learning with MATLAB

- Interactive environment
 - Visual tools for exploratory data analysis
 - Easy to evaluate and choose best algorithm
 - Apps available to help you get started
(*e.g., neural network tool, curve fitting tool*)
- Multiple algorithms to choose from
 - Classification
 - Regression
 - Clustering



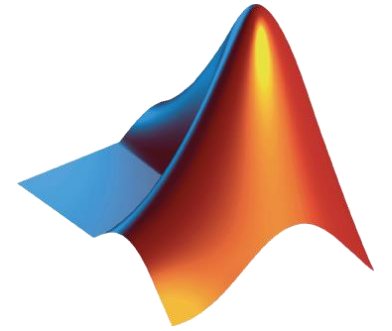
Deployment



- Royalty-free deployment
- Point-and-click workflow
- Unified process for desktop and server apps

Key Takeaways

Consider Machine Learning when:



- Hand written rules and equations are too complex
 - *Face recognition, speech recognition, recognizing patterns*
- Rules of a task are constantly changing
 - *Fraud detection from transactions, anomaly in sensor data*
- Nature of the data changes and the program needs to adapt
 - *Automated trading, energy demand forecasting, predicting shopping trends*



**“Essentially, all models are wrong,
but some are useful.”**

– George Box

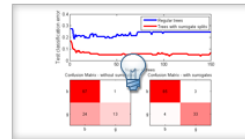
Learn More: Machine Learning with MATLAB

mathworks.com/machine-learning

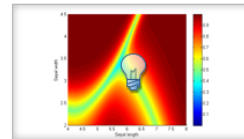
Classification Examples



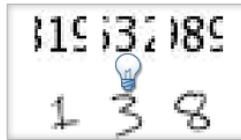
Basket Selection Using Stepwise Regression



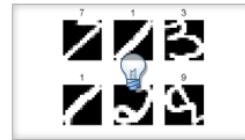
Classification in the Presence of Missing Data



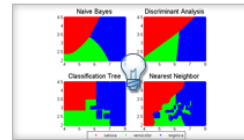
Classification Probability



Digit Classification Using HOG Features

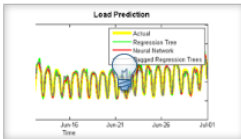


Handwriting Recognition Using Bagged Classification Trees

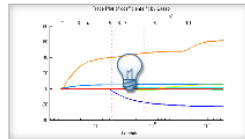


Visualize Decision Surfaces for Different Classifiers

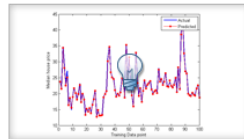
Regression Examples



Electricity Load Forecasting

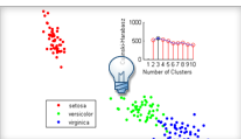


Lasso Regularization



Regression with Boosted Decision Trees

Clustering Examples



Cluster Evaluation



Cluster Genes Using K-Means and Self-Organizing Maps



Color-Based Segmentation Using K-Means Clustering

Machine learning algorithms use computational methods to “learn” information directly from data without assuming a predetermined equation as a model. They can adaptively improve their performance as you increase the number of samples available for learning.

Machine learning algorithms are used in applications such as computational finance (credit scoring and algorithmic trading), computational biology (tumor detection, drug discovery, and DNA sequencing), energy production (price and load forecasting), natural language processing, speech and image recognition, and advertising and recommendation systems.

Machine learning is often used in big data applications, which have large datasets with many predictors (features) and are too complex for a simple parametric model. Examples of big data applications include forecasting electricity load with a neural network, or bond rating classification for credit risk using an ensemble of decision trees.

Classification

Build models to classify data into different categories.



Regression

Build models to predict continuous data.



Clustering

Find natural groupings and patterns in data.



Questions?