Machine Learning with MATLAB

Gerardo Hernández Correa
Application Engineer
What You Will Learn

- Overview of Machine Learning
- Algorithms available in MATLAB
- Overcoming machine learning challenges with MATLAB
What is Machine Learning?
Most common tool for Data Analytics modeling

Use features in the data to create a predictive model
Machine Learning for Data Analytics

"Data analytics solutions allow firms to discover, optimize, and deploy **predictive models** by analyzing **data sources** to improve **business outcomes**."

Forrester Research

**Challenges**
- Overloaded by data
- Competition for better decision-making
- Big Data buzz, but missing solution to make sense of it (**analytics**) and integrate with enterprise-wide applications (**deployment**)
Used Across Many Application Areas

- **Biology**: Tumor Detection, Drug Discovery
- **Financial Services**: Credit Scoring, Algorithm Trading, Bond Classification
- **Image & Video Processing**: Pattern Recognition
- **Audio Processing**: Speech Recognition
- **Energy**: Load, Price Forecasting, Trading
Challenges – Machine Learning

- Lots of data, with many variables (predictors)
- Data is too complex to know the governing equations
- Significant technical expertise required
- No “one size fits all” solution \(\Rightarrow\) requires an iterative approach
  - Try multiple algorithms, see what works best
  - Time consuming
MATLAB Solution

- Strong environment for **interactive** exploration
- **Algorithms** and **Apps** to get started
  - Clustering, Classification, Regression
  - Neural network app, Curve fitting app
- Easy to evaluate, **iterate** and choose the best algorithm
- **Parallel computing**
- **Integrated** with data and deployment for Data Analytics workflows
Overview – Machine Learning

Type of Learning

- **Unsupervised Learning**
  - Group and interpret data based only on input data

- **Supervised Learning**
  - Develop predictive model based on both input and output data

Categories of Algorithms

- **Clustering**
- **Classification**
- **Regression**
Unsupervised Learning

Clustering

- k-Means, Fuzzy C-Means
- Hierarchical
- Neural Networks
- Gaussian Mixture
- Hidden Markov Model
Supervised Learning

Regression

- Neural Networks
- Decision Trees
- Ensemble Methods
- Non-linear Reg. (GLM, Logistic)
- Linear Regression

Classification

- Support Vector Machines
- Discriminant Analysis
- Naive Bayes
- Nearest Neighbor
Supervised Learning - Workflow

Data
- Import Data
- Explore Data
- Prepare Data

Select Model

Train the Model
- Known data
- Known responses
- Model

Measure Accuracy

Use for Prediction
- Model
- Predicted Responses
- New Data

Speed up Computations
Classification
Overview

- What is classification?
  - Predicting the best group for each point
  - “Learns” from labeled observations
  - Uses input features

- Why use classification?
  - Accurately group data never seen before

- How is classification done?
  - Can use several algorithms to build a predictive model
  - Good training data is critical
Example – Bank Marketing Campaign

- **Goal:**
  - Predict if customer would subscribe to bank term deposit based on different attributes

- **Approach:**
  - Train a classifier using different models
  - Measure accuracy and compare models
  - Reduce model complexity
  - Use classifier for prediction

Data set downloaded from UCI Machine Learning repository
http://archive.ics.uci.edu/ml/datasets/Bank+Marketing
Example – Bank Marketing Campaign

- Numerous predictive models with rich documentation
- Interactive visualizations and apps to aid discovery
- Built-in parallel computing support
- Quick prototyping; Focus on modeling not programming
Clustering
Overview

- What is clustering?
  - Segment data into groups, based on data similarity

- Why use clustering?
  - Identify outliers
  - Resulting groups may be the matter of interest

- How is clustering done?
  - Can be achieved by various algorithms
  - It is an iterative process (*involving trial and error*)
Example – Clustering Corporate Bonds

- **Goal:**
  - Cluster similar corporate bonds together

- **Approach:**
  - Cluster the bonds data using distance-based and probability-based techniques
  - Evaluate clusters for validity
Example – Clustering Corporate Bonds

- Numerous clustering functions with rich documentation
- Interactive visualizations to aid discovery
- Viewable source; not a black box
- Rapid exploration & development
# MATLAB for Machine Learning

<table>
<thead>
<tr>
<th>Challenges</th>
<th>MATLAB Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (loss of productivity)</td>
<td><strong>Rapid analysis and application development</strong>&lt;br&gt;High productivity from data preparation, interactive exploration, visualizations.</td>
</tr>
<tr>
<td>Extract value from data</td>
<td><strong>Machine learning, Video, Image, and Financial</strong>&lt;br&gt;Depth and breadth of algorithms in classification, clustering, and regression</td>
</tr>
<tr>
<td>Computation speed</td>
<td><strong>Fast training and computation</strong>&lt;br&gt;Parallel computation, Optimized libraries</td>
</tr>
<tr>
<td>Time to deploy &amp; integrate</td>
<td><strong>Ease of deployment and leveraging enterprise</strong>&lt;br&gt;Push-button deployment into production</td>
</tr>
<tr>
<td>Technology risk</td>
<td><strong>High-quality libraries and support</strong>&lt;br&gt;Industry-standard algorithms in use in production&lt;br&gt;Access to support, training and advisory services when needed</td>
</tr>
</tbody>
</table>
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 Stacked 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.