Machine Learning for Financial Applications

Dr. Alexander Diethert, Application Engineer
Challenges for Analysts

How to work with your data to

... analyze for insights?
... identify patterns?
... develop predictive models?
... deploy to collaborators?
... innovate?
Customer Examples

MATLAB Used to Predict Financial Crises in Emerging Markets

“Because MATLAB is both powerful and easy to use, I felt confident that the Bank of Indonesia would be able to implement the MATLAB programs and use them as an early warning system for financial distress.”

— Dr. Paul McNelis, Georgetown University

Machine Learning and Visualisation in the Context of a Large Enterprise
Arjun Viswanathan, Citi

Gas Natural Fenosa Predicts Energy Supply and Demand

“Because we need to rapidly respond to shifting production constraints and changing demands, we cannot depend on closed or proprietary solutions. With MathWorks tools we get more accurate results—and we have the flexibility to develop, update, and optimize our models in response to changing needs.”

— Angel Caballero, Gas Natural Fenosa

Machine Learning and Applications in Finance
Christian Hesse, Deutsche Bank and University College London

Portfolio Allocation with Machine Learning and MATLAB Distributed Computing Server on Microsoft Azure Cloud
Emilio Llorente-Cano and James Mann, Aberdeen Asset Management
Agenda

Introduction

- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary
Computational Finance Workflow

Access
- Files
- Databases
- Datafeeds

Research and Quantify
- Data Analysis & Visualization

Share
- Reporting
- Applications
- Production

Automate
- Files
- Databases
- Datafeeds
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*)
Challenges – Machine Learning

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

- Introduction

Unsupervised Learning: Clustering Bond Data

- Supervised Learning: Create Rating System

- Neural Network Based Time Series Forecasting

Outlook: Distributing MATLAB Applications

- Summary
Unsupervised Learning

Clustering

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

Clustering

- k-Means, Fuzzy C-Means
- Hierarchical
- Neural Networks
- Gaussian Mixture
- Hidden Markov Model
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
Example – Clustering Corporate Bonds

- Numerous clustering functions with rich documentation
- Interactive visualizations to aid discovery
- Viewable source; not a black box
- Rapid exploration & development
Classification Learner App
Agenda

- Introduction

- Unsupervised Learning: Clustering Bond Data

Supervised Learning: Create Rating System

- Neural Network Based Time Series Forecasting

- Outlook: Distributing MATLAB Applications

- Summary
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

**Train the Model**
- Known data
- Known responses
- Model

**Use for Prediction**
- Model
- Predicted Responses
- New Data

**Select Model**

**Measure Accuracy**

**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: Calibrating a rating system

Key tasks
1. Perform *ad hoc* data analysis (linear regression)
2. Improve existing analysis (logistic regression)
3. Create a robust classifier: TreeBagger
Regression Learner App

Predictions: model 4

Dataset: cartable  Observations: 406  Size: 30 kB  Predictors: 7  Response: MPG  Validation: 5-fold Cross-Validation
Agenda

- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary
Neural Network Use Cases

- Function Approximation and Nonlinear Regression
- Pattern Recognition and Classification
- Clustering
  - Self-Organizing Maps
  - Competitive Layers
- **Time Series** and Dynamic Systems
  - Modeling and Prediction with NARX and Time-Delay Networks
  - Creating Simulink Models
Neural Network Apps

Welcome to the Neural Network Time Series Tool.
Solve a nonlinear time series problem with a dynamic neural network.

Introduction
Prediction is a kind of dynamic filtering, in which past values of one or more time series are used to predict future values. Dynamic neural networks, which include tapped delay lines, are used for nonlinear filtering and prediction.

There are many applications for prediction. For example, a financial analyst might want to predict the future value of a stock, bond or other financial instrument. An engineer might want to predict the impending failure of a jet engine.

Predictive models are also used for system identification (or dynamic modelling), in which you build dynamic models of physical systems. These dynamic models are important for analysis, simulation, monitoring and control of a variety of systems, including manufacturing systems, chemical processes, robotics and aerospace systems.

This tool allows you to solve three kinds of nonlinear time series problems shown in the right panel. Choose one and click [Next].

Select a Problem
- **Nonlinear Autoregressive with External (Exogenous) Input (NARX)**
  Predict series $y(t)$ given $d$ past values of $y(t)$ and another series $x(t)$.
  $$y(t) = f(x(t-1), ..., x(t-d), y(t-1), ..., y(t-d))$$

- **Nonlinear Autoregressive (NAR)**
  Predict series $y(t)$ given $d$ past values of $y(t)$.
  $$y(t) = f(y(t-1), ..., y(t-d))$$

- **Nonlinear Input-Output**
  Predict series $y(t)$ given $d$ past values of series $x(t)$.
  **Important Note**: NARX solutions are more accurate than this solution. Only use this solution if past values of $y(t)$ will not be available when deployed.
  $$y(t) = f(x(t-1), ..., x(t-d))$$

To continue, click [Next].
Agenda

 Introduction
 Unsupervised Learning: Clustering Bond Data
 Supervised Learning: Create Rating System
 Neural Network Based Time Series Forecasting

Outlook: Distributing MATLAB Applications

 Summary
Deploy your MATLAB code to people who don’t need MATLAB
Sharing Standalone Applications

1. Application Author
   - MATLAB
     - Toolboxes
2. MATLAB Compiler
   - Standalone Application
   - Excel Add-in
   - Hadoop
3. End User
   - MATLAB Runtime
Which Product will Fit Your Needs?

**MATLAB Compiler** for sharing MATLAB programs without integration programming

**MATLAB Compiler SDK** provides implementation and platform flexibility for software developers

**MATLAB Production Server** provides the most efficient development path for secure and scalable web and enterprise applications
The Range of Application Platforms

Standalone Applications

Web and Enterprise Applications

| Scale of Distribution | Application Complexity | Enterprise Integration |
MATLAB Production Server

Web Applications

Web Server

Application Server

MATLAB Component

MATLAB Production Server

Application 1

Application 2

Application 3

Database Server

Warehouse

Trade Execution; Batch Applications

Desktop Applications

MATLAB Component

Web Applications

Web Server

Application Server

MATLAB Production Server

Application 1

Application 2

Application 3

Database Server

Warehouse

Trade Execution; Batch Applications

Desktop Applications
Agenda

- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications

Summary
# 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></td>
</tr>
<tr>
<td></td>
<td>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></td>
</tr>
<tr>
<td></td>
<td>Depth and breadth of algorithms in classification, clustering, and regression</td>
</tr>
<tr>
<td>Computation speed</td>
<td><strong>Fast training and computation</strong></td>
</tr>
<tr>
<td></td>
<td>Parallel computation, Optimized libraries</td>
</tr>
<tr>
<td>Time to deploy &amp; integrate</td>
<td><strong>Ease of deployment and leveraging enterprise</strong></td>
</tr>
<tr>
<td></td>
<td>Push-button deployment into production</td>
</tr>
<tr>
<td>Technology risk</td>
<td><strong>High-quality libraries and support</strong></td>
</tr>
<tr>
<td></td>
<td>Industry-standard algorithms in use in production</td>
</tr>
<tr>
<td></td>
<td>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 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.
Thank you!