MathWorks Seminars at

Session 1: 1:00 – 2:30 pm
Data Analytics and Machine Learning using MATLAB

Break: 15 minutes

Session 2: 2:45 – 4:15 pm
Optimizing and Accelerating MATLAB Code
Data Analytics and Machine Learning using MATLAB

Tom McHugh  Account Manager
Saket Kharsikar  Application Engineer
What is Data Analytics?

*Turn large volumes of complex data into actionable information*

- **Descriptive**
  - What happened?

- **Diagnostic**
  - Why did it happen?

- **Predictive**
  - What will happen?

- **Prescriptive**
  - What should be done?
Data Analytics Workflow

Access and Explore Data
- Files
- Databases
- Sensors

Preprocess Data
- Working with Messy Data
- Data Reduction/Transformation
- Feature Extraction

Develop Predictive Models
- Model Creation e.g. Machine Learning
- Parameter Optimization
- Model Validation

Integrate Analytics with Systems
- Desktop Apps
- Enterprise Scale Systems
- Embedded Devices and Hardware
Data Analytics Workflow

Access and Explore Data

Files
Databases
Sensors

Preprocess Data

Working with Messy Data
Data Reduction/Transformation
Feature Extraction

Develop Predictive Models

Model Creation e.g. Machine Learning
Parameter Optimization
Model Validation

Integrate Analytics with Systems

Desktop Apps
Enterprise Scale Systems
Embedded Devices and Hardware
Data Analytics Workflow

Access and Explore Data

Files
- Repositories – SQL, NoSQL, etc.
- File I/O – Text, Spreadsheet, etc.
- Web Sources – RESTful, JSON, etc.

Databases

Sensors
- Real-Time Sources – Sensors, GPS, etc.
- File I/O – Image, Audio, etc.
- Communication Protocols – OPC (OLE for Process Control), CAN (Controller Area Network), etc.

Preprocess Data

Business and Transactional Data

- Repositories – SQL, NoSQL, etc.
- File I/O – Text, Spreadsheet, etc.
- Web Sources – RESTful, JSON, etc.

Engineering, Scientific and Field Data

- Real-Time Sources – Sensors, GPS, etc.
- File I/O – Image, Audio, etc.
- Communication Protocols – OPC (OLE for Process Control), CAN (Controller Area Network), etc.

Working with Messy Data

Data Reduction/Transformation

Feature Extraction
Data Analytics Workflow

Access and Explore Data

Files

Databases

Sensors

Preprocess Data

Working with Messy Data

Data Reduction/Transformation

Feature Extraction

Challenges

- Data aggregation
  - Different sources (files, web, etc.)
  - Different types (images, text, audio, etc.)

- Data clean up
  - Poorly formatted files
  - Irregularly sampled data
  - Redundant data, outliers, missing data etc.

- Data specific processing
  - Signals: Smoothing, resampling, denoising, Wavelet transforms, etc.
  - Images: Image registration, morphological filtering, deblurring, etc.

- Dealing with out of memory data (big data)
Case Study: Day-Ahead Load Forecasting

- **Goal:**
  - Implement a tool for *easy* and *accurate* computation of day-ahead system load forecast

- **Requirements:**
  - Acquire and clean data from multiple sources
  - Accurate predictive model
  - Easily deploy to production environment
Techniques to Handle Missing Data

- List-wise deletion
  - Unbiased estimates
  - Reduces sample size

- Implementation options
  - Built in to many MATLAB functions
  - Manual filtering
Techniques to Handle Missing Data

Substitution – replace missing data points with a reasonable approximation

Easy to model

Too important to exclude
Merge Different Sets of Data

- Join along a common axis

- Popular Joins:
  - Inner
  - Full Outer
  - Left Outer
  - Right Outer
# Full Outer Join

## First Data Set

<table>
<thead>
<tr>
<th>Key</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>7</td>
<td>1.7</td>
</tr>
<tr>
<td>9</td>
<td>1.9</td>
</tr>
</tbody>
</table>

## Second Data Set

<table>
<thead>
<tr>
<th>Key</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>0.7</td>
<td>0.8</td>
</tr>
</tbody>
</table>

## Joined Data Set

<table>
<thead>
<tr>
<th>Key</th>
<th>B</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>NaN</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>1.4</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td>5</td>
<td>NaN</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>1.7</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>9</td>
<td>1.9</td>
<td>NaN</td>
<td>NaN</td>
</tr>
</tbody>
</table>
Data Analytics Workflow

1. MATLAB Analytics work with business and engineering data

Access and Explore Data

- Files
- Databases
- Sensors

Preprocess Data

- Working with Messy Data
- Data Reduction/Transformation
- Feature Extraction

- Point and click tools to access variety of data sources
- High-performance environment for big data
- Built-in algorithms for data preprocessing including sensor, image, audio, video and other real-time data
Learn More: Big Data with MATLAB

www.mathworks.com/discovery/matlab-mapreduce-hadoop.html

MapReduce on the Desktop
Explore and analyze big data sets on your desktop with the MapReduce programming technique built into MATLAB.

Creating algorithms using MapReduce: max, mean, mean by group, histograms, covariance and related quantities, summary statistics by group, logistic regression, tall skinny QR

» Get started with MATLAB MapReduce
» MapReduce design patterns
» Use MATLAB MapReduce with relational databases

MapReduce on Hadoop
Execute MATLAB MapReduce based algorithms within Hadoop MapReduce to explore and analyze data that is stored and managed on Hadoop, using MATLAB Distributed Computing Server.

» Run MATLAB MapReduce on Hadoop

Create applications and libraries based upon MATLAB MapReduce for deployment within production instances of Hadoop, using MATLAB Compiler.

» Deploy MATLAB MapReduce applications to Hadoop
Data Analytics Workflow

Access and Explore Data

Files

Databases

Sensors

Preprocess Data

Working with Messy Data

Data Reduction/Transformation

Feature Extraction

Develop Predictive Models

Model Creation e.g. Machine Learning

Parameter Optimization

Model Validation

Integrate Analytics with Systems

Desktop Apps

Enterprise Scale Systems

Embedded Devices and Hardware
Data Analytics Workflow

**Challenges**

- Lack of data science expertise
- Feature Extraction – How to transform data to best represent the system?
  - Requires subject matter expertise
  - No right way of designing features
- Feature Selection – What attributes or subset of data to use?
  - Entails a lot of iteration – Trial and error
  - Difficult to evaluate features
- Model Development
  - Many different models
  - Model Validation and Tuning
- Time required to conduct the analysis

**Preprocess Data**

- Working with Messy Data
- Data Reduction/Transformation
- Feature Extraction

**Develop Predictive Models**

- Model Creation e.g. Machine Learning
- Parameter Optimization
- Model Validation
Machine Learning
Characteristics and Examples

- **Characteristics**
  - Lots of 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*)
Overview – Machine Learning

**Type of Learning**

- **Supervised Learning**
  - Develop *predictive model* based on both input and output data
- **Unsupervised Learning**
  - Group and interpret data based only on input data

**Categories of Algorithms**

- Classification
- Regression
- Clustering
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
Unsupervised Learning

Clustering

- k-Means, Fuzzy C-Means
- Hierarchical
- Neural Networks
- Gaussian Mixture
- Hidden Markov Model
Data Analytics Workflow

MATLAB enables domain experts to do Data Science

Preprocess Data

- Working with Messy Data
- Data Reduction/Transformation
- Feature Extraction

Apps

- Easy to use apps
- Wide breadth of tools to facilitate domain specific analysis
- Examples/videos to get started

Language

- Automatic MATLAB code generation
- High speed processing of large data sets

Develop Predictive Models

- Model Creation e.g. Machine Learning
- Parameter Optimization
- Model Validation

Apps

- Easy to use apps
- Wide breadth of tools to facilitate domain specific analysis
- Examples/videos to get started

Language

- Automatic MATLAB code generation
- High speed processing of large data sets

Apps

- Easy to use apps
- Wide breadth of tools to facilitate domain specific analysis
- Examples/videos to get started

Language

- Automatic MATLAB code generation
- High speed processing of large data sets

Apps

- Easy to use apps
- Wide breadth of tools to facilitate domain specific analysis
- Examples/videos to get started

Language

- Automatic MATLAB code generation
- High speed processing of large data sets
Learn More: Machine Learning with MATLAB

mathworks.com/machine-learning
Data Analytics Workflow

Access and Explore Data

Preprocess Data

Develop Predictive Models

Integrate Analytics with Systems

1. Files
   - Databases
   - Sensors

2. Working with Messy Data
   - Data Reduction/Transformation
   - Feature Extraction

3. Model Creation e.g. Machine Learning
   - Parameter Optimization
   - Model Validation

Desktop Apps

Enterprise Scale Systems

Embedded Devices and Hardware
Data Analytics Workflow

Develop Predictive Models

Model Creation e.g. Machine Learning

Parameter Optimization

Model Validation

Integrate Analytics with Systems

Desktop Apps

Enterprise Scale Systems

Embedded Devices and Hardware

Challenges

- End user: Operators, Analysts, Administrative Staff, customers etc.
- Different target platforms:
  - Cluster or Cloud environment
  - Standalone desktop applications
  - Server based Web and enterprise systems
  - Embedded hardware
- Different Interfaces: C++, Java, Python, .NET etc.
- Need to translate analytics to production environment
Integrate analytics with systems

MATLAB Analytics
run anywhere

Embedded Hardware
C, C++, HDL, PLC

Enterprise Systems
Standalone Application, Excel Add-in, Hadoop/Spark, C/C++, Java, Python, .NET, MATLAB Production Server

MATLAB
Production
Server

C, C++, HDL, PLC

for k=1:max
x = fft(dat)
y = 20*log1

for k=1:max
x = fft(dat)
y = 20*log1

MATLAB Runtime

3
Deployed Analytics

MATLAB Production Server

Web Application Server

- Apache Tomcat
- Web Server/Webservice

MATLAB Production Server

MATLAB Desktop

- Train in MATLAB
- Predictive Models
- Weather Data
- Energy Data

Weather Data

Energy Data

Predictive Models
Learn More: Application Deployment with MATLAB

www.mathworks.com/solutions/desktop-web-deployment/

Deploying MATLAB Code as an Executable or Software Component

Using MathWorks application deployment products, you can eliminate the costly and error-prone work of recoding your MATLAB algorithms in another programming language. Because you maintain your source code in MATLAB, you can easily develop and update your algorithms and automatically package them as standalone executables or software components for integration in environments such as C, C++, Java™, .NET, and Excel®.

MATLAB Compiler packages your MATLAB applications as encrypted standalone executables or C/C++ shared libraries. MATLAB builder products work in conjunction with MATLAB Compiler to create standard components for use with Java, .NET, or Excel. These executables and components can be deployed royalty-free on operating systems supported by MATLAB.

Deploying MATLAB Code as a Web Application

Using MathWorks application deployment products, you can develop MATLAB-based components for the Web that execute mathematical computations and generate interactive graphics. After developing an algorithm in MATLAB, you can automatically create a standard component designed to integrate in a Web application using MATLAB builder products for either Java or .NET.

Once you place the component on a Web server, your users access the application through a Web browser and do not need to install additional software on their desktop computers.

The Java and .NET components created by the deployment tools can be used in conjunction with standard Web technologies such as ASP.NET, SDA, Sawg, JavaScript, and HTML.
Key Takeaways

1. MATLAB Analytics work with **business** and **engineering data**
2. MATLAB enables **domain experts to do Data Science**
3. MATLAB Analytics **run anywhere**

- Utilize all of your data.
- Apply advanced analytics techniques.
- Operationalize analytics to enterprise systems and embedded devices.