Machine Learning
Proven Applications and New Features

Seth DeLand

MathWorks
How to Get Started with Machine Learning?

get started with machine learning

About 611,000,000 results (0.63 seconds)
Kinesis Health Technologies
Predicting a patient’s fall risk with machine learning.
Machine Learning + X
Machine Learning

+ 

Industry Knowledge  Application Knowledge

Your Own Expertise
Examples of Successful Machine Learning Applications

- Fleet Data Analytics
- Energy Forecasting
- Manufacturing Analytics

New Capabilities
- MATLAB apps
- AutoML
- Signal Processing with Machine Learning
- C/C++ Code Generation
Examples of Successful Machine Learning Applications

- Fleet Data Analytics
- Energy Forecasting
- Manufacturing Analytics
Fleet Data Analytics

Design Decisions

Test Plans
What Level of Data?

Equipment

Trip/Session

Messages

Signals

Time – Value pairs
For each (trip, day, serial #, customer, etc) in the fleet data set, calculate some Key Performance Indicator (KPI*) given parameters XYZ".

Across All (data) in the fleet data set, calculate descriptive statistics of specific variables (min, max, median, count, etc.) to summarize and visualize (histograms).
Scale to Large Collections of Data with Datastore

Create a datastore from all CSV files

```
    ds = datastore('*.csv')
```

Read a single file of data

```
    data = read(ds);
```

Reset the datastore back to the first file

```
    reset(ds);
```

Find the maximum value of “Y” in each file

```
    X = [];
    while hasdata(ds)
        data = read(ds);
        X(end+1) = max(data.Y);
    end
```

Available Datastores

<table>
<thead>
<tr>
<th>General</th>
<th>datastore</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>spreadsheetDatastore</td>
</tr>
<tr>
<td></td>
<td>tabularTextDatastore</td>
</tr>
<tr>
<td></td>
<td>fileDatastore</td>
</tr>
<tr>
<td>Database</td>
<td>databaseDatastore</td>
</tr>
<tr>
<td>Image</td>
<td>imageDatastore</td>
</tr>
<tr>
<td></td>
<td>denoisingImageDatastore</td>
</tr>
<tr>
<td></td>
<td>randomPatchExtractionDatastore</td>
</tr>
<tr>
<td></td>
<td>pixelLabelDatastore</td>
</tr>
<tr>
<td></td>
<td>augmentedImageDatastore</td>
</tr>
<tr>
<td>Audio</td>
<td>audioDatastore</td>
</tr>
<tr>
<td>Predictive Maintenance</td>
<td>fileEnsembleDatastore</td>
</tr>
<tr>
<td></td>
<td>simulationEnsembleDatastore</td>
</tr>
<tr>
<td>Simulink</td>
<td>SimulationDatastore</td>
</tr>
<tr>
<td>Automotive</td>
<td>mdfDatastore</td>
</tr>
<tr>
<td>Custom</td>
<td>subclass matlab.io.Datastore</td>
</tr>
<tr>
<td>Transformed</td>
<td>transform an existing datastore</td>
</tr>
</tbody>
</table>
Performing “Across All” Calculations with Tall

Create a datastore from a collection of CSV files, and select the “Time” and “EngineSpeedRPM” variables.

```matlab
ds = datastore('EngineData*.csv', 'SelectedVariableNames', ['Time', 'EngineSpeedRPM']);
```

Create tall table:

```matlab
t = tall(ds);
```

Convert to tall timetable:

```matlab
tt = table2timetable(t);
```

Plot EngineSpeedRPM vs. Time:

```matlab
plot(tt.Time, tt.EngineSpeedRPM)
```

- Visualizations
- Data preprocessing
- Machine Learning
Exploring Fleet Data with Unsupervised Learning
Unsupervised Learning for Operational Mode Clustering

Plot the raw data:

```matlab
figure;
plot(t.Speed_OBD_,t.EngineRPM,'.k')
xlabel('Vehicle Speed');
ylabel('Engine Speed');
```

Cluster the data with the K-Means algorithm:

```matlab
X = [t.Speed_OBD_,t.EngineRPM];
IDX = kmeans(X,5,"Distance","cosine");
```

Plot results of the clustering:

```matlab
gscatter(t.Speed_OBD_,t.EngineRPM,IDX);
xlabel('Vehicle Speed');
ylabel('Engine Speed');
```
Deploying Fleet Analytics

“Cold Storage”

Historic data:
- Batch processing
- Large data on cluster
- Explore long term trends
- Build models

“Hot Storage”

Streaming data:
- Near real-time
- Test and implement model for new data
- Stream processing

Vehicle data, driver profiles

Spark

kafka

MATLAB EXPO
Fleet Analytics in Practice: Volkswagen Data Lab

Develop technology building block for tailoring car features and services to individual

- Driver and Fleet Safety
- Driver Coaching
- Driver-Specific Insurance

Data sources

- Logged CAN bus data and travel record

Results

- Proof-of-concept model for “telematic fingerprint”
- Basis for the “pay-as-you-drive” concept

Source: “Connected Car – Fahrererkennung mit MATLAB”
Julia Fumbarev, Volkswagen Data Lab
MATLAB EXPO Germany, June 27, 2017, Munich Germany
Machine Learning + X

Fleet Analytics

**Equipment Expertise**
- Design Specs
- Operating Modes
- Operating Conditions

**Machine Learning**
- Statistical Analysis
- Unsupervised Learning

Energy Forecasting

Electricity Grid Expertise
- Seasonality
- Weather Effects
- Generator Characteristics

Manufacturing Analytics

Manufacturing Expertise
- Process Equipment
- Process Variables
- Performance Metrics

Anomaly Detection

Regression

Classification
Examples of Successful Machine Learning Applications

- Fleet Data Analytics
- Energy Forecasting
- Manufacturing Analytics
The Need for Energy Forecasts

Wind

Demand

Solar

Price
How Energy Forecasting Works

Historical Data

Electricity Demand

Weather

Electricity Prices

Combine

Preprocess

Features

Machine Learning

load wind 24hr
temp day 1week
month
Building Forecast Models with Regression Techniques

MACHINE LEARNING

SUPERVISED LEARNING

UNSUPERVISED LEARNING

CLUSTERING

K-Means, K-Medoids, Fuzzy C-Means
Hierarchical
Gaussian Mixture
Neural Networks
Hidden Markov Model
Using Energy Forecasting Models

New Data

Electricity Demand

Weather

Electricity Prices

Combine

Features

load  wind  24hr
temp  day  1week
day  month

Trained Machine Learning Model

Forecast
Deploying Energy Forecasts

Dashboards for operators and traders

API for App Developers
Combining Forecasting with Optimization

“When should I operate my generators to maximize the return on my investment?”

Optimization Problem:

Minimize:

Cost of generating electricity

Constraints:
1) Meet forecasted demand
2) Operational constraints
3) Etc.
Challenge
Maximize margins in energy trading by predicting available supply and peak demand

Solution
Use MATLAB to build and optimize models that incorporate historical data, weather forecasts, and regulatory rules

Results
- Response time reduced by months
- Productivity doubled
- Program maintenance simplified

"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

Link to user story
Machine Learning + X

Fleet Analytics

- Equipment Expertise
  - Design Specs
  - Operating Modes
  - Operating Conditions

- Machine Learning
  - Statistical Analysis
  - Unsupervised Learning

Energy Forecasting

- Electrical Grid Expertise
  - Seasonality
  - Weather Effects
  - Generator Characteristics

- Machine Learning
  - Time Series Modeling
  - Regression
Machine Learning apps

- Try out many models
- Compare Results
- Get to a reasonable model without worrying about the details

Perform Hyperparameter Optimization in apps
AutoML

- Build many machine learning models
- Find a good model without becoming an expert

**fitcauto**

Model Selection

Hyper-parameter Optimization

Decision Tree? SVM? KNN? Ensemble? ...

Import Data → Preprocess Data → Extract Features → Train Model → Deploy & Integrate

Wavelet Scattering

Feature Selection

MATLAB EXPO
AutoML “in action”

% Step 1: apply Wavelet scattering to extract features
sf = waveletScattering('SignalLength',N, 'SamplingFrequency',50);
Wfeatures = featureMatrix(sf,thisSignal(1:N), 'Transform', 'Log');
% do this across signals <thisSignal> and accumulate <allFeatures> with labels

% Step 2: select top <featN> features according to feature ranking, e.g. MMR
[mrrrFeatures, scores] = fscmrrr(allFeatures, 'class');
trainFeatures = allFeatures(:, [mrrrFeatures(1:numPredictorsToUse);true]);

% Step 3: Select optimized model from 100 iterations of 1-step model selection
modelAuto = fitauto(trainFeatures, 'class', 'Learners', 'all',
'MaxNumberOfIterations', 100);
Examples of Successful Machine Learning Applications

- Fleet Data Analytics
- Energy Forecasting
- Manufacturing Analytics
What is Manufacturing Analytics?

**Definition:** Apply modeling (AI) to **process** and **sensor data** to maximize operational performance

**Key Use Cases:**

1. **Automate** the **monitoring** of manufacturing process
2. Ensure **product quality**
3. **Optimize yield** of complex production processes
Challenges in Applying AI to Manufacturing

Lots of Data – much in “Data Historians” (SCADA, LIMS, OSIsoft PI)

Reliable measurements or modeling
- Sensor failures
- Hidden variables

Use of many different tools
- Limited Predictive modeling
- Handle streaming data
- Customization
Uncover Hidden Variables with Process Modeling

Plant Production History (2010-2015)

Catalyst Aging

pretty big ➔ 6307200x37 tall table
Case Study: Anomaly Detection
Case Study: Anomaly Detection

1. Cluster with DBSCAN

2. One-class SVM

[Images of data clusters and anomaly detection models]
Deployment

Integration with Data Historians
- OPC Toolbox (Database tbx via ODBC or JDBC) connects with PI Server

Customize Analytics Delivery
- Accessing insights via GUI critical for plant staff and process engineers
- Build a custom dashboard with App Designer
Machine Learning + X

Fleet Analytics
- Equipment Expertise
  - Design Specs
  - Operating Modes
  - Operating Conditions
- Machine Learning
  - Statistical Analysis
  - Unsupervised Learning

Energy Forecasting
- Electrical Grid Expertise
  - Seasonality
  - Weather Effects
  - Generator Characteristics
- Machine Learning
  - Time Series Modeling
  - Regression

 Manufacturing Analytics
- Manufacturing Expertise
  - Process Equipment
  - Variables & Set Points
  - Parameter Impact
- Machine Learning
  - Anomaly Detection
  - Regression
  - Multivariate Statistics
Data Preprocessing

- Detrending
- Smoothing
- Resampling
- Filtering

Feature Engineering

- Bandwidth measurements
- Spectral statistics

Frequency domain

- Power Spectrum
- Find All Peaks

Time domain

- Find signal patterns
Predicting a patient’s fall risk with machine learning.
From Desktop to Production

Reasons for Updates:

- Found a better model
- New data became available
- Business needs change
- ...
Automatic C/C++ Code Generation

1. Prediction for most Classification and Regression models
2. Update deployed models without regenerating code
   - SVM, Decision Trees, Linear Models

1. Fixed-Point support
   - SVM, Decision Trees, Ensemble of Trees
   - Shallow Neural Network (through Simulink)

1. Integrate with Simulink models as MATLAB Function Block

Integrate MATLAB with Other Languages
Examples of Successful Machine Learning Applications

- Fleet Data Analytics
- Energy Forecasting
- Manufacturing Analytics

New Capabilities
- MATLAB apps
- AutoML
- Signal Processing with Machine Learning
- C/C++ Code Generation
Machine Learning

+ Fleet Data Analytics
  Industry Knowledge
  Manufacturing Analytics
  Medical Devices

X Signal Processing
  Energy Forecasting
  Application Knowledge
  Mining

C/C++ Code Generation

Apps
AutoML
Learn More

Get Started for Free

MATLAB Onramp
Get started quickly with the basics of MATLAB.
» Details and launch

Machine Learning Onramp
An interactive introduction to practical machine learning methods for classification problems.
» Details and launch

Deep Learning Onramp
Get started with deep learning techniques to perform image recognition.
» Details and launch

Training Courses

MATLAB Fundamentals (3 days)
MATLAB for Data Processing and Visualization (1 day)
Processing Big Data with MATLAB (1 day)
Statistical Methods in MATLAB (2 days)
Machine Learning with MATLAB (2 days)
Signal Preprocessing and Feature Extraction with MATLAB (1 day)
Deep Learning with MATLAB (2 days)
Accelerating and Parallelizing MATLAB Code (2 days)

Practical Data Science with MATLAB Specialization

- Exploratory Data Analysis
- Data Processing and Feature Engineering
- Predictive Modeling and Machine Learning
- Data Science Project