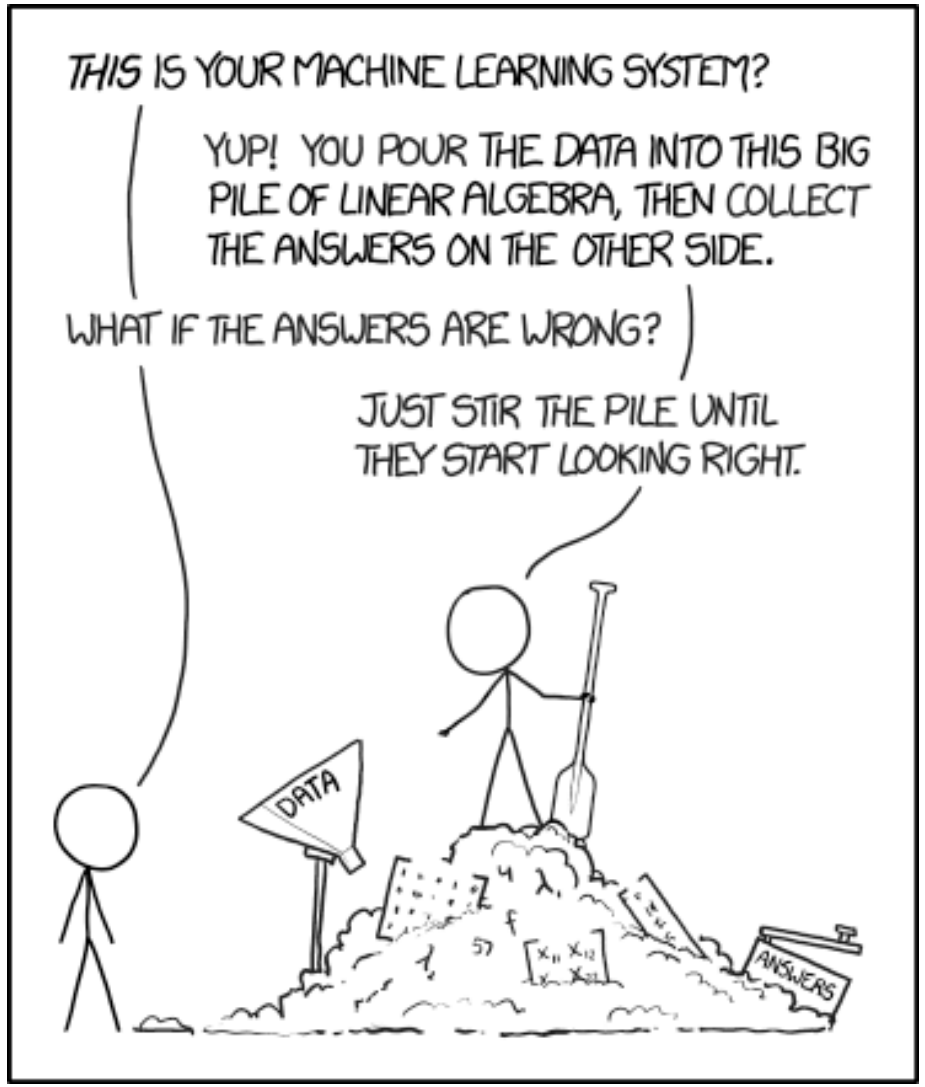
Machine Learning with MATLAB

A hands-on MATLAB workshop

Reece Teramoto
Application Engineer, MathWorks

What's Machine Learning About?



Source: https://xkcd.com/1838/

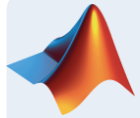
Internet of Shit Retweeted

Computer Facts
@computerfact

concerned parent: if all your friends jumped off a bridge would you follow them?
machine learning algorithm: yes.

2:20 PM · Mar 15, 2018

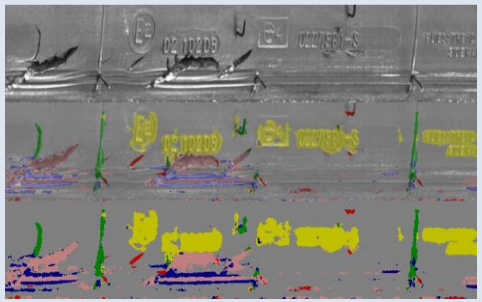
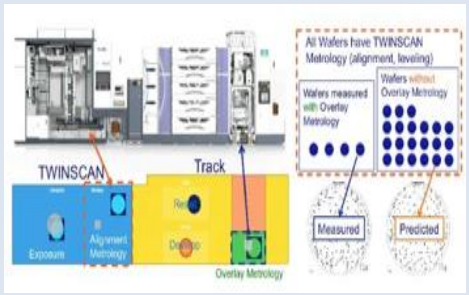
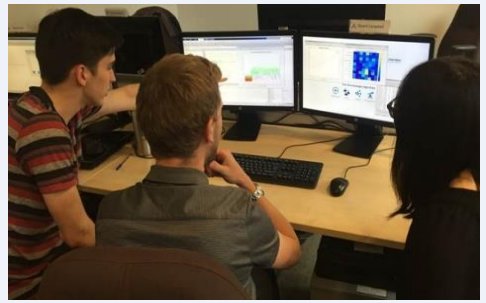
Agenda



Machine learning introduction

- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

Machine Learning is Everywhere

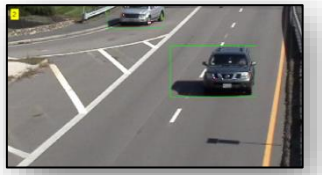
Automobile	Industrial Automation	CES & Aero Defense	Energy & Finance
			
Tire Wear BRIDGESTONE	Overlay metrology improvement ASML	Telecom customer churn prediction Cognizant	Forecasting & Risk Analysis HORIZON WIND POWER
			
Detect Oversteer BMW	Building energy use optimization BuildingIQ	Engine Health (Pred Maintenance) SAFRAN	Portfolio Allocation Aberdeen Asset management

What is Machine Learning?

Ability to learn from data without being explicitly programmed

Solution is too complex for hand written rules or equations

Speech Recognition


Object Recognition


Engine Health Monitoring

learn complex non-linear relationships

Solution needs to adapt with changing data

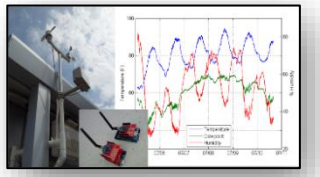

Weather Forecasting


Energy Load Forecasting

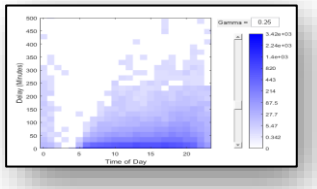

Stock Market Prediction

update as more data becomes available

Solution needs to scale

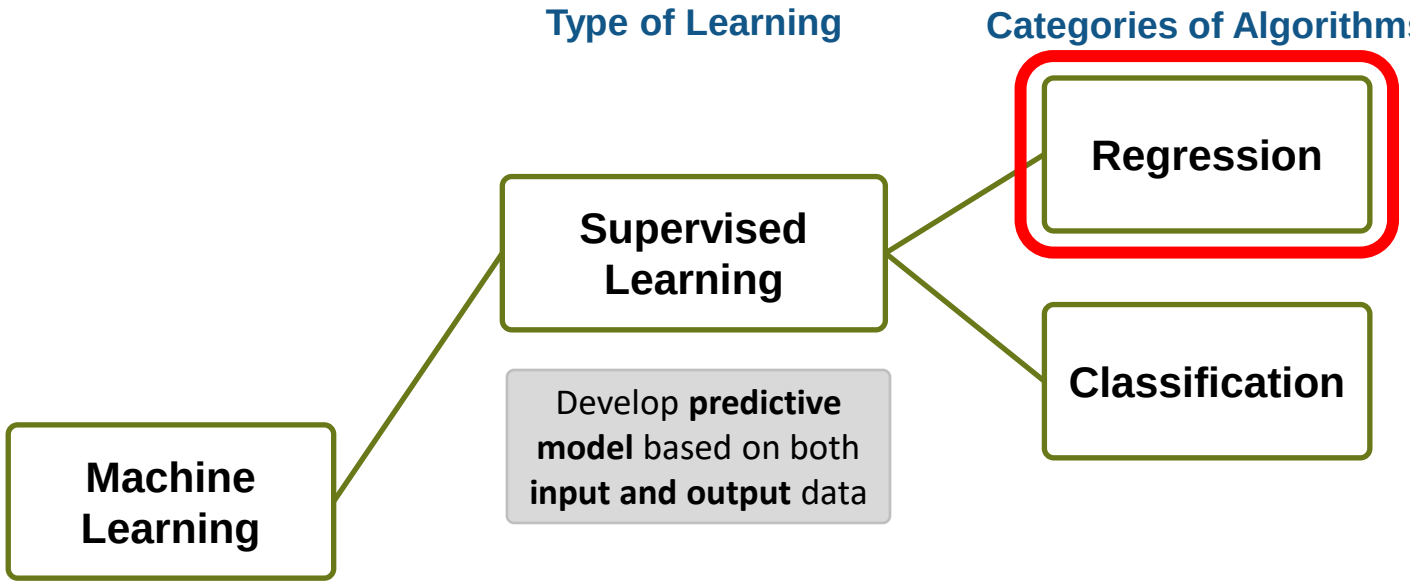

IoT Analytics

Taxi Availability


Airline Flight Delays

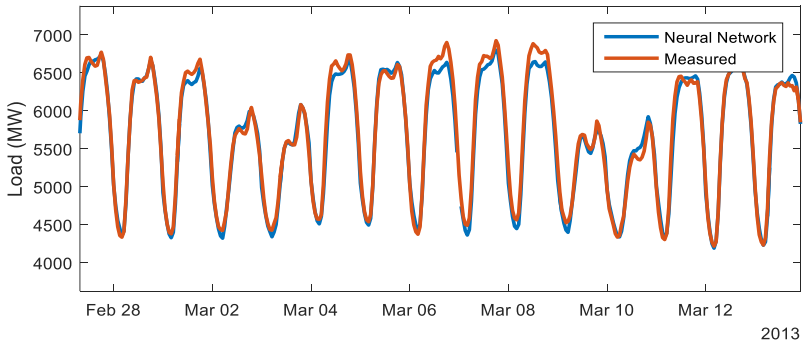
learn efficiently from very large data sets

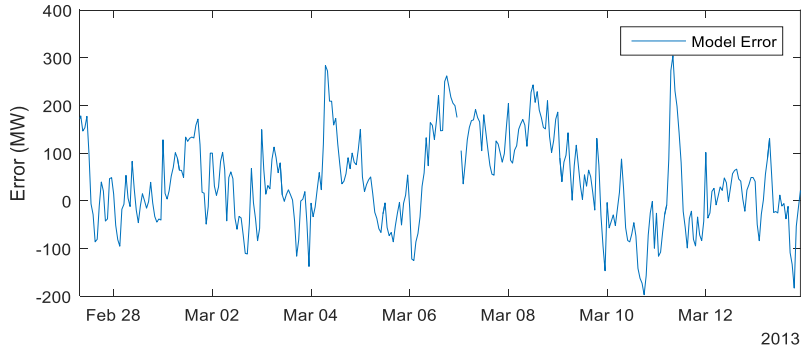
Types of Machine Learning



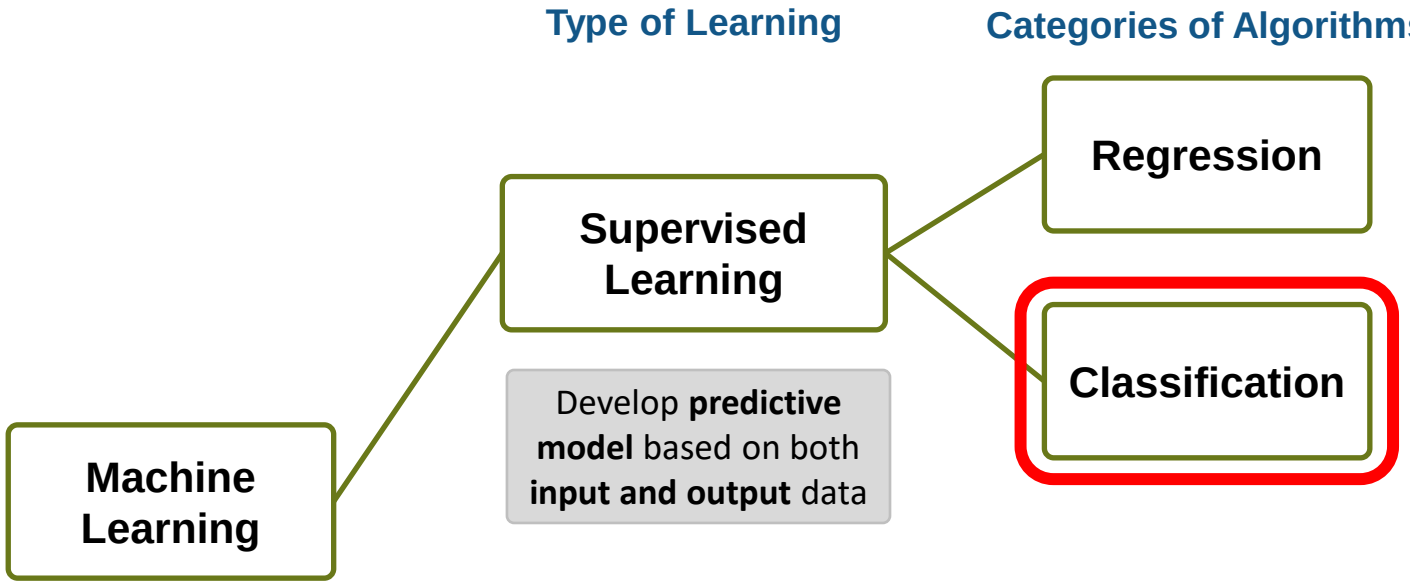
Objective:

Easy and accurate computation of day-ahead system load forecast





Types of Machine Learning



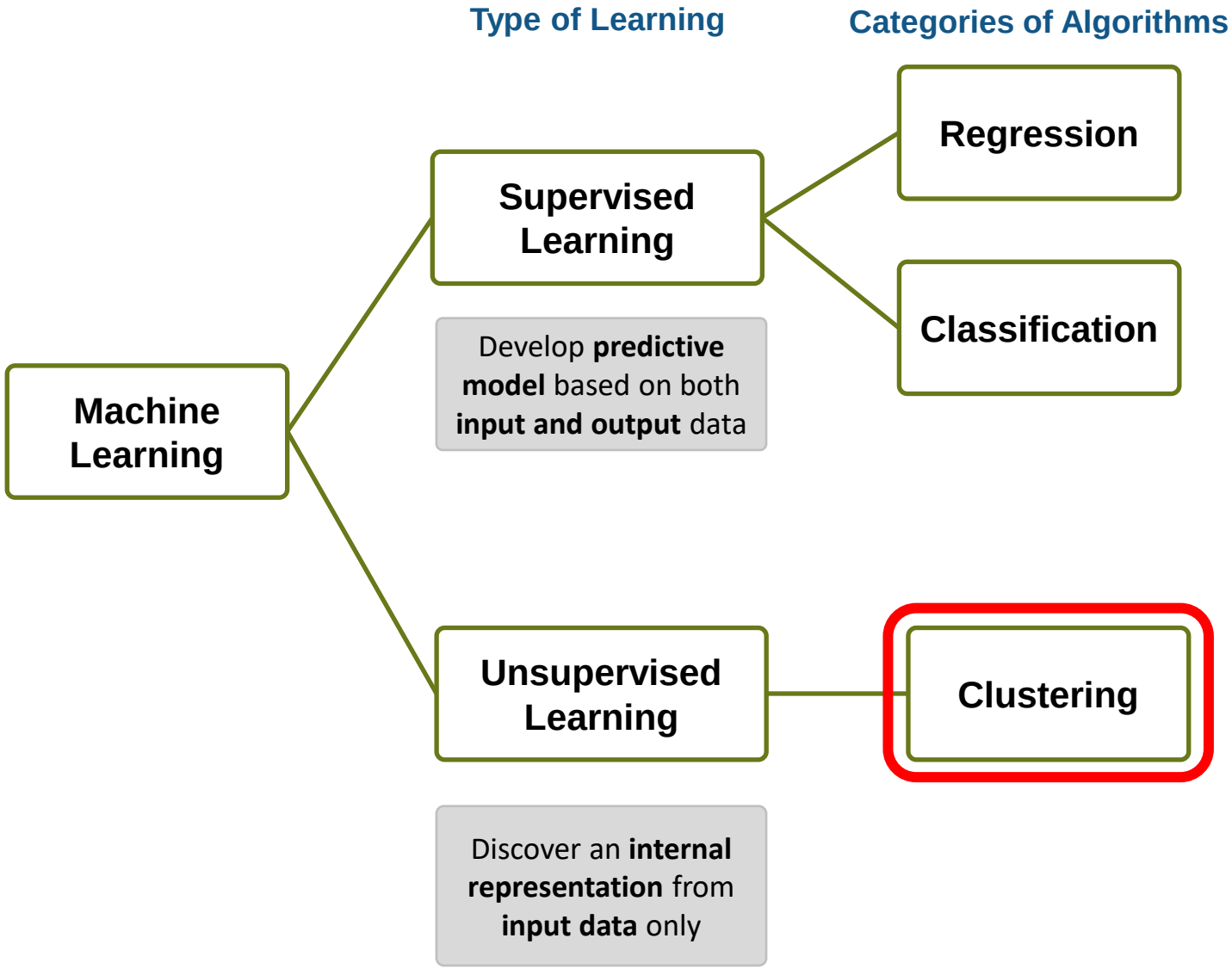
Objective:

Train a classifier to classify human activity from sensor data

Data:

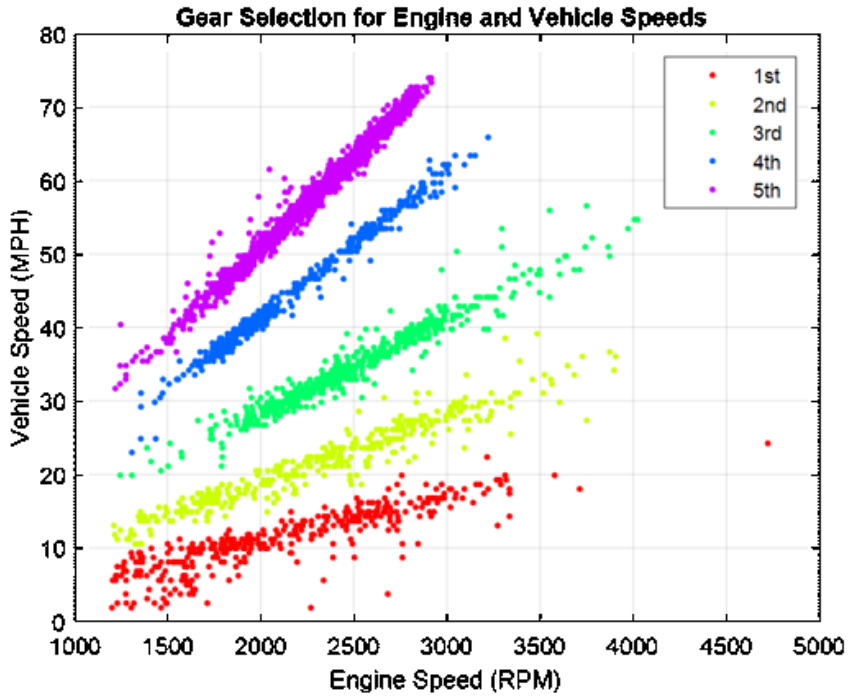
Inputs	3-axial Accelerometer 3-axial Gyroscope 
Outputs	

Types of Machine Learning



Objective:

Given data for engine speed and vehicle speed, identify clusters



Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

Exercise 1: Predicting Fuel Economy

Regression

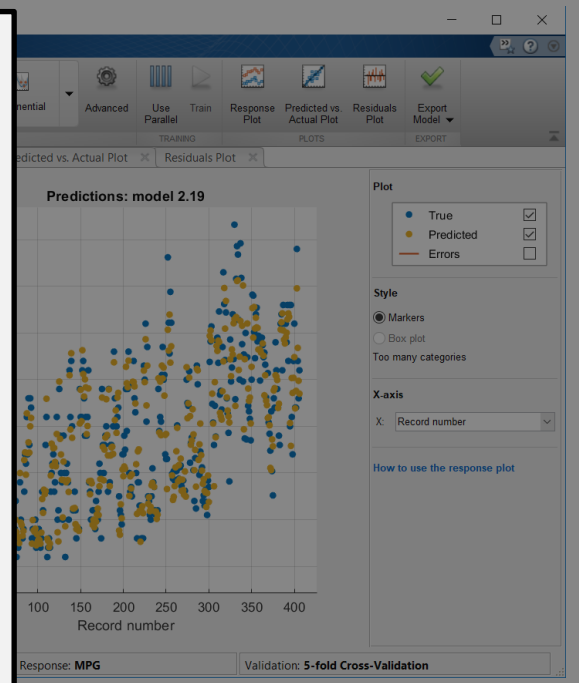
Goal: Study drivers

Approach:

- Load data in MA
- Use the Regres
  multiple regress
- Create a model
  new car given c
  weight, etc

Let's try it out!

Exercise: ***Predicting Fuel Economy***
in folder 01-RegressionModels



“essentially, all models are wrong,
but some are useful”
– George Box

Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

Machine Learning Workflow

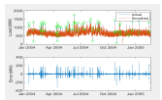
Access and Explore Data

- Files
- Databases
- Sensors

- Data Diversity
- Data clean up
- Working with big data

Preprocess Data

- Noisy Data
- Data Transformation
- Feature Extraction



- Data specific processing
- Feature Extraction
- Feature Selection

Develop Predictive Models

- Model Creation
- Parameter Optimization
- Model Validation

- Many different models
- Model tuning
- Computationally intensive

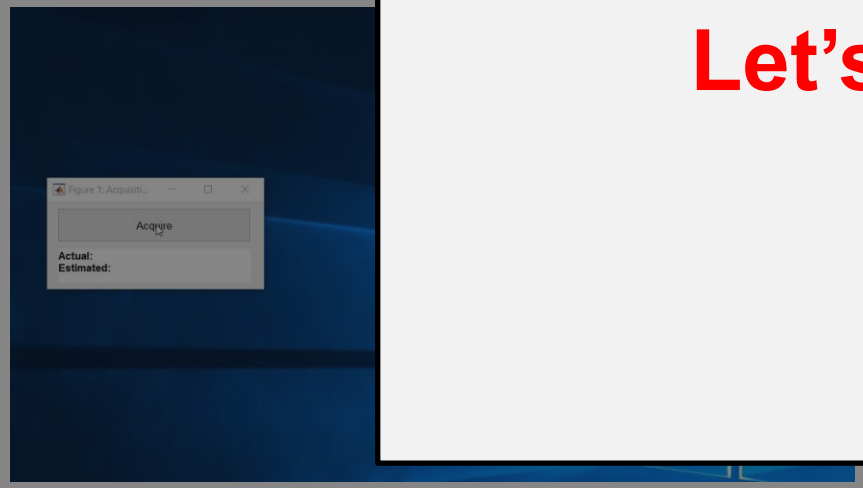
Integrate Analytics with Systems

- Desktop Apps
- Enterprise Systems
- Embedded Devices

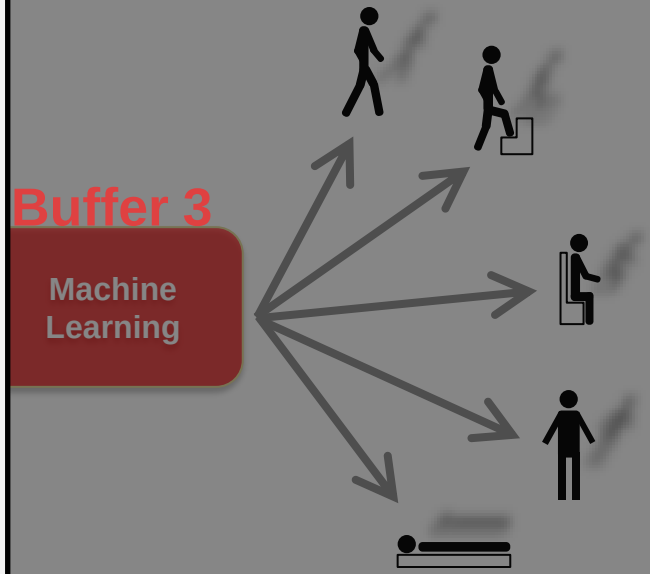
- Different end users
- Different target platforms
- Different Interfaces

Exercise 2: Human Activity Learning using Smartphones

Goal: Create a mo
activity from sens

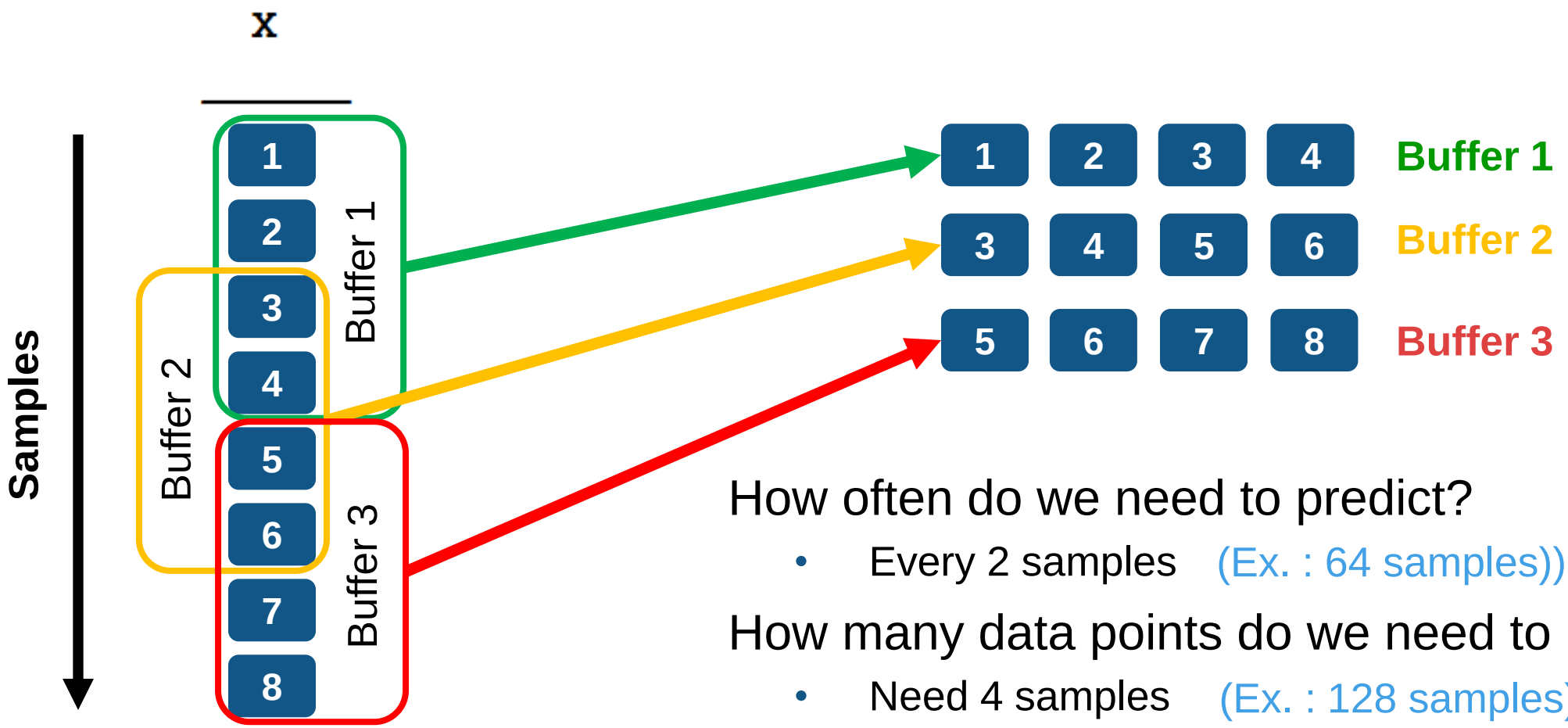


Let's observe!



Dataset courtesy of:
Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. *Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine.* International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012 http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Signal Buffering - Simple Example



How often do we need to predict?

- Every 2 samples (Ex. : 64 samples))

How many data points do we need to predict?

- Need 4 samples (Ex. : 128 samples))
- Create overlapping buffers of 4 points (Ex. : 64 samples))

Compute features (e.g. mean) on each buffer

Exercise 2: Human Activity Learning using Smartphones

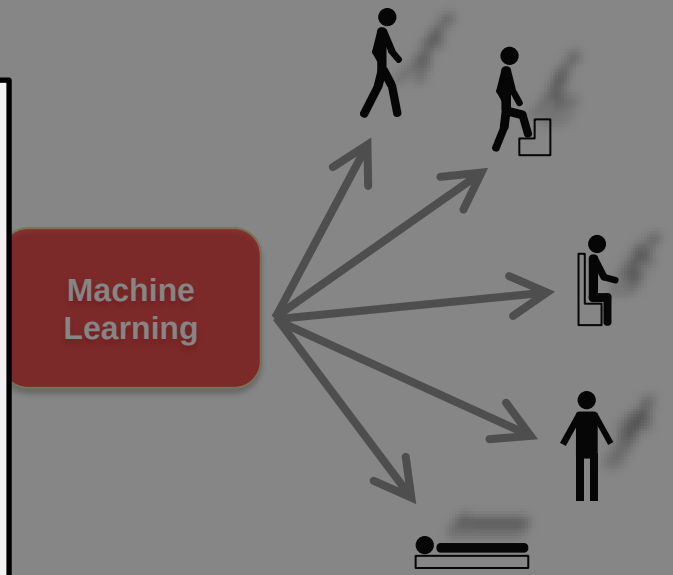
Goal: Create a m
activity from sens

Approach:

- Load buffered
- Extract statisti
- Train a machin (interactively)
- Train a machin (programmatically)
- Optimize model using hyperparameter tuning

Let's try it out!

Exercise:
humanActivityClassification.mlx
in folder 02-ClassificationModels

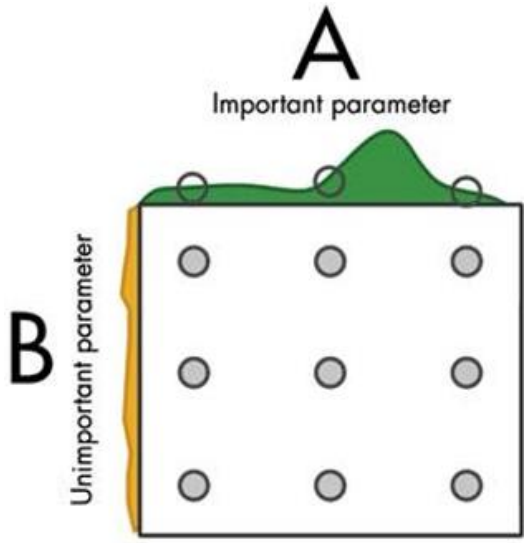


ects	30
	25
ta	3
testData	2

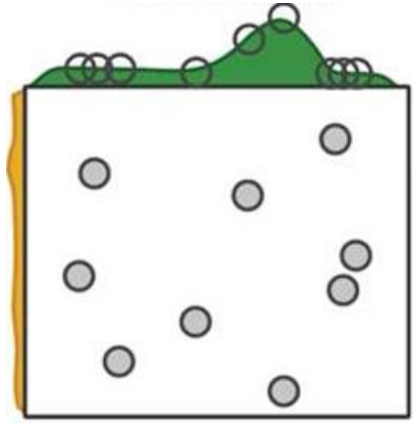
Dataset courtesy of:
Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. *Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine.* International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012 http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Hyperparameter Tuning

Standard:
Grid Search

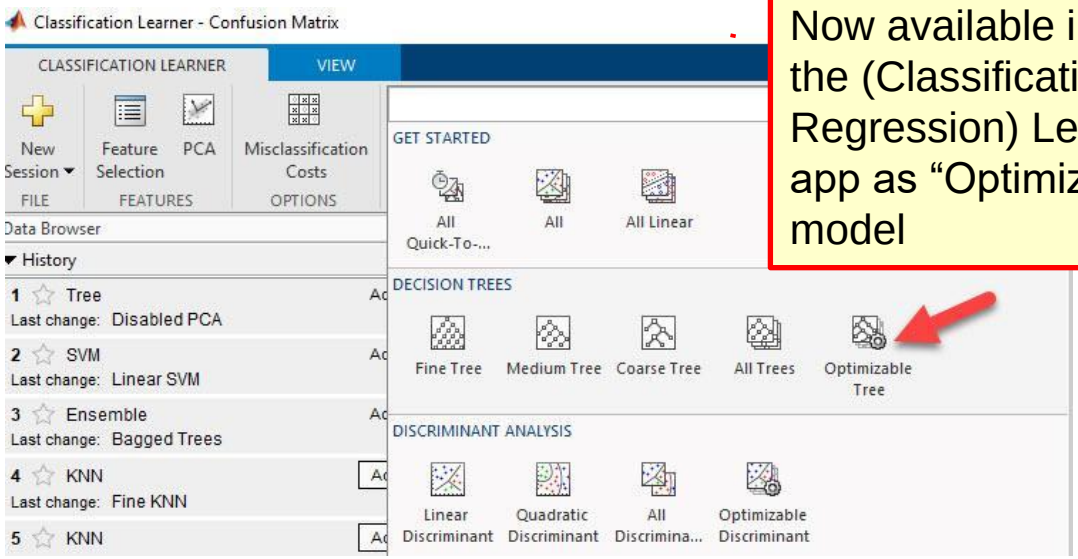


Better:
Random Search



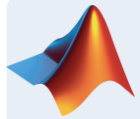
Best: Bayesian Optimization

- Bayesian model indicates impact of change
- Model picks “good” point to try next
- Much more efficient!
- Scale to multi-cores (using PCT) for larger datasets



Now available inside the (Classification/ Regression) Learner app as “Optimizable” model

Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
-  Feature extraction and feature selection
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

Feature Engineering

Using domain knowledge to create features for machine learning algorithms

"… is the art part of data science"

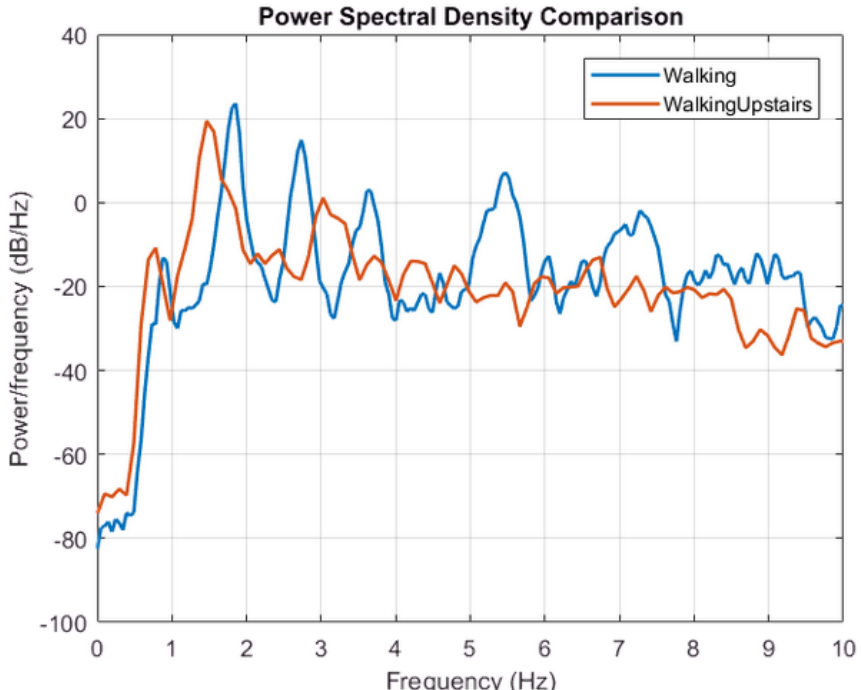
Sergey Yurgenson
(Kaggle Master)

Feature transformation: high dimensionality

Feature selection: subset of relevant features

Possible feature engineering ideas:

- Additional statistics – PCA, NCA etc.
- Signal Processing Techniques – power spectral density, wavelets etc.
- Image Processing Techniques – bag of words, pixel intensity etc.
- Get creative!



Exercise 3 – Feature Engineering for human activity

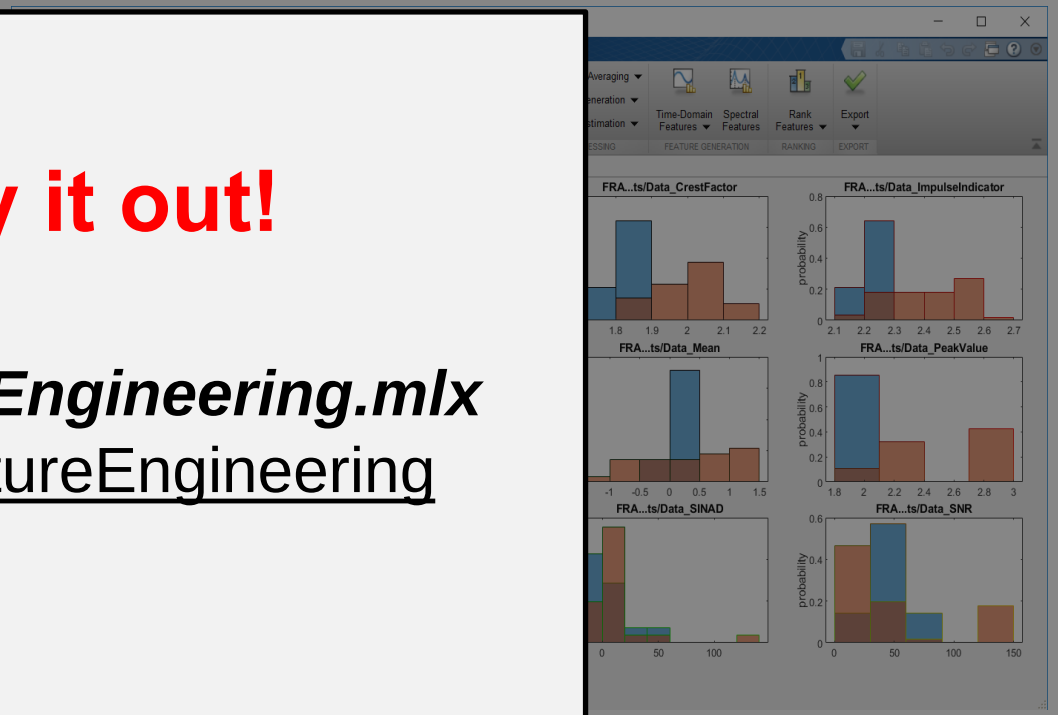
Goal: Explore different techniques for feature engineering

Approach:

- Use signal processi extract time domain
- Use feature selectio reduce the set of fea relevant
- Browse examples i documentation for different applications

Let's try it out!

Exercise: ***featureEngineering.mlx***
in folder 03-FeatureEngineering



Automated Feature Generation with Wavelet Scattering

Wavelet Scattering Framework [Bruna and Mallat 2013]

- Automatic Feature Extraction
- Great starting point if you don't have a lot of data
- Reduces data dimensionality and provides compact features

Additional Resources:

- Wavelet scattering for ECG [doc example]
- Applying Deep Learning to Signals [3 min video]
- Blog about Wavelet scattering on towardsdatascience.com

Diagnostic Feature Designer App

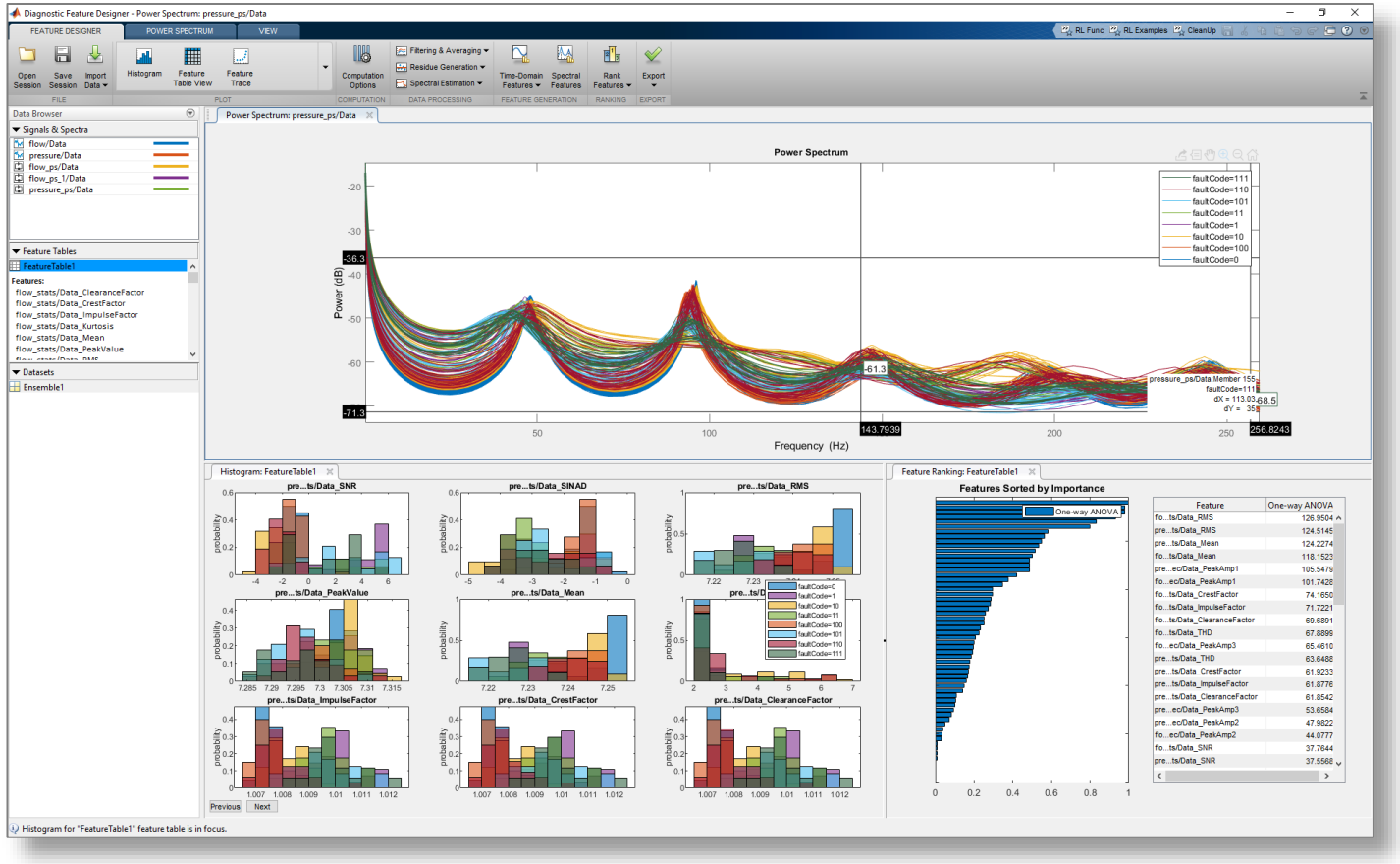
Predictive Maintenance Toolbox R2018b and R2019a

Extract, visualize, and rank features from sensor data

Use both statistical and dynamic modeling methods

Work with out-of-memory data

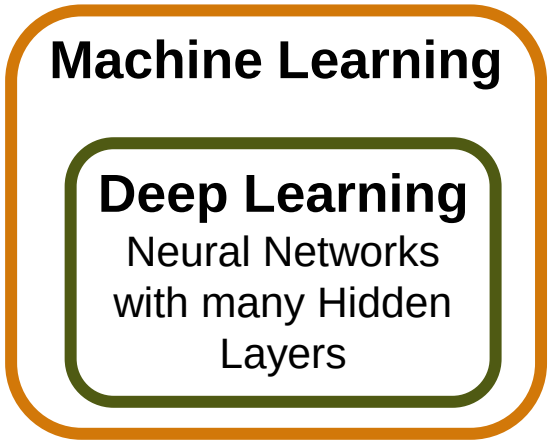
Explore and discover techniques without writing MATLAB code



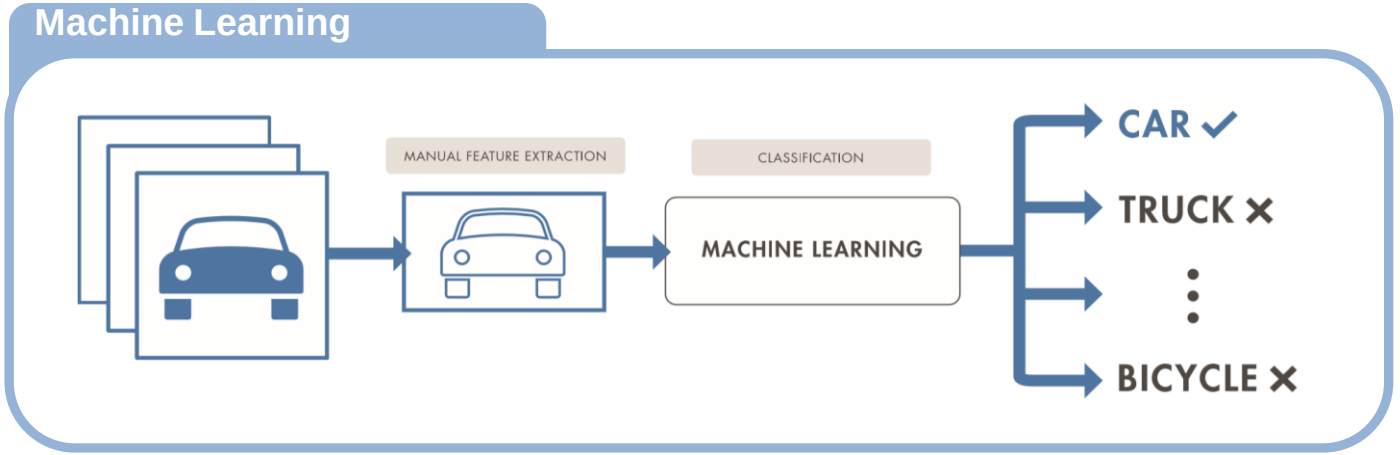
What have we discussed so far…

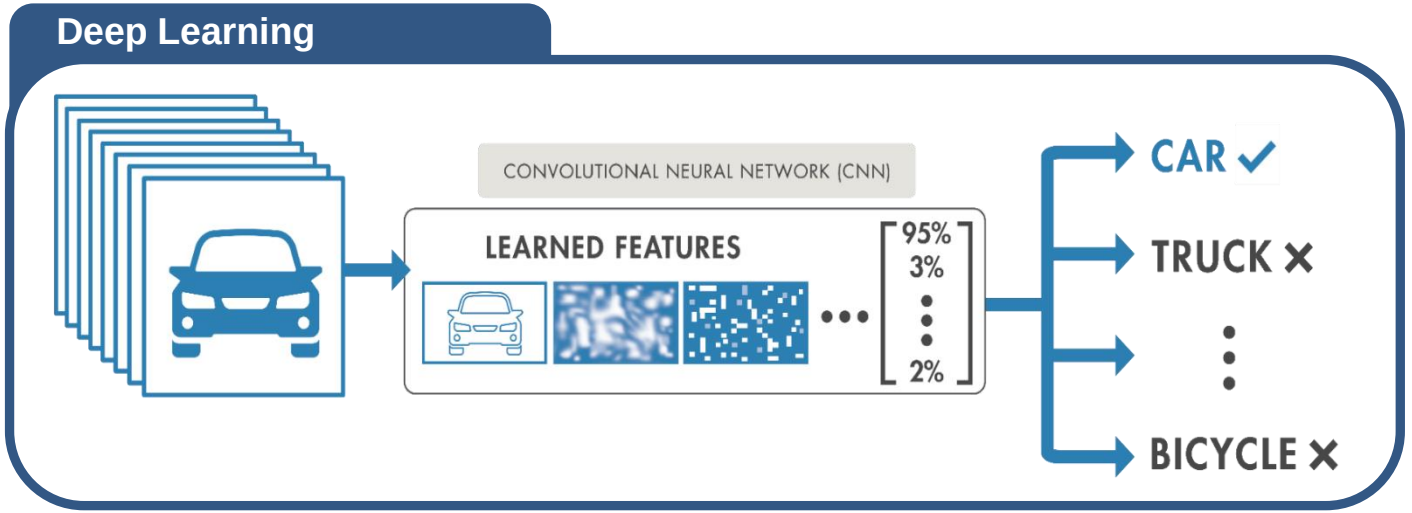
- Fundamentals of machine learning
- Various machine learning models for classification and regression
- Optimizing models leveraging hyperparameter tuning and feature selection
- Advanced signal processing and feature extraction techniques

Beyond traditional Machine Learning: Deep Learning

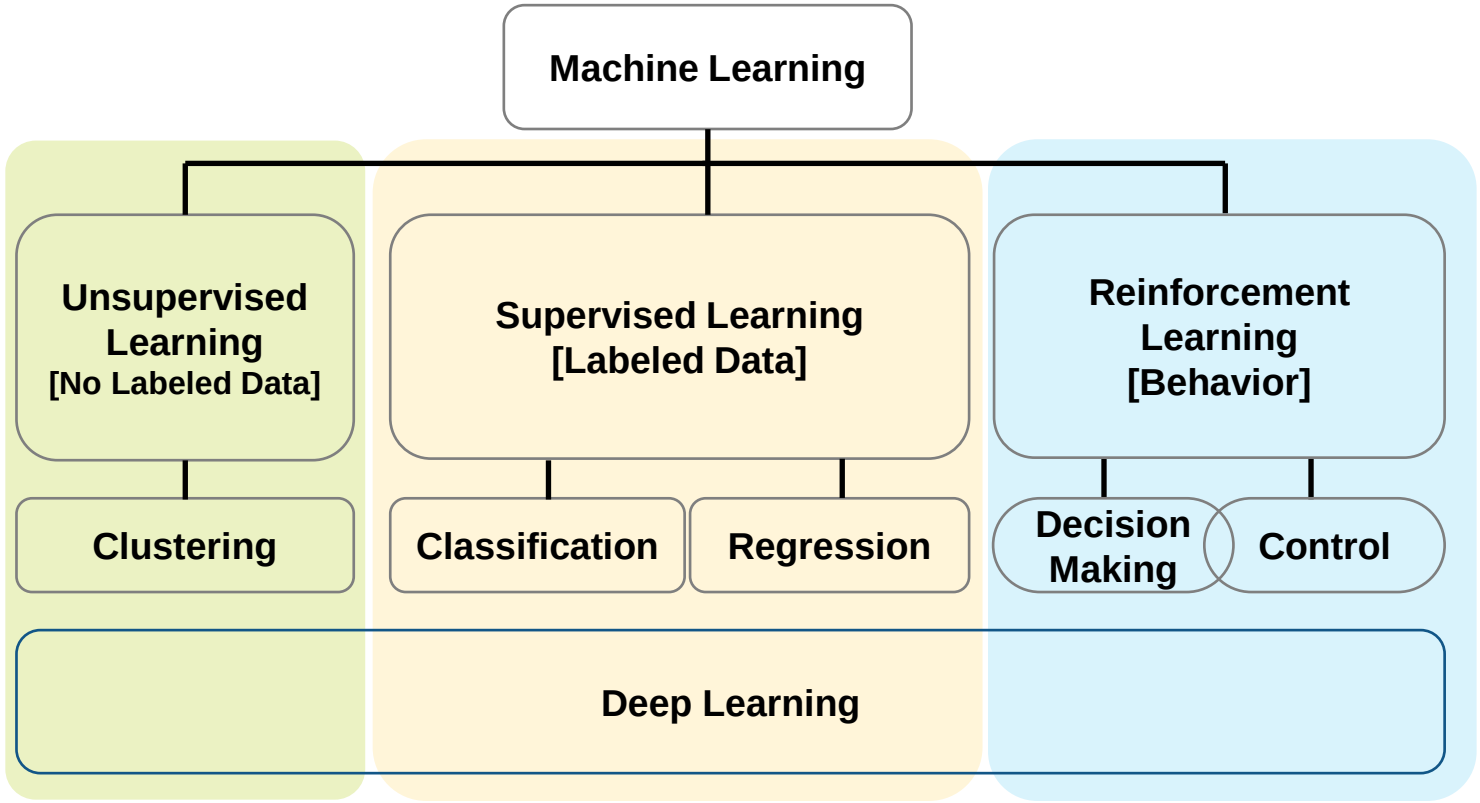


- Learns directly from data
- More Data = better model
- Computationally Intensive
- **Not interpretable**





Beyond Machine Learning: Reinforcement Learning



Reinforcement learning:

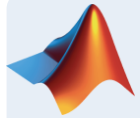
Learning through trial & error [*interaction data*]

Complex problems typically need deep learning

It's about learning a **behavior** or accomplishing a **task**

Resources, sessions, workshops and training available

Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
-  Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

Clustering

What is clustering?

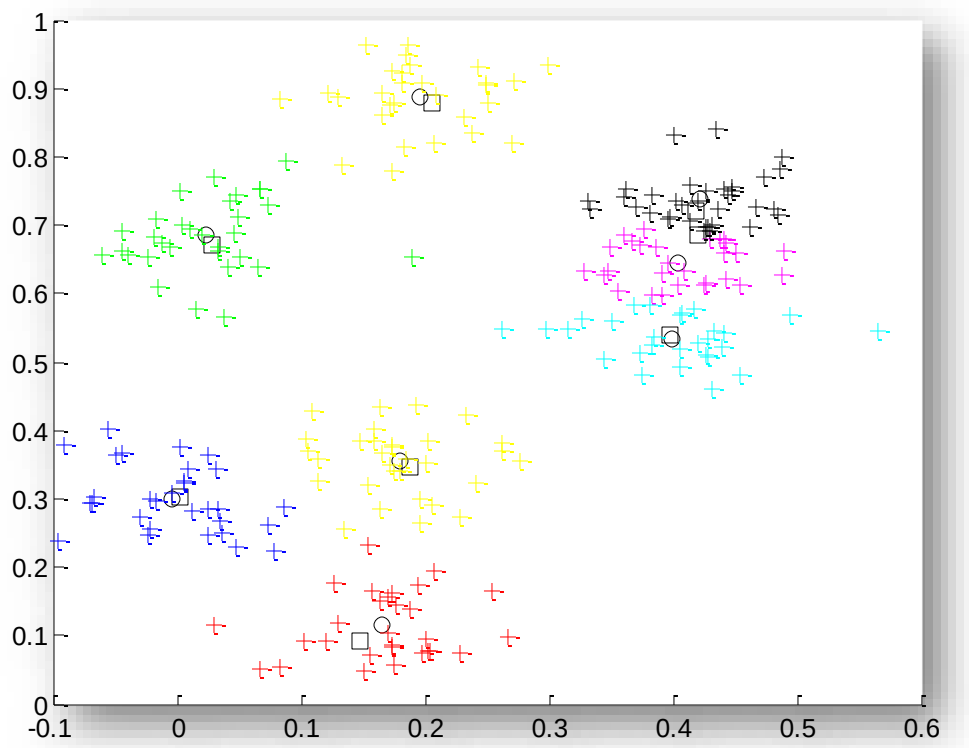
Segment data into groups,
based on data similarity

Why use clustering?

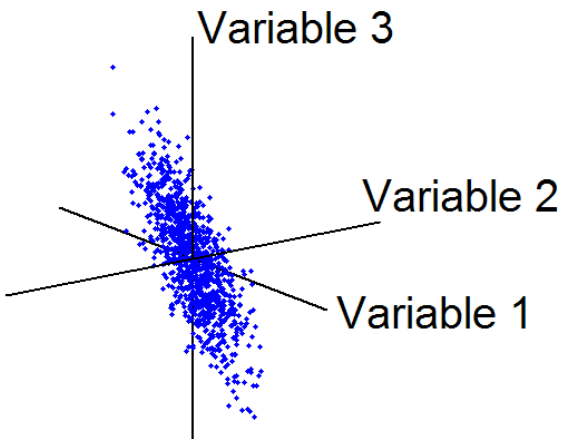
- Identify outliers
- Discover patterns of interest

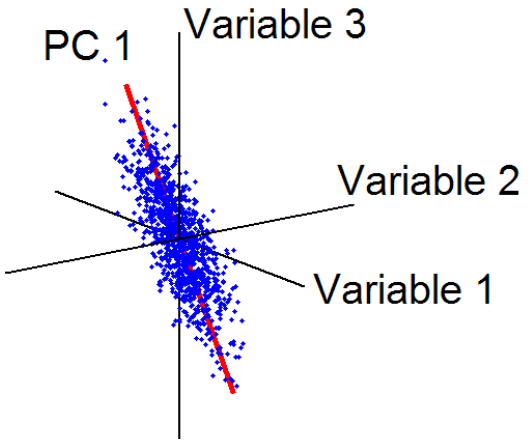
How is clustering done?

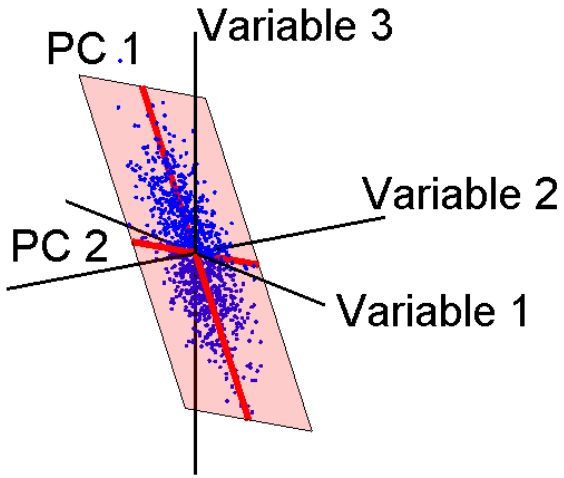
- Can be achieved by various algorithms
- It is an iterative process *(involving trial and error)*

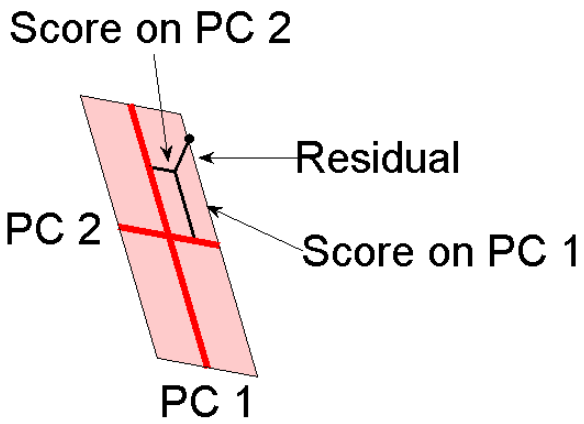


Principal Components Analysis (PCA)









Exercise 4: Clustering Human Activity

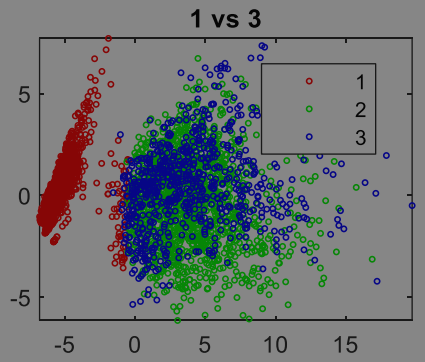
Goal: Find natural
large number of fea
activity

Approach:

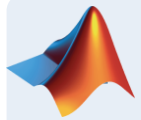
- Reduce dimens
  structure of data
- Evaluate differe
  identify groups

Let's try it out!

Exercise: ***clusteringHumanActivity.mlx***
in folder 04-UnsupervisedLearning



Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
- Unsupervised learning (optional)
-  Working with big data (optional)
- Deploying Machine Learning Algorithms

Big Data in MATLAB: Tall Arrays

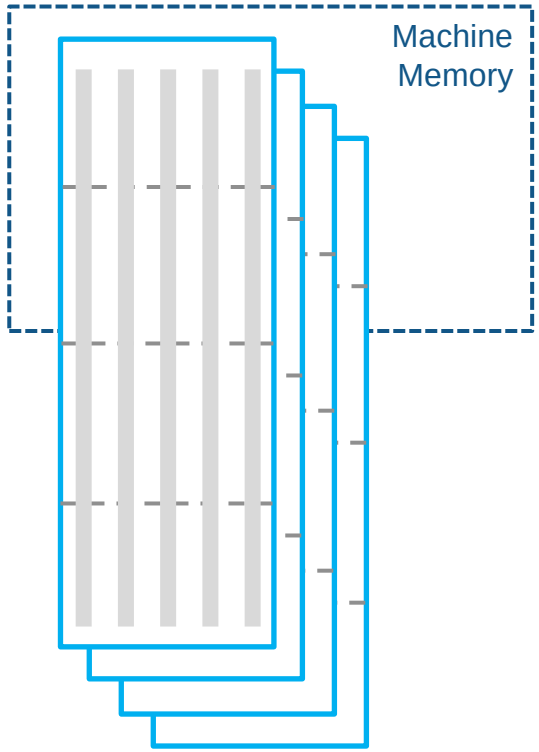
Extends the “array” data type to out-of-memory

- Use like a regular (in-memory) array in supported functions
- (With some setup) Scales processing to clusters with Spark

Applicable when:

- Data is **columnar** – with **many** rows
- Overall data size is **too big to fit into memory**
- Operations are mathematical/statistical in nature

Hundreds of functions supported in MATLAB and Statistics and Machine Learning Toolbox



Tall Data

Big Data Without Big Changes

One file

Access Data

```
measured = readtable('PumpData.csv');
measured = table2timetable(measured);
```

Preprocess Data

Select data of interest

```
measured = measured(timerange(seconds(1),seconds(2)),:)
```

Work with missing data

```
measured = fillmissing(measured,'linear');
```

Calculate statistics

```
m = mean(measured.Speed);
s = std(measured.Speed);
```

One hundred files

Access Data

```
measured = datastore('PumpData*.csv');
measured = tall(measured);
measured = table2timetable(measured);
```

Preprocess Data

Select data of interest

```
measured = measured(timerange(seconds(1),seconds(2)),:)
```

Work with missing data

```
measured = fillmissing(measured,'linear');
```

Calculate statistics

```
m = mean(measured.Speed);
s = std(measured.Speed);
```

```
[m,s] = gather(m,s);
```

Exercise 5: Predicting Tips for Cab Drivers

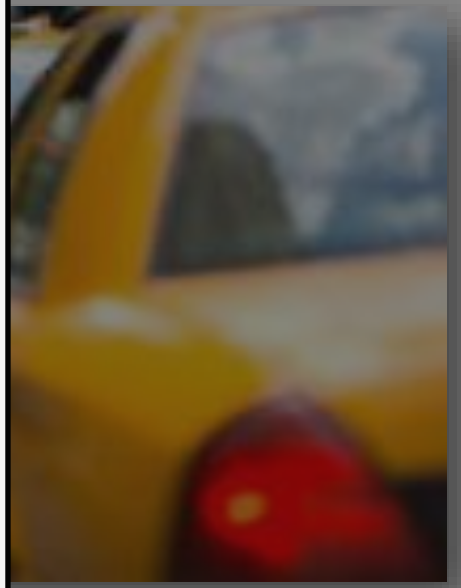
Goal: Create a model on a (simulated) large dataset

Approach:

- Access data spread
- Preprocess and Exp
- Train and validate a model

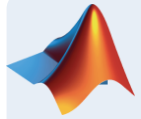
Let's try it out!

Exercise: ***predictDriverTip.mlx***
in folder 05-BigData



Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
- Unsupervised learning (optional)
- Working with big data (optional)



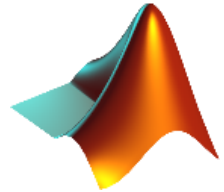
Deploying Machine Learning Algorithms

Deploying MATLAB Algorithms

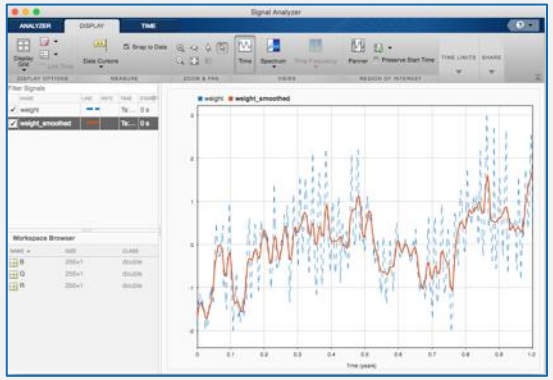
- Royalty-free deployment
- Point-and-click workflow
- desktop and server apps

Let's observe!

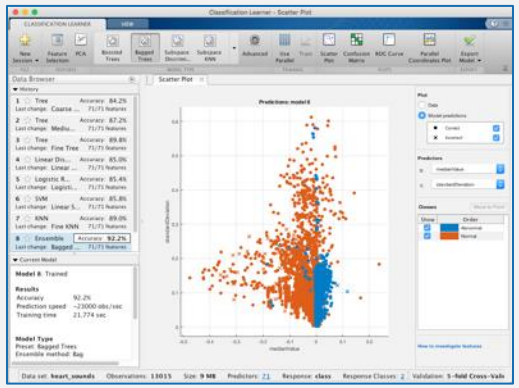
Summary: Complete Machine Learning Workflow



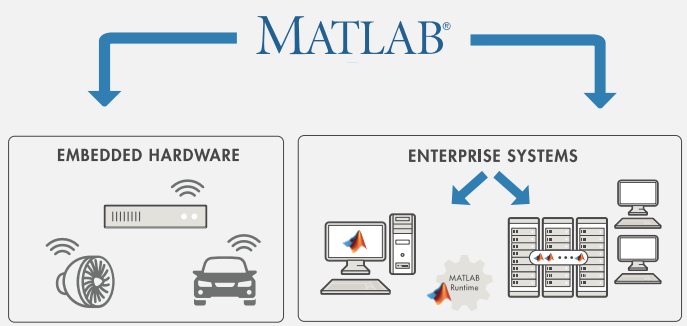
Access and explore data | Preprocessing | Feature Engineering | Model Training | Model Tuning | Integrate Analytics



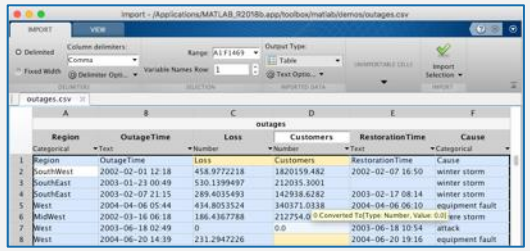
Datatypes and tools for missing data, outliers, time-alignment, etc.



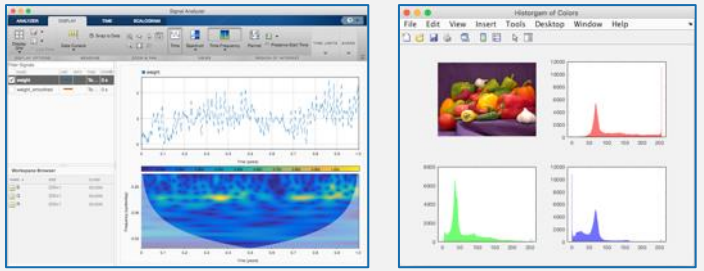
Machine Learning apps



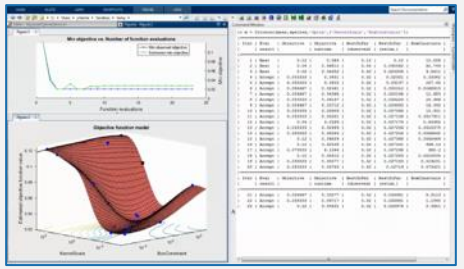
C/C++ Code Generation and Enterprise IT Integration



Text files, spreadsheets, databases, binary files, data feeds, web, cloud storage



Domain-specific techniques for Signals, Images, Video, Audio, and Text



Automated Parameter Tuning

Where to go from here?

- Finish what you didn’t get to - Continue exploring:
  - Keep using **MATLAB Online**: https://matlab.mathworks.com (but no GPU!)
  - Your existing desktop MATLAB license (but need to copy content)
- Where to find content? **MATLAB Drive** drive.matlab.com (250MB)
- Apply this to YOUR work

Don’t Forget:

- Fill out **feedback** (on back of your setup instructions)

Resources

Machine Learning Intro Tech talks

Machine Learning with MATLAB:

- Overview
- Cheat sheet
- Introductory eBook
- Mastering Machine Learning eBook
- Try the Classification Learner App in a browser

Deep learning onramp course

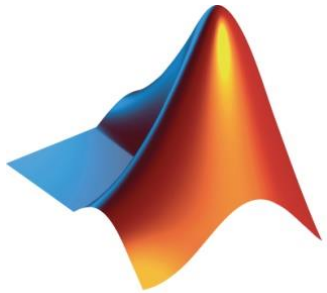
MathWorks® can help you do Machine Learning

Free resources:

- Guided evaluations with a MathWorks machine learning engineer
- Proof-of-concept projects
- Seminars and technical deep dives
- Hands-on Workshop for Deep or Reinforcement Learning workshop

More options:

- Technical support
- Advanced customer support
- Installation, enterprise, and cloud deployment
- Consulting services
- Machine/Deep Learning Paid Training



MathWorks®

Accelerating the pace of engineering and science