Hands-On Virtual Lab: Machine Learning

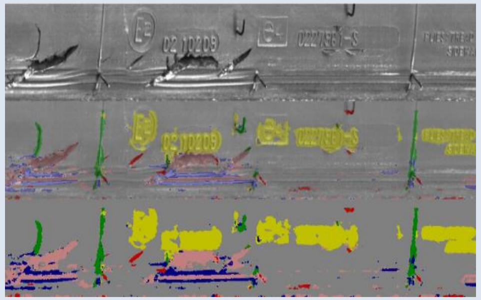
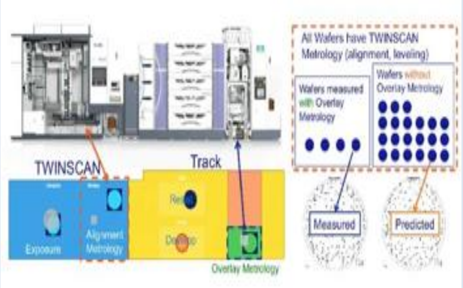
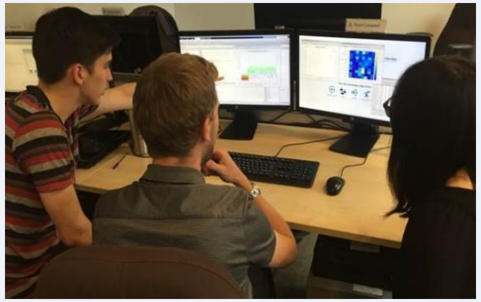
Reece Teramoto

Application Engineer, MathWorks

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

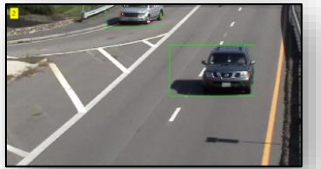
Automobile	Industrial Automation	CES & Aero Defense	Energy & Finance
 Tire Wear BRIDGESTONE	 Overlay metrology improvement ASML	 Telecom customer churn prediction Cognizant	 Forecasting & Risk Analysis HORIZON WIND POWER
 Detect Oversteer BMW	 Building energy use optimization BuildingIQ	 Engine Health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

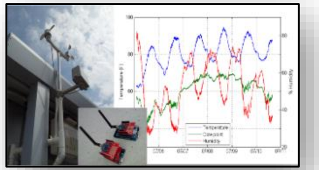

Weather Forecasting


Energy Load Forecasting

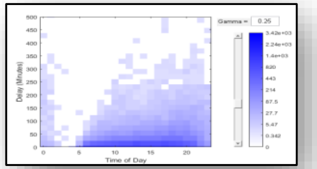

Stock Market Prediction

update as more data becomes available

Solution needs to scale

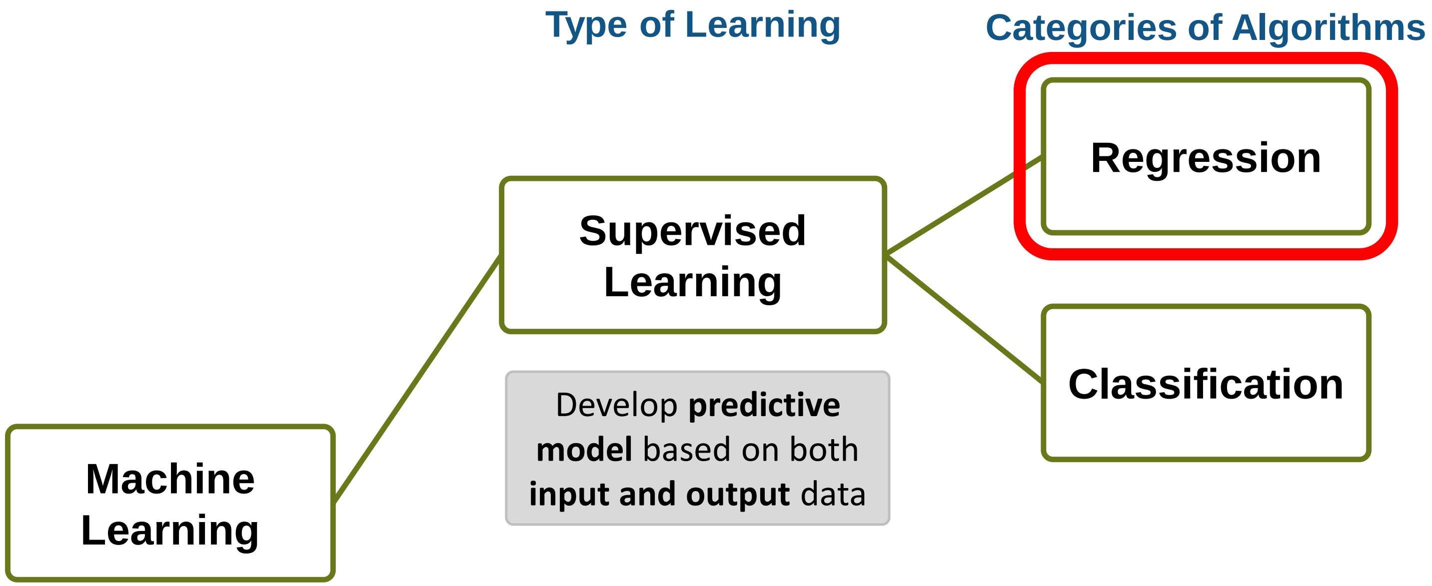

IoT Analytics

Taxi Availability


Airline Flight Delays

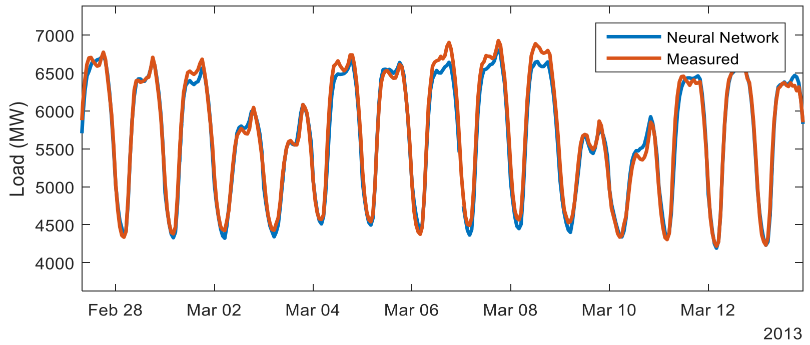
learn efficiently from very large data sets

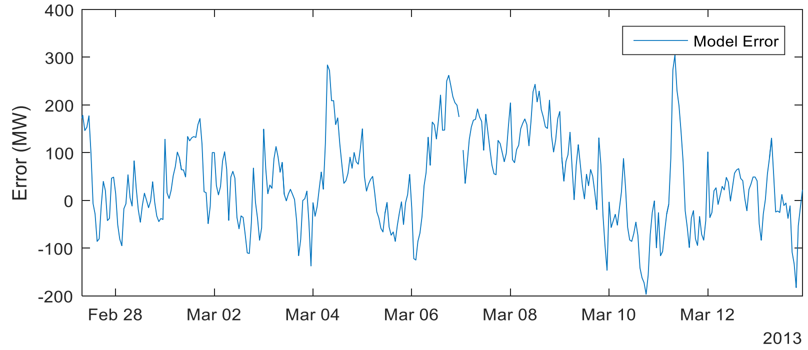
Types of Machine Learning



Objective:

Easy and accurate computation of day-ahead system load forecast





Types of Machine Learning

Type of Learning | **Categories of Algorithms**

Machine Learning

Supervised Learning

Develop **predictive model** based on both **input and output** data

Regression

Classification

Objective:

Train a classifier to classify human activity from sensor data

Data:

Inputs	3-axial Accelerometer 3-axial Gyroscope 
Outputs	

Types of Machine Learning

Type of Learning | **Categories of Algorithms**

Machine Learning

- **Supervised Learning**
  - **Regression**
  - **Classification**

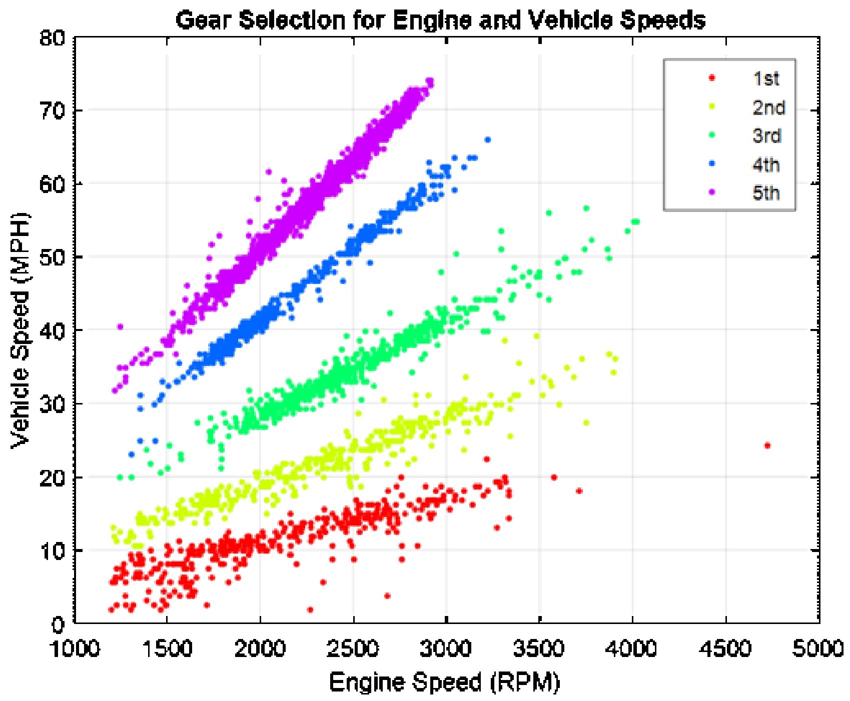
  Develop **predictive model** based on both **input and output** data

- **Unsupervised Learning**
  - **Clustering**

  Discover an **internal representation** from **input data** only

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 of

– Build initial models

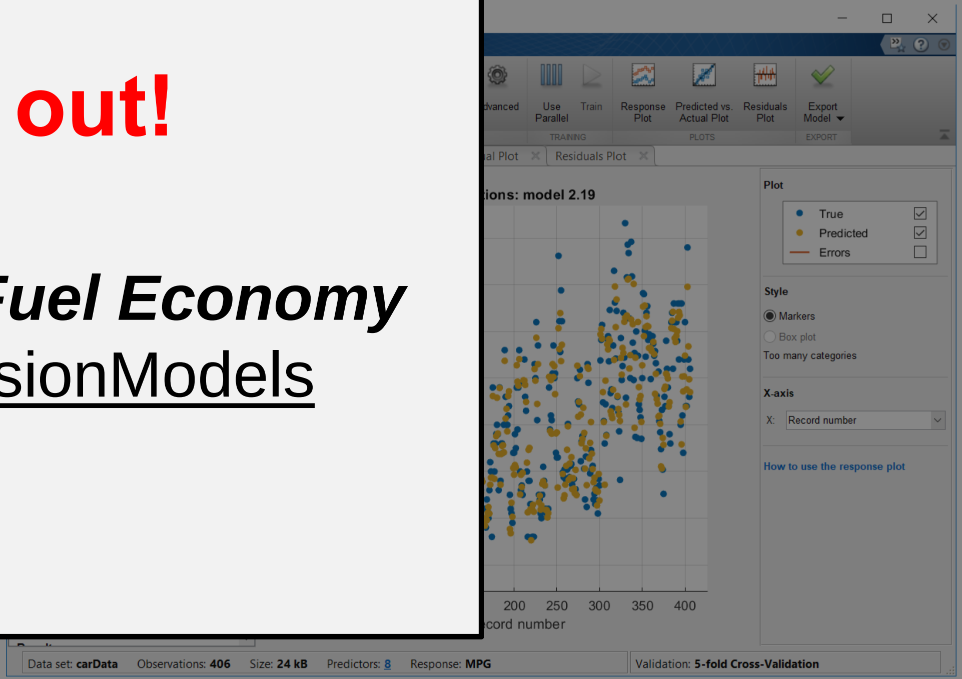
– Don't need to be a

Approach:

– Load data in MATL

– Use the Regression
multiple regression

– Create a model which can predict mpg for a new car given characteristics like horsepower, 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

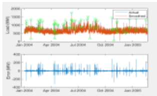
Databases

Sensors

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

Preprocess Data

Noisy Data



Data Transformation

Feature Extraction

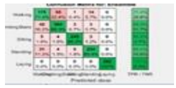
- Data specific processing
- Feature Extraction
- Feature Selection

Develop Predictive Models

Model Creation

Parameter Optimization

Model Validation



- Many different models
- Model tuning
- Computationally intensive

Integrate Analytics with Systems

Desktop Apps

Enterprise Systems

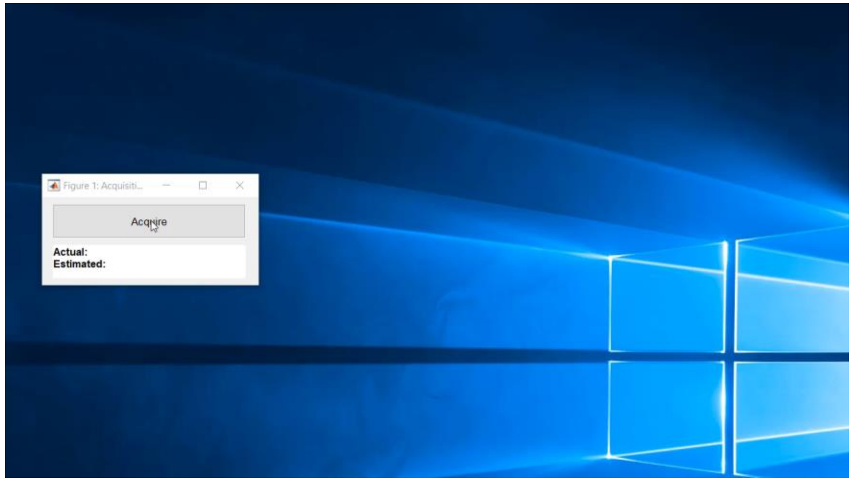
MATLAB Excel .NET .exe C/C++ Java .dll

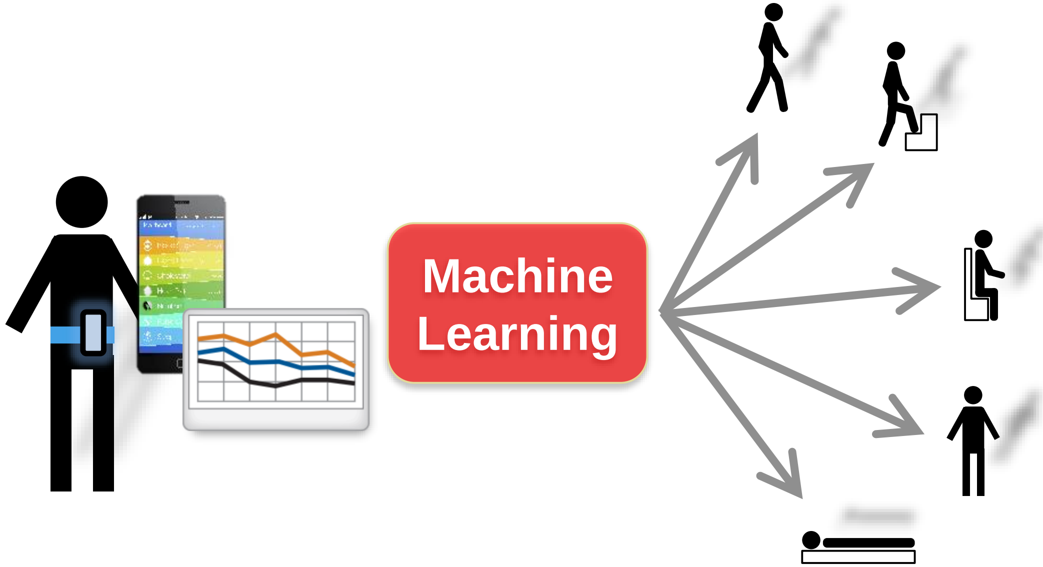
Embedded Devices

- Different end users
- Different target platforms
- Different Interfaces

Human Activity Learning using Smartphones

Example task: Create a model to classify human activity from sensor data

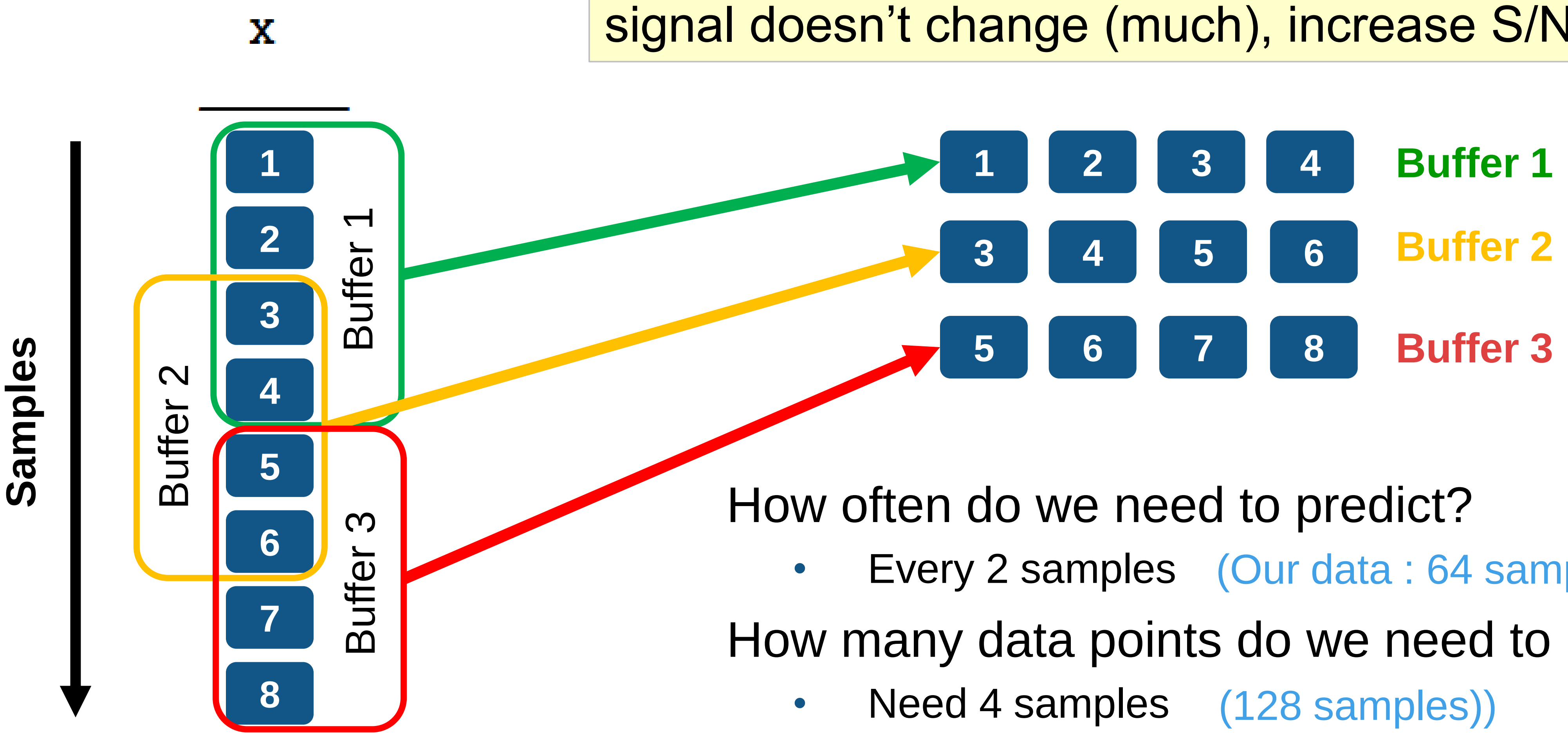




Dataset courtesy of:
Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. *Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine.* International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec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

Signal Buffering

Why? – Calculate features on "chunks" during which signal doesn't change (much), increase S/N (in feature)!



How often do we need to predict?

- Every 2 samples (Our data : 64 samples))

How many data points do we need to predict?

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

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

Exercise 2: Human Activity Learning using Smartphones

Goal: create initial mod

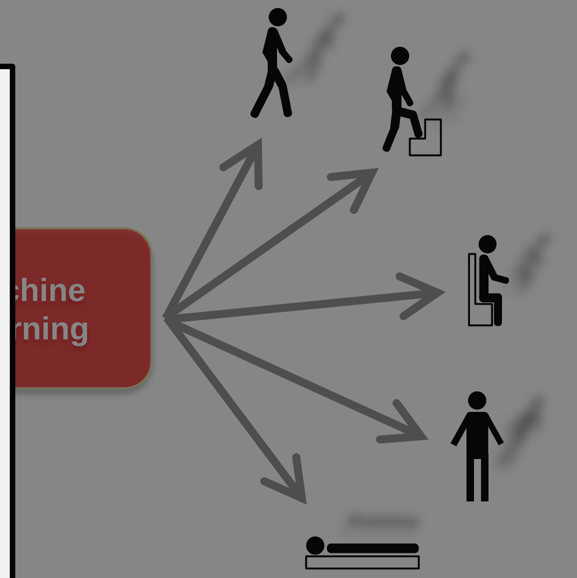
- Buffering helps a lot
- Hyperparameter tuni

Let's try it out!

Exercise:
humanActivityClassification.mlx
in folder <u>02-ClassificationModels</u>

Approach:

- Load buffered data
- Extract statistical fea
- Compare various ma (interactively)
- Optimize model using hyperparameter tuning



	30
	25
validateData	3
testData	2

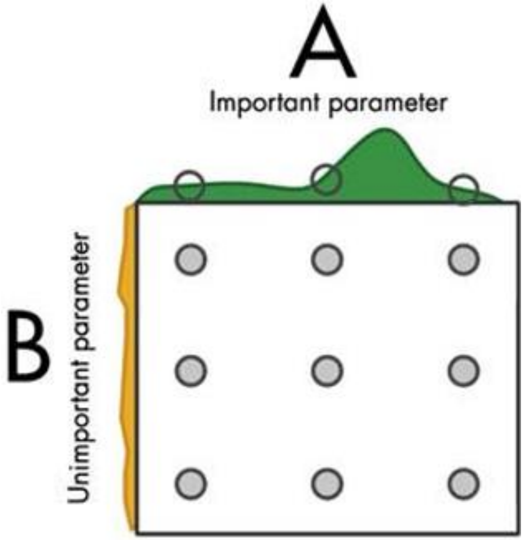
Combined to held-out validation set

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

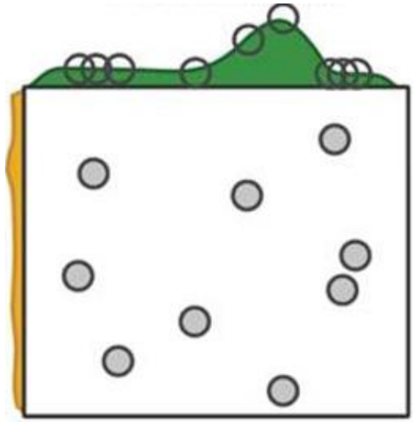
Hyperparameter Tuning

Why? – Model “knobs” (hyperparameters) need to be set properly for optimal performance

Standard:
Grid Search

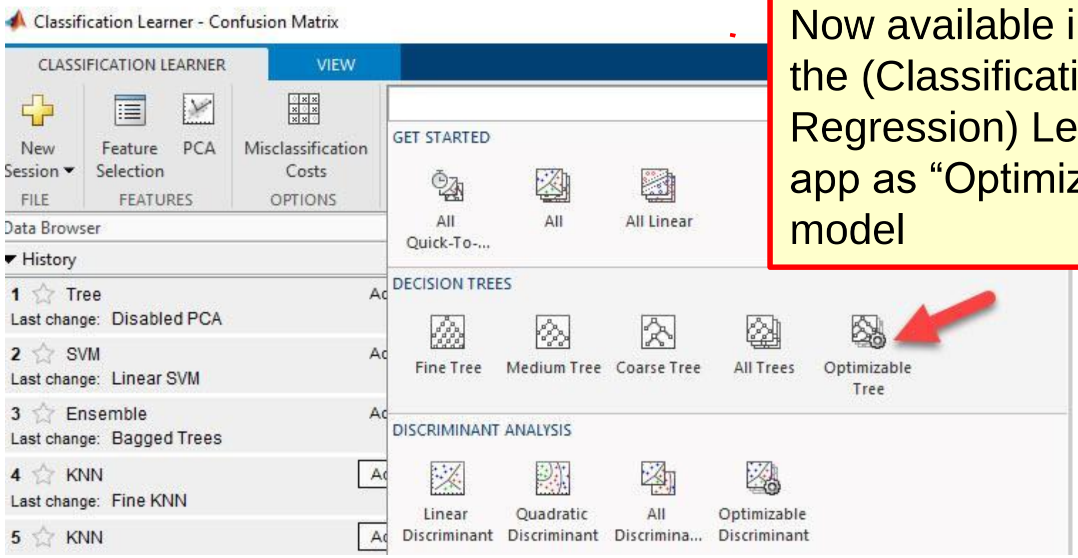


Better:
Random Search



Best: Bayesian Optimization

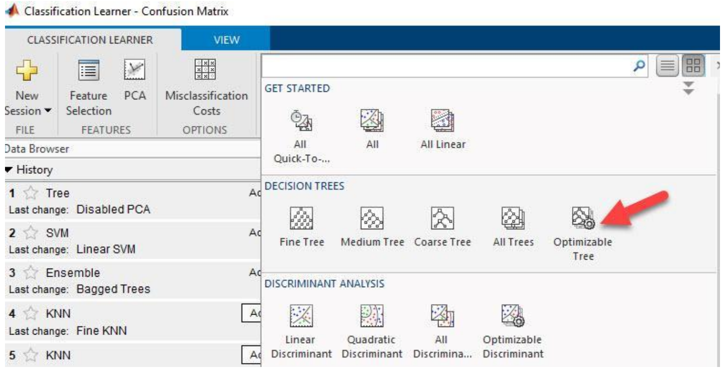
- Bayesian model indicates impact of change
- Model picks “good” point to try next
- Much more efficient!
- Scale to multi-cores (using PCT) for larger datasets



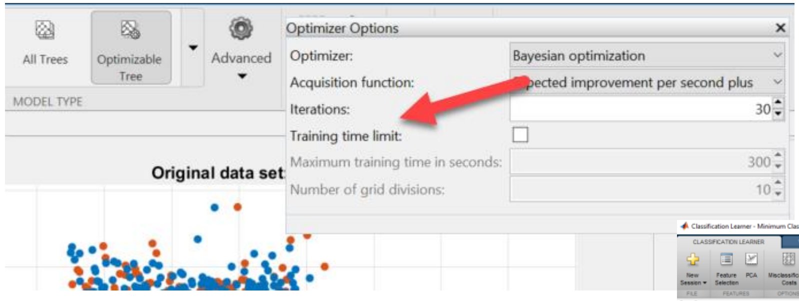
Now available inside the (Classification/ Regression) Learner app as “Optimizable” model

Hyperparameter Tuning Workflow inside Learner Apps

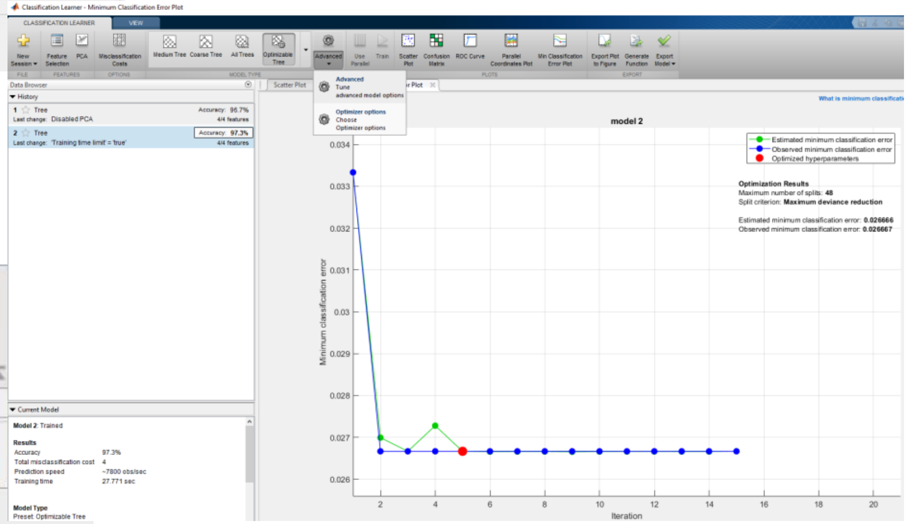
1. Choose “Optimizable” model from gallery



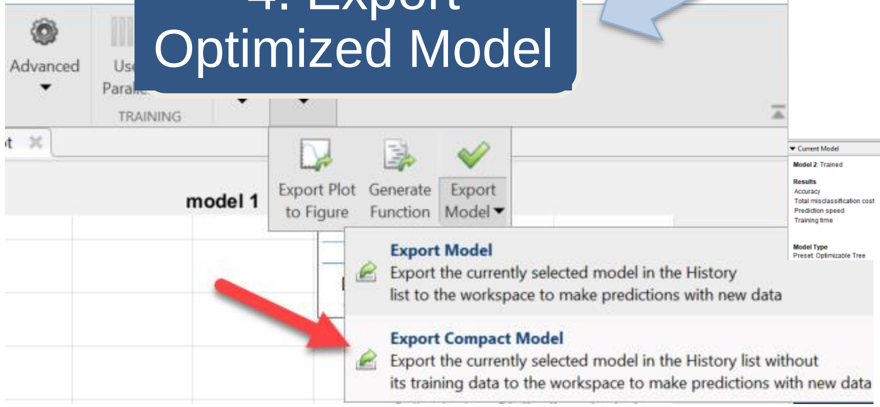
2. Adjust Optimizer Options (control runtime!)



3. “Train”: Bayesian Optimization iterates



4. Export Optimized Model



5. Iterate OR Prepare for Integration

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"

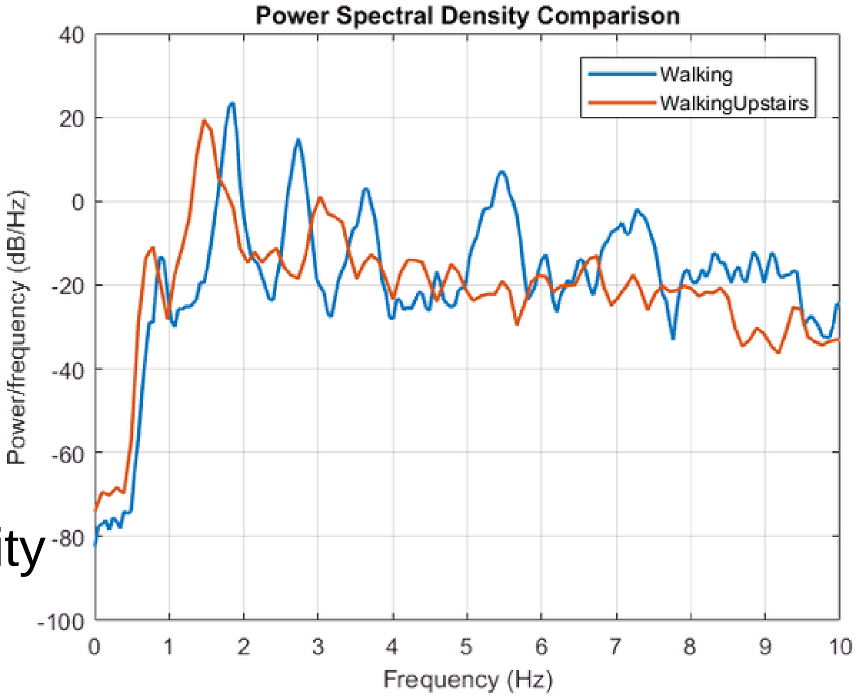
Sergey Yurgenson
(Kaggle Master)

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!



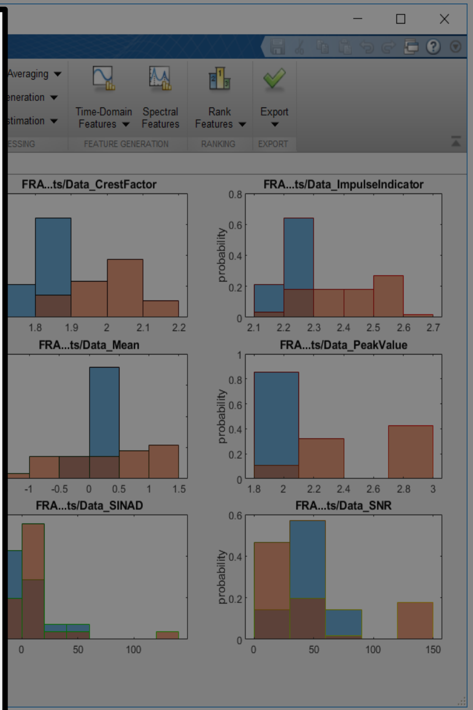
[How to use Diagnostic Feature Designer](#)[12 min video]

Exercise 3 – Feature Engineering for human activity

Goal: Explore different techniques for feature engineering

Approach:

– Use signal processin… extract time domain …
– Use feature selectio… reduce the set of fea… relevant
– Browse examples in MATLAB documentation for different applications



Let's try it out!

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

Feature Generation with Wavelet Scattering

Why? – Obtain good features “automagically”, without domain knowledge

What are Wavelets?

- Instead of decomposing signal into complete sinus waves, decompose into “wavelets”
- Tech Talks explaining WaveLets [4 videos]
- This conceptually looks like this:

Better than Spectrograms because can vary in scale!

Wavelet Scattering Framework [Bruna and Mallat 2013]

- Automatic Feature Extraction
- Reduces data dimensionality and provides compact features
- Works with both Signal and Image data [Texture example, Digit Classification]

Wavelet Scattering Nuts and Bolts

Pseudo-Code:

```
sf = waveletScattering(SignalLength);
Loop over signal
   waveletFeature = featureMatrix(sf,signal)
   Append  waveletFeature  to feature table
   Add labels
end
```

Additional Resources:

- Wavelet scattering Tech talk [4 min video]
- Wavelet scattering for ECG [doc example]
- Blog about Wavelet scattering on towardsdatascience.com

Diagnostic Feature Designer App

Predictive Maintenance Toolbox R2018b and R2019a

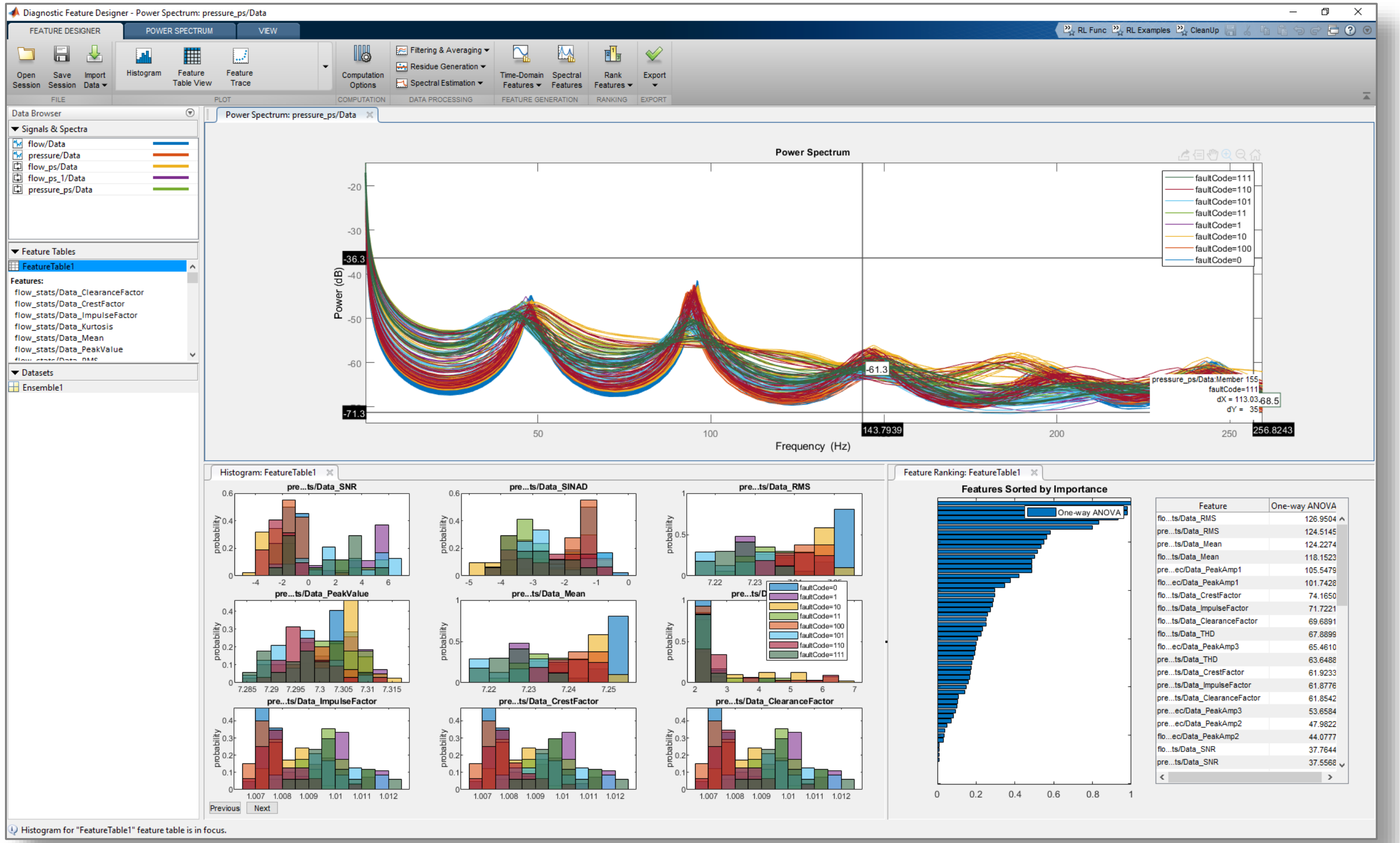
Why? – Empower signal domain expert to try all his favorite features.

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



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

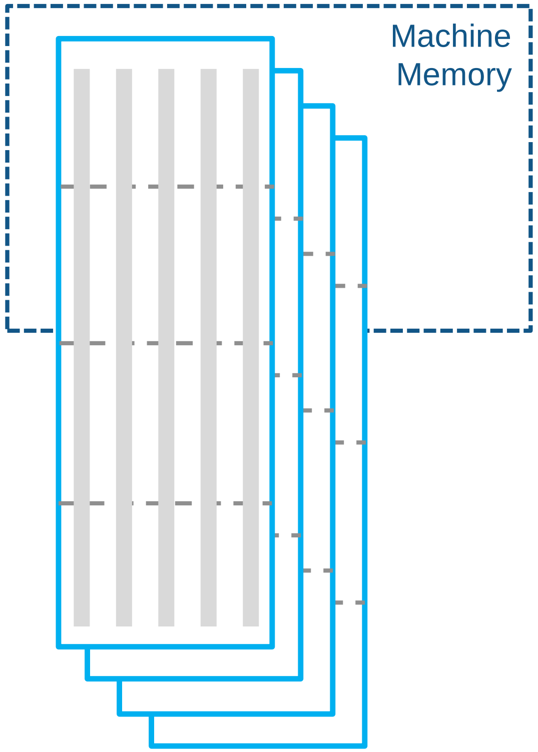
Extends the "array" data type to out-of-memory

- 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);
```

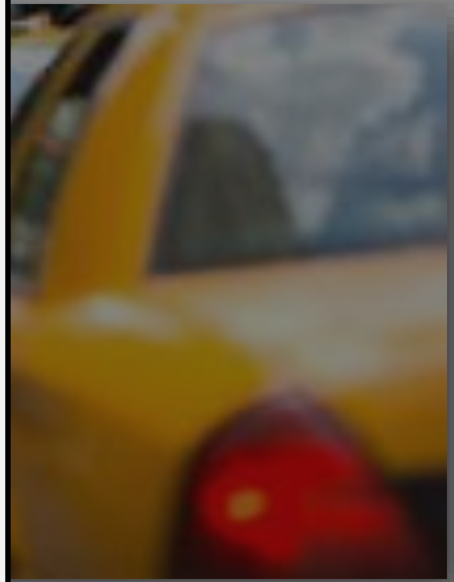
```
[m,s] = gather(m,s);
```

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

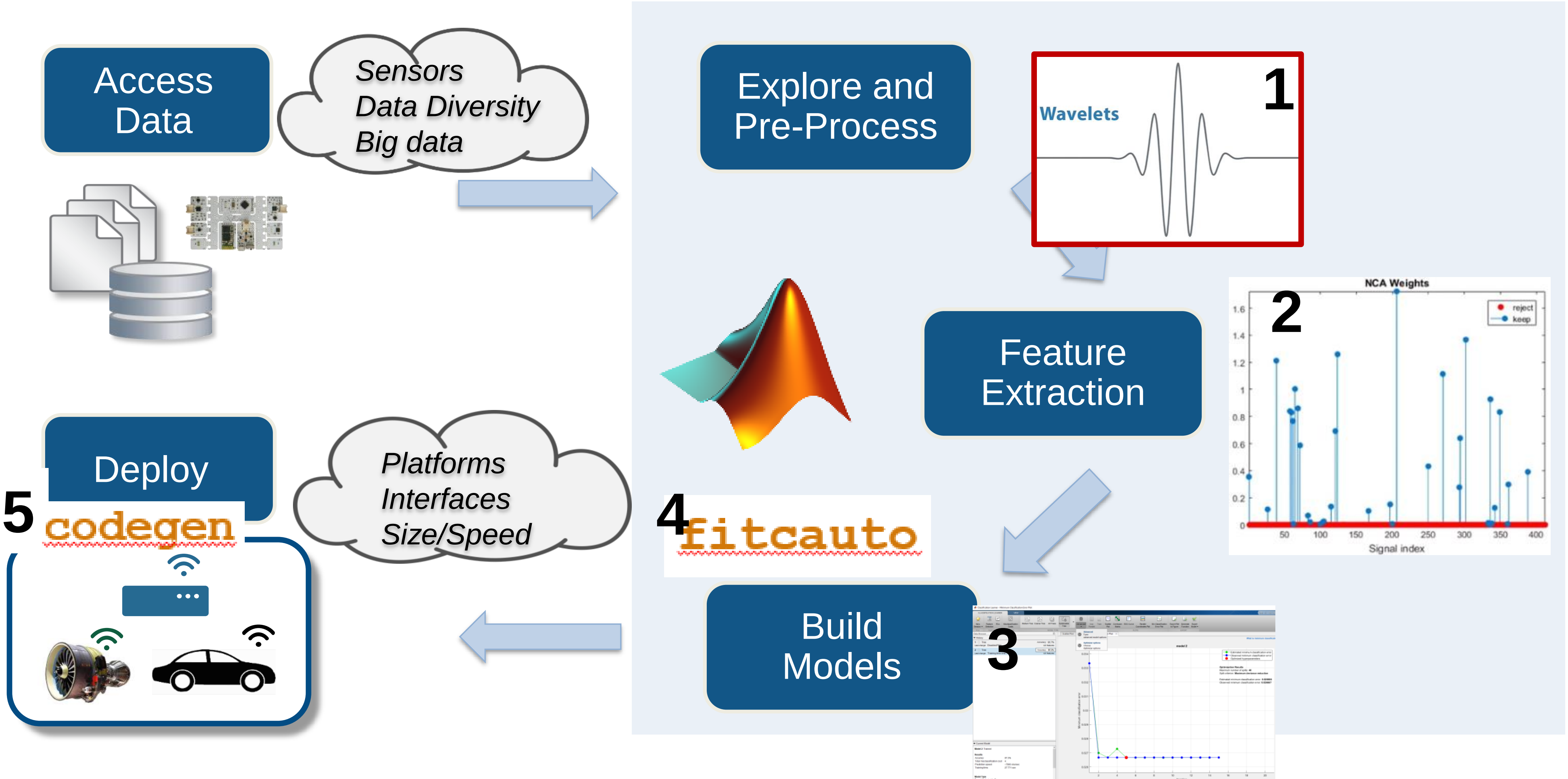


Let's try it out!

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

What's our AutoML?

Automate main steps to minimize expertise needed and increase productivity

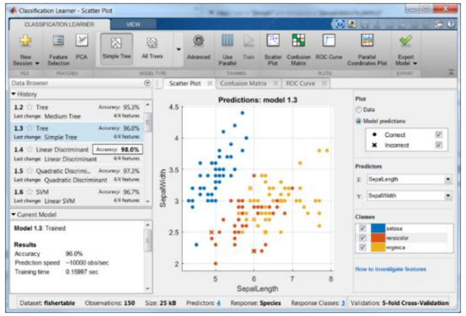


Summary: Complete Machine Learning Workflow

1. Easy to Learn and Use
2. Engineer Features & Optimize Model
3. Deploy Anywhere: Embedded Device and Enterprise IT/OT

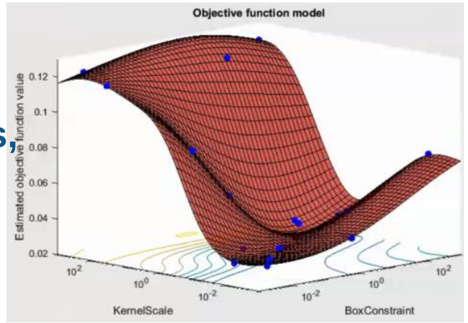
Learner Apps

Train Classification & Regression models without coding



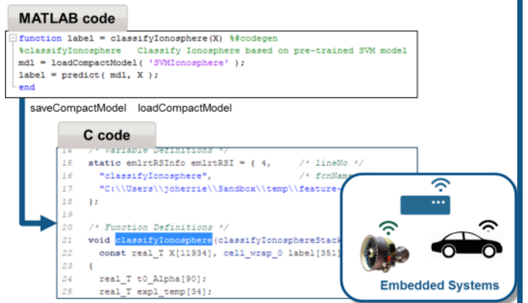
Automation

Optimize hyperparameters, Generate and Select features

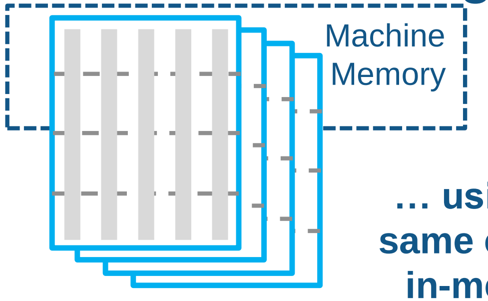


C Code Generation

Deploy not just prediction, but also preprocessing



Scale to Big Data

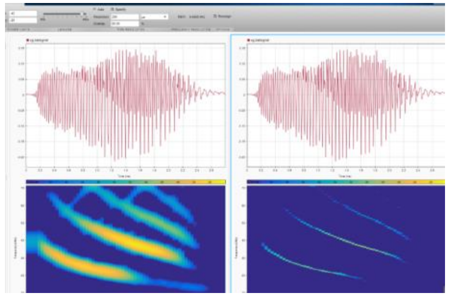


… using the same code as in-memory

Tall Data

Domain Support

Predictive Maintenance, Text, Signal and Image Processing,



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
- Take a paid training on Machine Learning or Big Data

Resources

Machine Learning Onramp (2 hr online introduction)

Machine Learning with MATLAB:

– Overview, Cheat sheet
– Machine Learning Intro (Tech talk videos)
– Machine Learning with MATLAB Introduction (eBook)
– Mastering Machine Learning (eBook)
– Applied Machine Learning (Tech Talk videos)
– Practical Data Science with MATLAB (Coursera Specialization)

Machine and Deep Learning

– Deep vs. Machine Learning: Choosing the Best Approach (eBook)
– Deep learning Onramp (2hr online introduction)

MathWorks® can help you do Machine Learning

Free resources:

- Guided evaluations
- Proof-of-concept projects
- Seminars
- Other Hands-on workshops

More options:

- Paid Training (2-day Machine Learning, 1-day Big Data, see Appendix)
- Advanced customer support
- Enterprise and cloud deployment
- Consulting services

MathWorks®

MathWorks®

Accelerating the pace of engineering and science