MathWorks

Machine Learning with MATLAB on SNIC

A hands-on MATLAB workshop

Antti Löytynoja, *Senior Application Engineer*, **MathWorks**
Emelie Andersson, *Application Engineer*, **MathWorks**
Sagar Zade, *Customer Success Engineer*, **MathWorks**
Anders Sjöström, *System Administrator*, **Lund University**

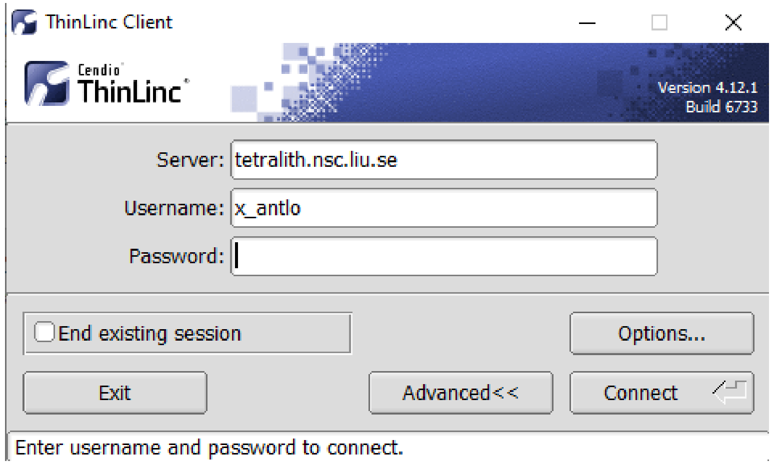
How can you participate?

You have three options:

1. Run exercises yourself on Tetralith (you need to have SNIC and Tetralith credentials)
    1. Bonus: you have access to at least 8 cores (and even more outside the workshop)
2. Run exercises on your own computer (if you don't have access to Tetralith)
    1. Download exercise material: **tinyurl.com/rebrxrfz**
3. Get your popcorn and just follow along

In case you want to run on Tetralith

- Login to Tetralith using Thinlinc



- Open terminal window
  - Copy workshop material:
  - `cp -r /home/x_antlo/Public/ExerciseFiles .`
  - Get on a compute node:
  - `interactive --reservation=matlab-monday -n8 -t 03:00:00`
  - This gives you 8 cores for three hours on a node that has been reserved for us

- Start MATLAB (from terminal)
  - `module load MATLAB/R2021a-nsc1`
  - `matlab -softwareopengl`

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.

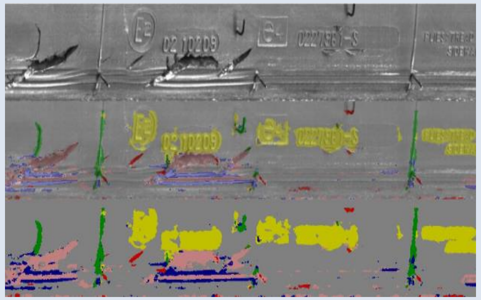
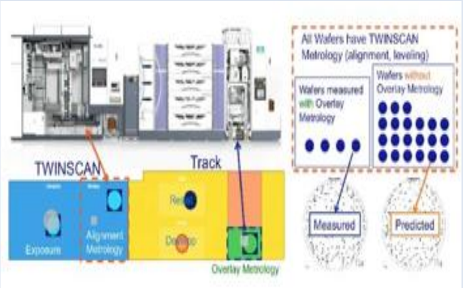
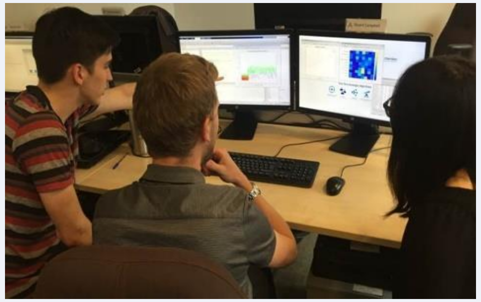
2:20 PM · Mar 15, 2018

Agenda

Machine learning introduction

- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
  - Feature engineering
  - AutoML
  - Interpretability
- Unsupervised learning (optional)
- Working with big data (optional)
- Text Analytics (demo)

Machine Learning is Everywhere

Automobile	Manufacturing	Comms & Operations	Energy & Finance
 Tire Wear	 Overlay metrology improvement	 Telecom customer churn prediction	 Forecasting & Risk Analysis
 Detect Oversteer	 Monitor Deployed Compressors using Digital Twin	 Building energy use optimization	 Portfolio Allocation

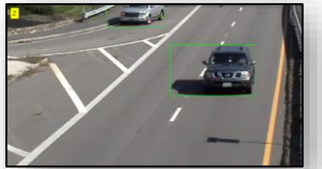
9

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

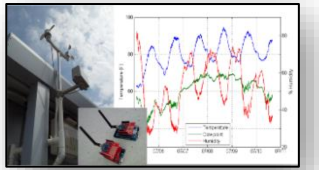
Weather Forecasting

Energy Load Forecasting

Stock Market Prediction

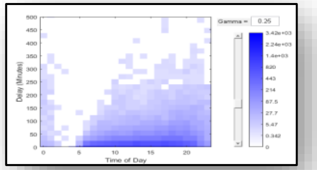
update as more data becomes available

Solution needs to scale



IoT Analytics

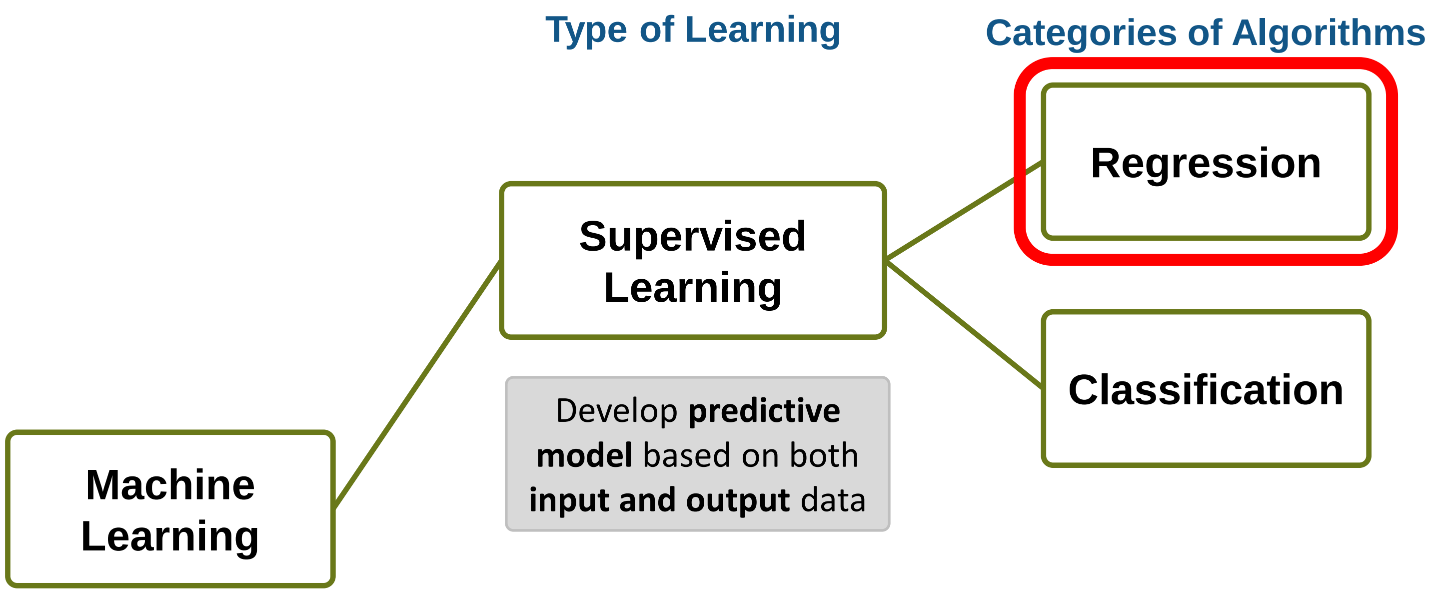
Taxi Availability



Airline Flight Delays

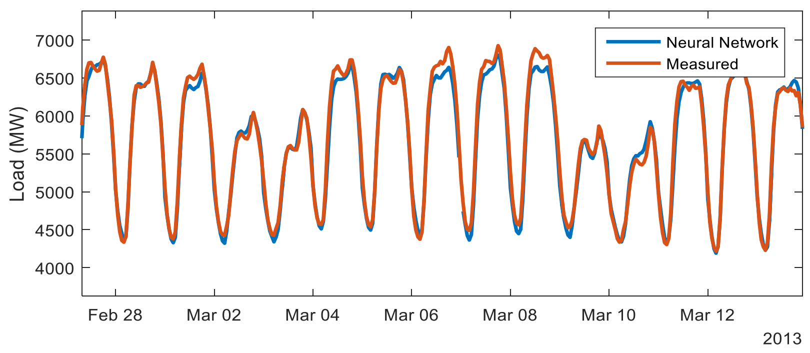
learn efficiently from very large data sets

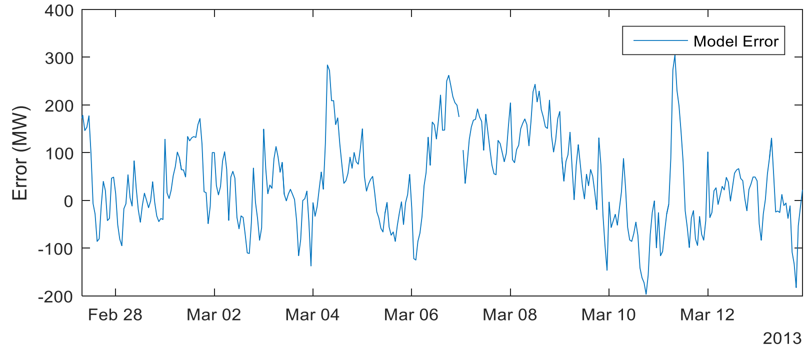
Types of Machine Learning



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

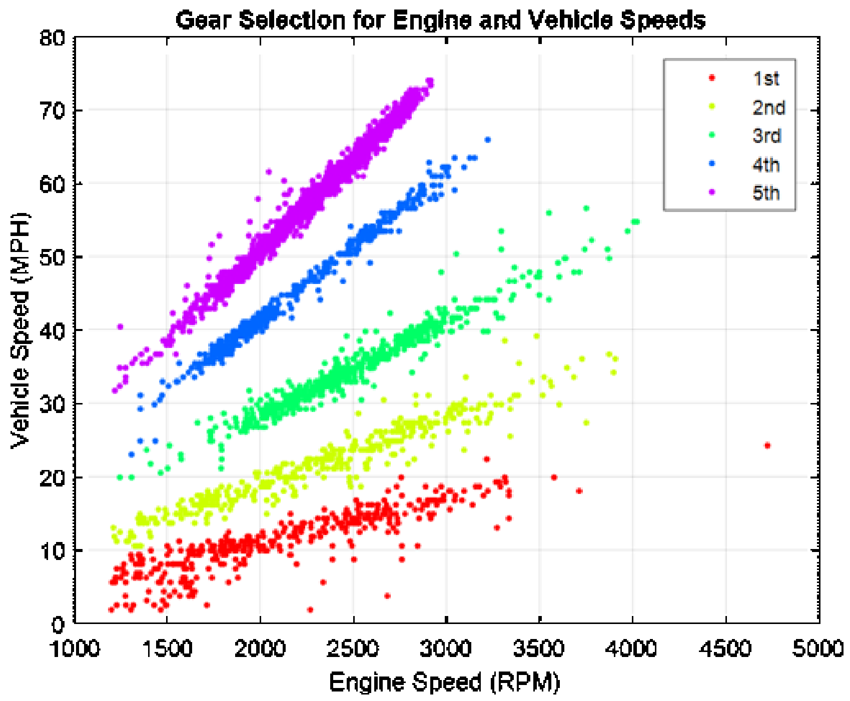
Type of Learning | **Categories of Algorithms**

Machine Learning

- **Supervised Learning**
  - Develop **predictive model** based on both **input and output** data
  - **Regression**
  - **Classification**
- **Unsupervised Learning**
  - Discover an **internal representation** from **input data** only
  - **Clustering**

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

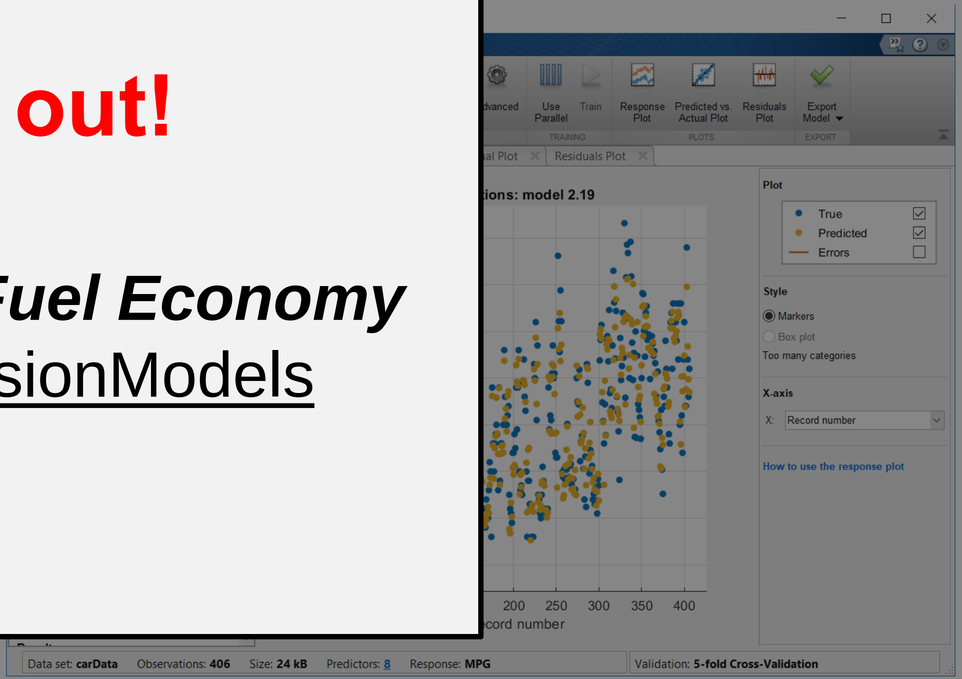
Regression

Goal: Study drivers of

- Build initial models
- Don't need to be a

Approach:

- Load data in MATL
- Use the Regressio
  multiple regressio
- 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

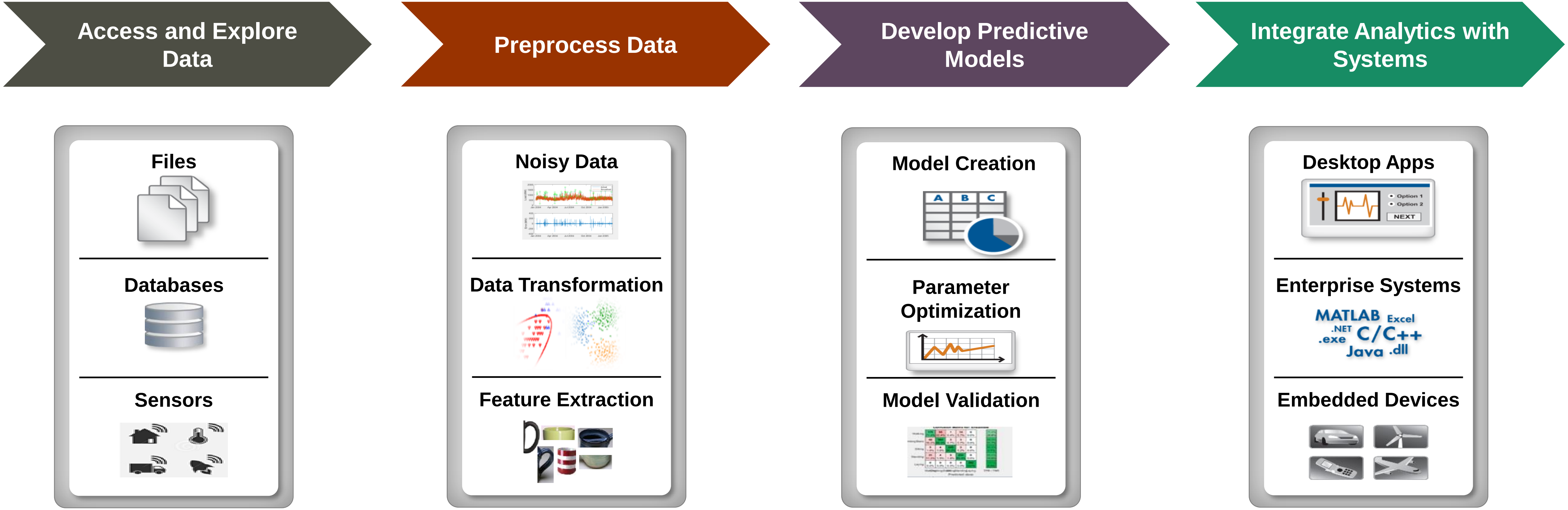
“essentially, all models are wrong, but some are useful”

– 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



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

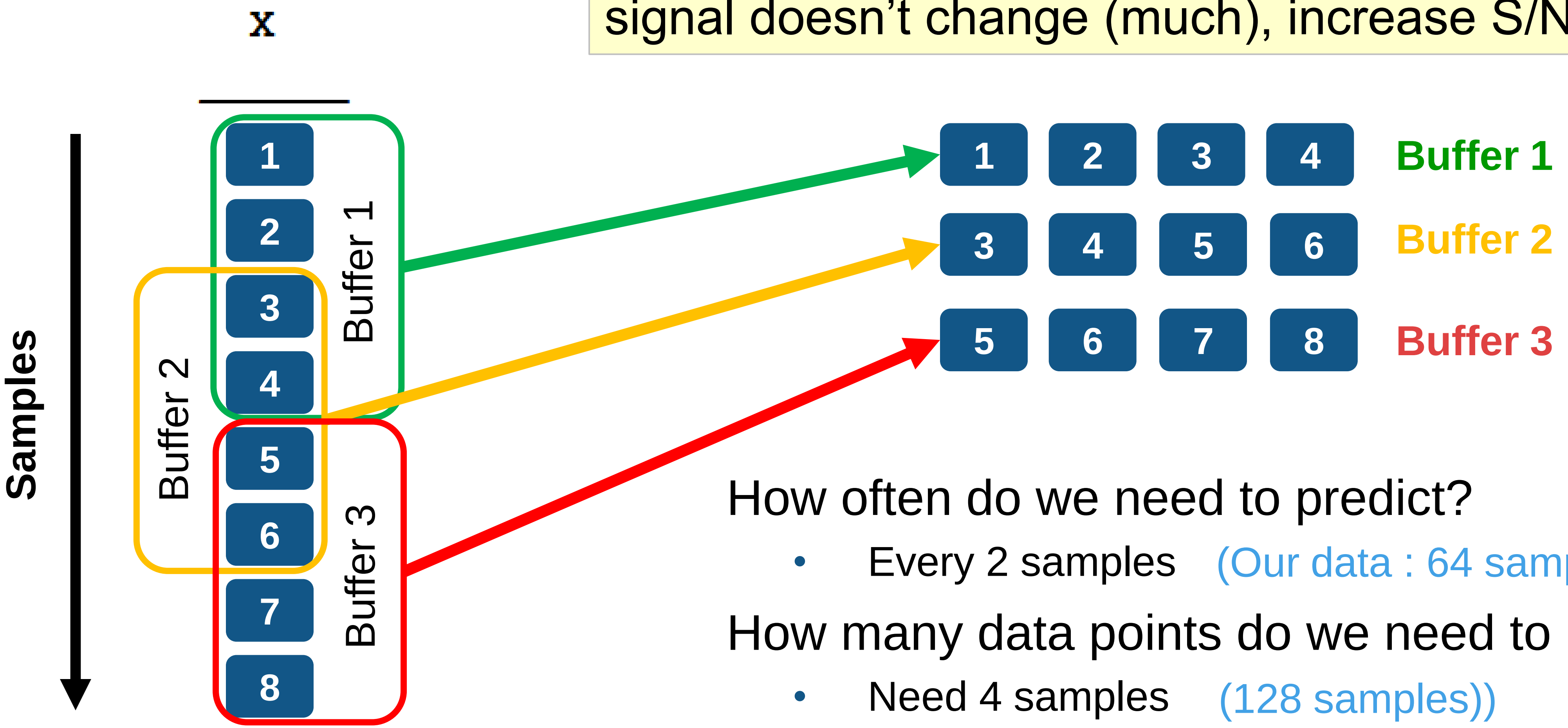
- Data specific processing
- Feature Extraction
- Feature Selection

- Many different models
- Model tuning
- Computationally intensive

- Different end users
- Different target platforms
- Different Interfaces

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

MathWorks

Exercise 2: Human Activity Learning using Smartphones

Goal: create initial mod

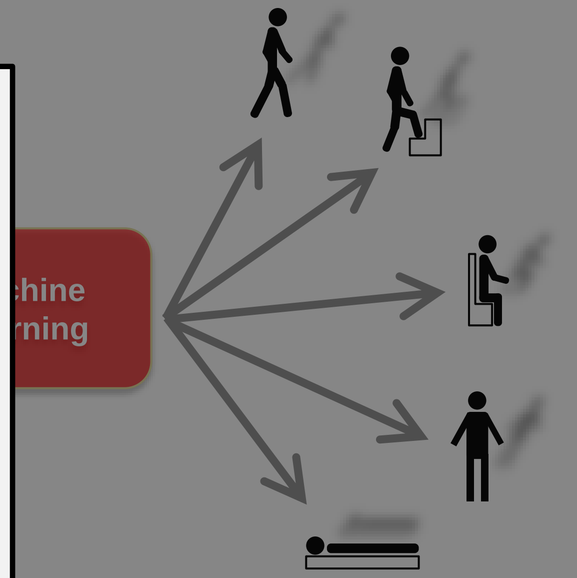
- Buffering helps a lot
- Hyperparameter tuni

Approach:

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

Let's try it out!

Exercise:
humanActivityClassification.mlx
in folder 02-ClassificationModels



	30
	25
validateData	3
testData	2

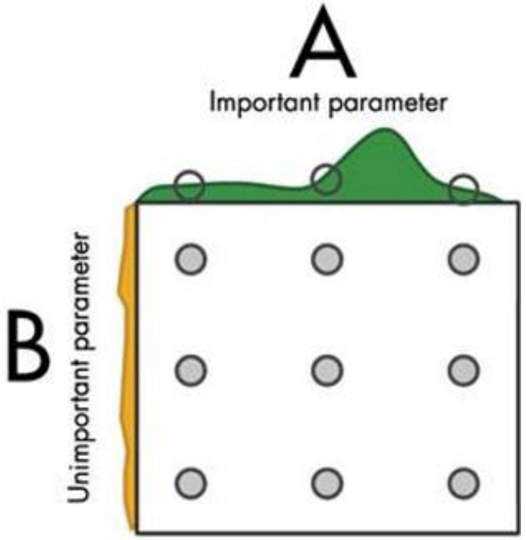
} Combined to held-out validation set

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

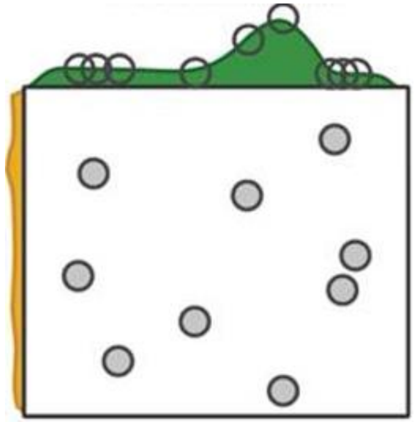
Hyperparameter Tuning

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

Standard:
Grid Search

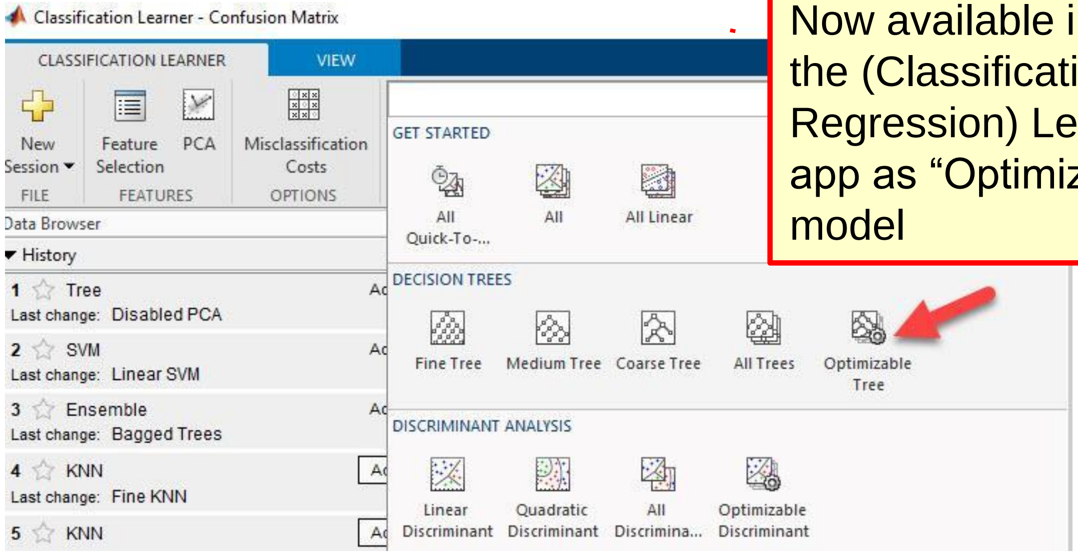


Better:
Random Search



Best: Bayesian Optimization

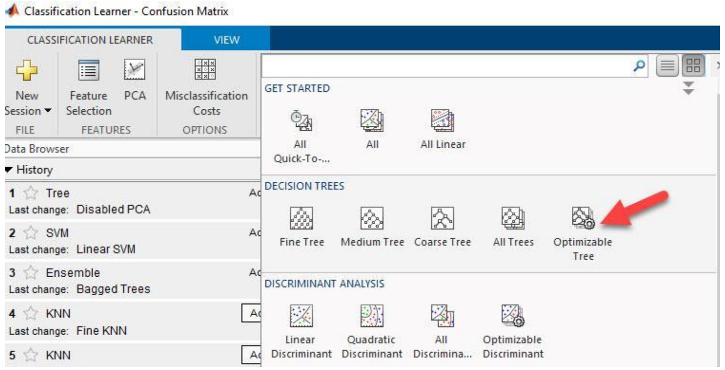
- Bayesian model indicates impact of change
- Model picks “good” point to try next
- Much more efficient!
- Scale to multi-cores (using PCT) for larger datasets



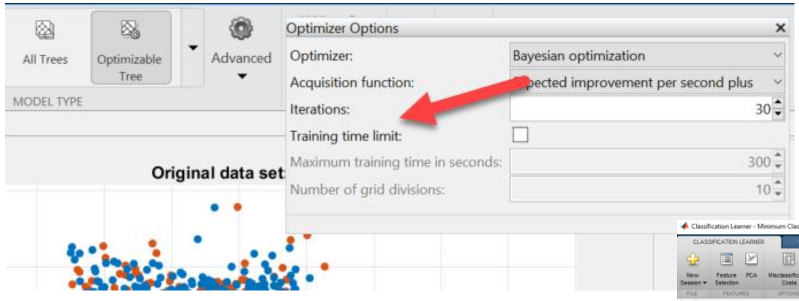
Now available inside the (Classification/ Regression) Learner app as “Optimizable” model

Hyperparameter Tuning Workflow inside Learner Apps

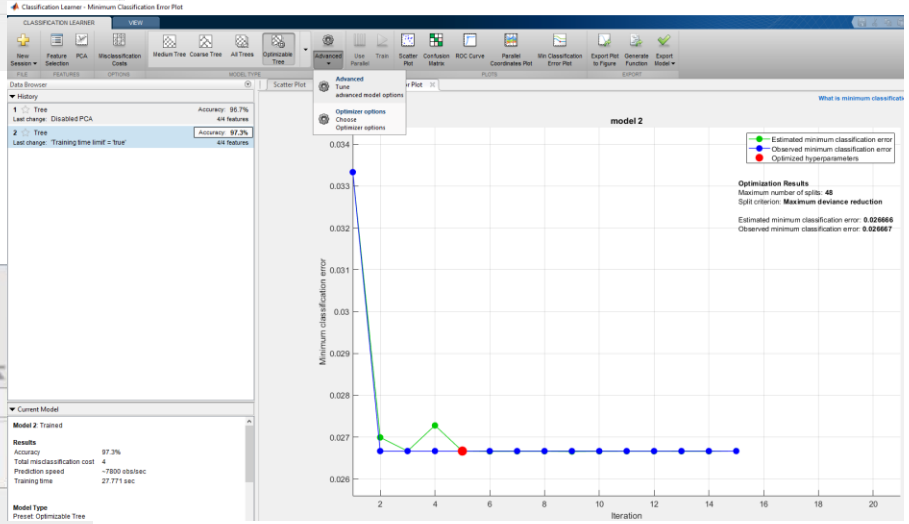
1. Choose “Optimizable” model from gallery



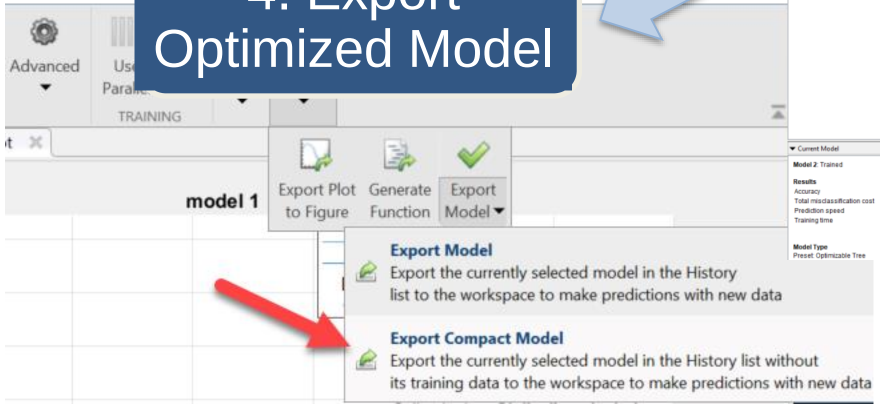
2. Adjust Optimizer Options (control runtime!)



3. “Train”: Bayesian Optimization iterates



4. Export Optimized Model



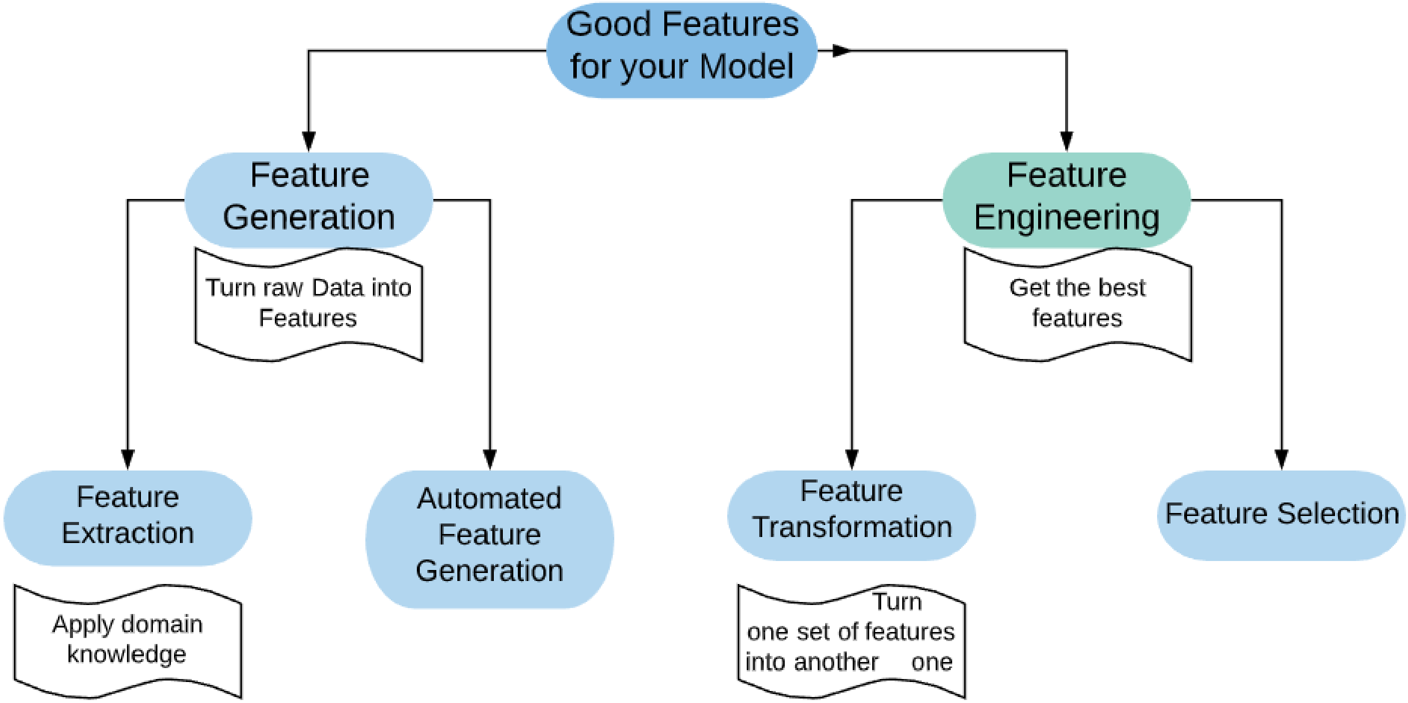
5. Iterate OR Prepare for Integration

Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
  - Feature engineering
  - AutoML
  - Interpretability
- 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)

What could you try?

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

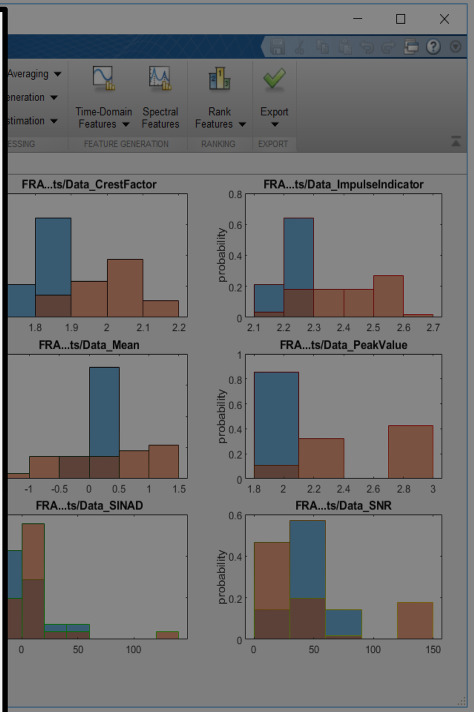
[How to use Diagnostic Feature Designer](#)
[12 min video]

Exercise 3 – Feature Engineering for human activity

Goal: Explore different techniques for feature engineering

Approach:

- Use signal processi[…] 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

Diagnostic Feature Designer App

Predictive Maintenance Toolbox R2018b and R2019a

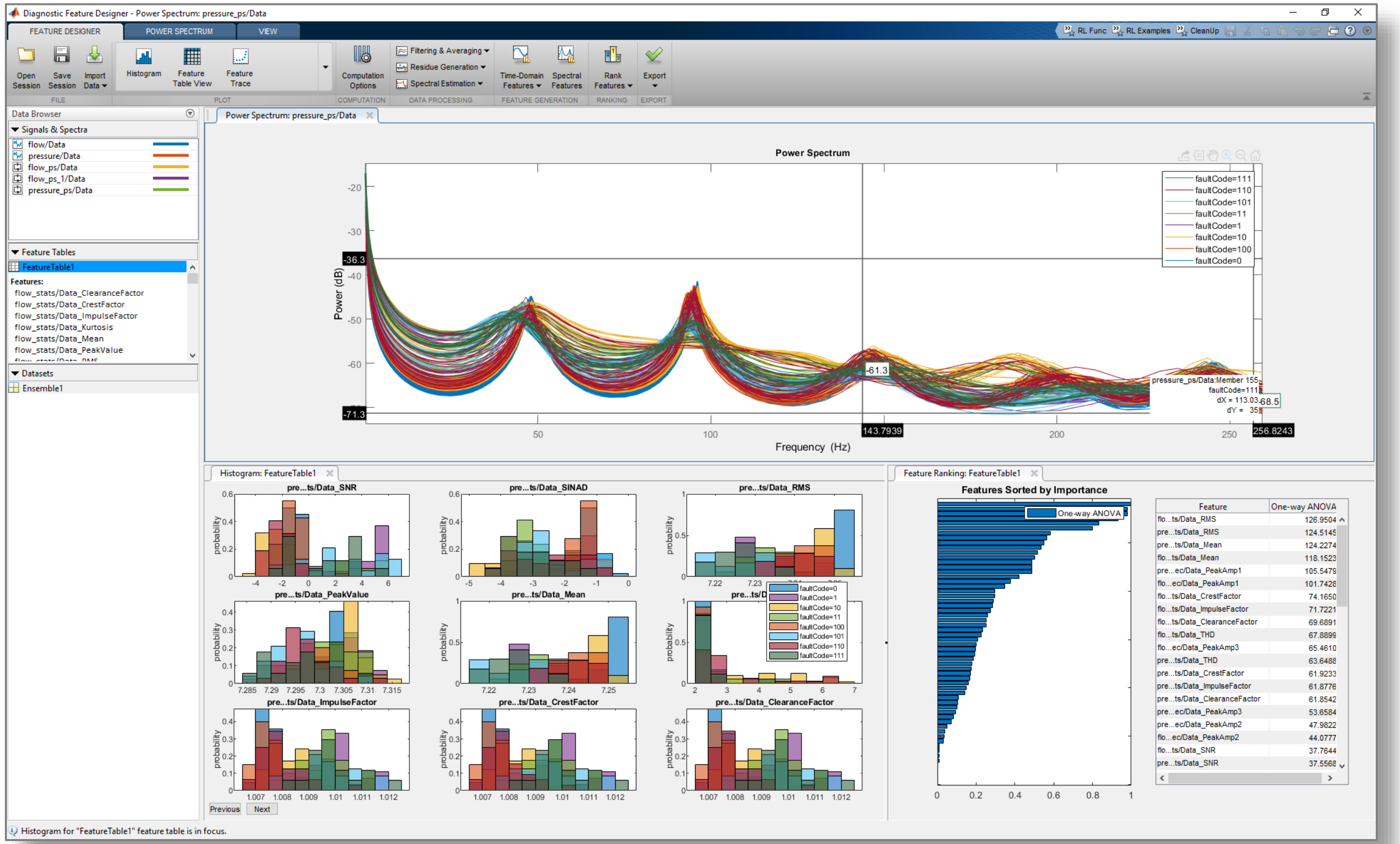
Why? – Empower signal domain expert to try all his favorite features.

Extract, visualize, and rank features from sensor data

Use both statistical and dynamic modeling methods

Work with out-of-memory data

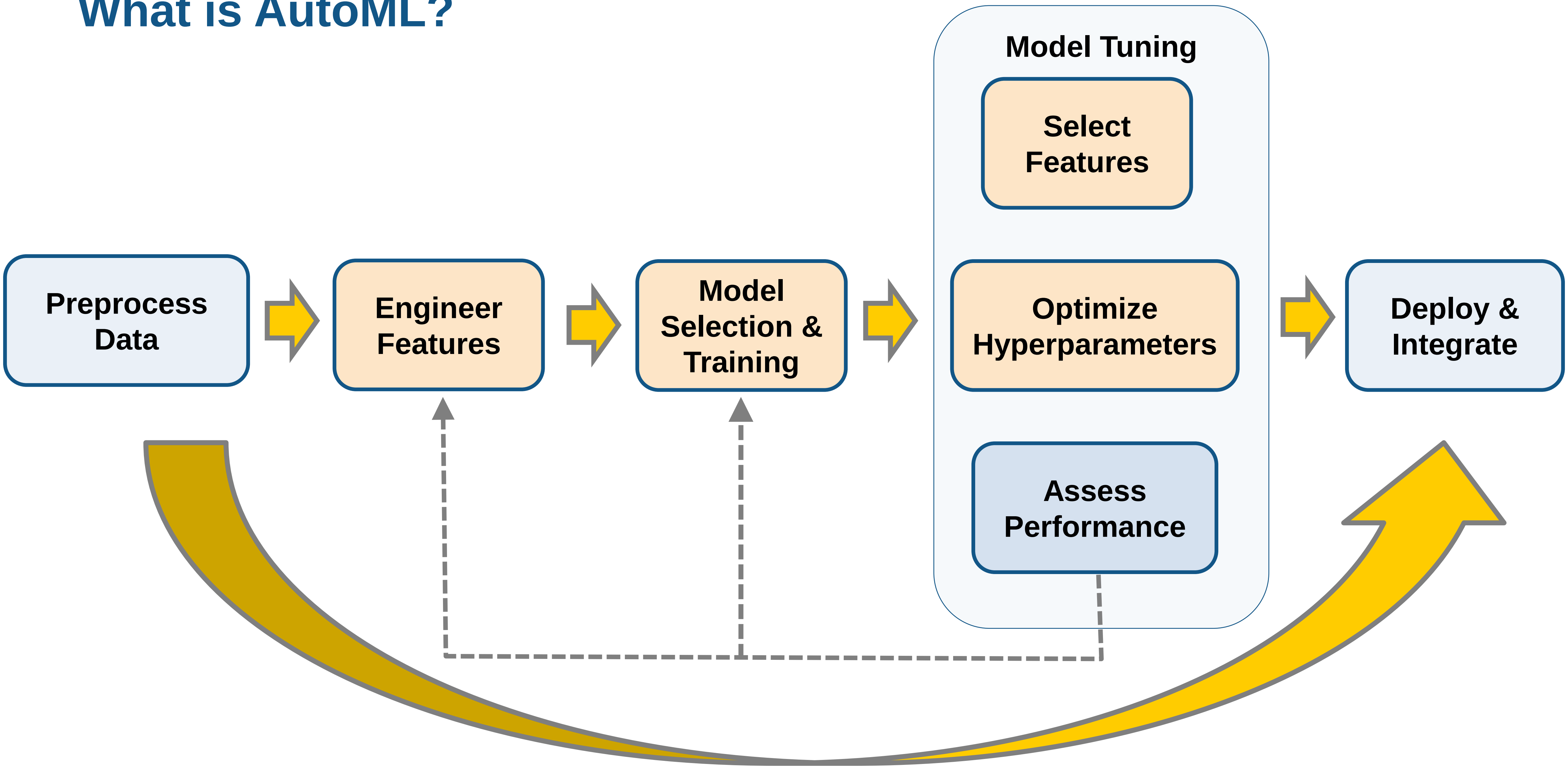
Explore and discover techniques without writing MATLAB code



Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
  - Feature engineering
  - AutoML
  - Interpretability
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

What is AutoML?



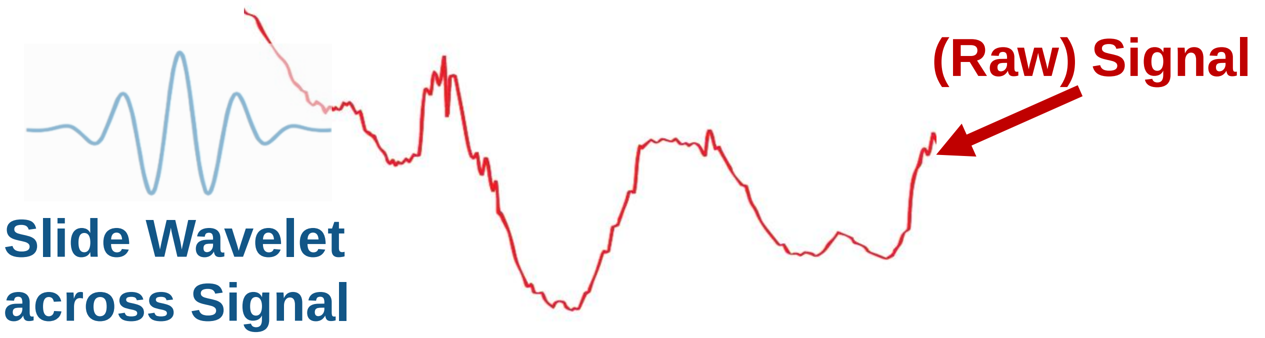
Feature Generation with Wavelet Scattering

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

Agenda

- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
  - Feature engineering
  - AutoML
  - Interpretability
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

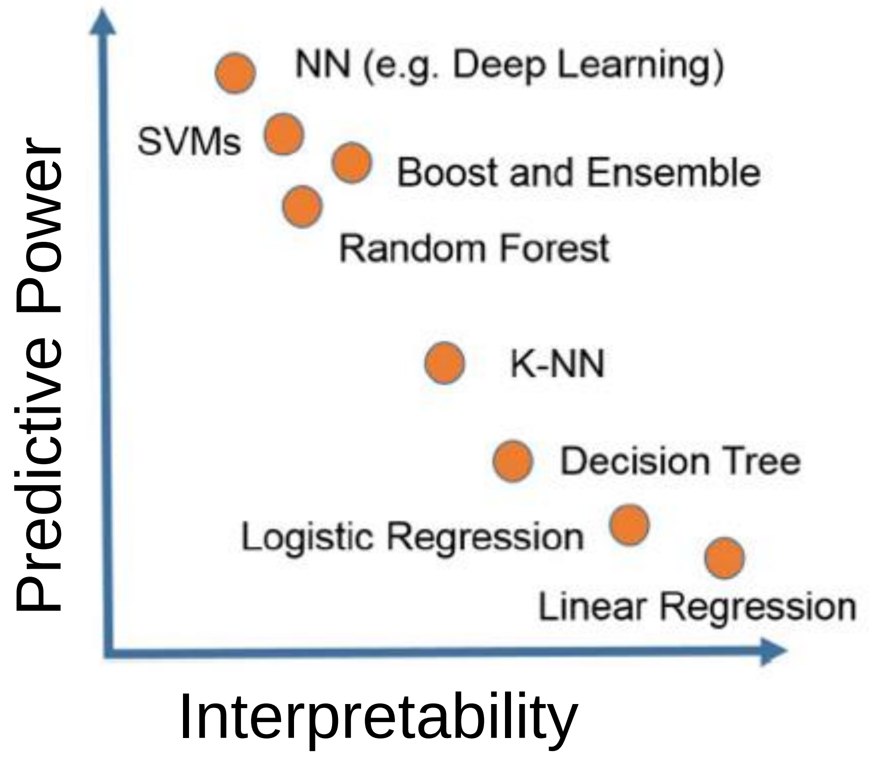
Interpretability and Explainability

Both terms describe the process of making “blackbox” models understandable

- “Interpretability”: primarily “classic” machine learning, causality of specific model decisions
- “**Explainable AI**” often refers to AI=Deep Learning, sometimes explaining how model works

Why Interpretability?

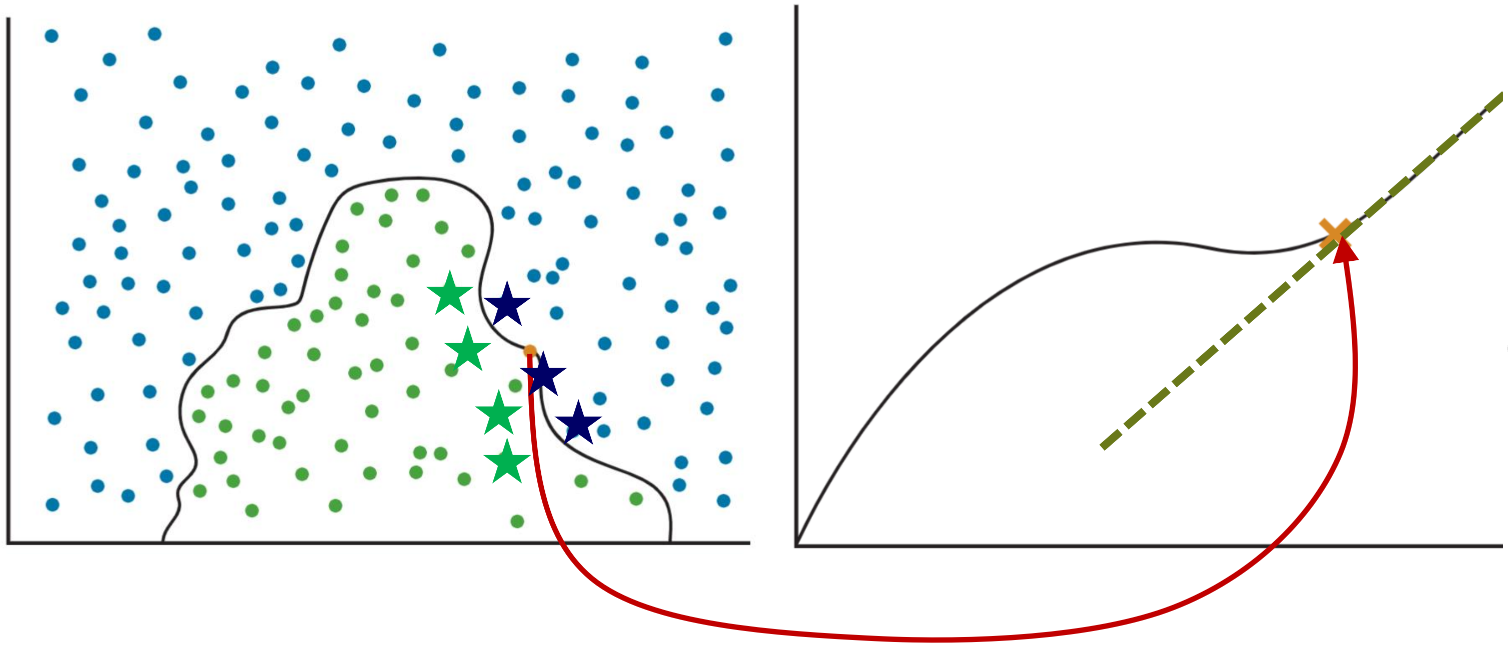
1. Overcome “blackbox” model
   - Not acceptable by company guidelines
   - Build trust for users unfamiliar with machine learning
   - Pick model that looks at “right” evidence
2. Regulatory requirements (Finance, Europe’s GDPR):
3. Debug models

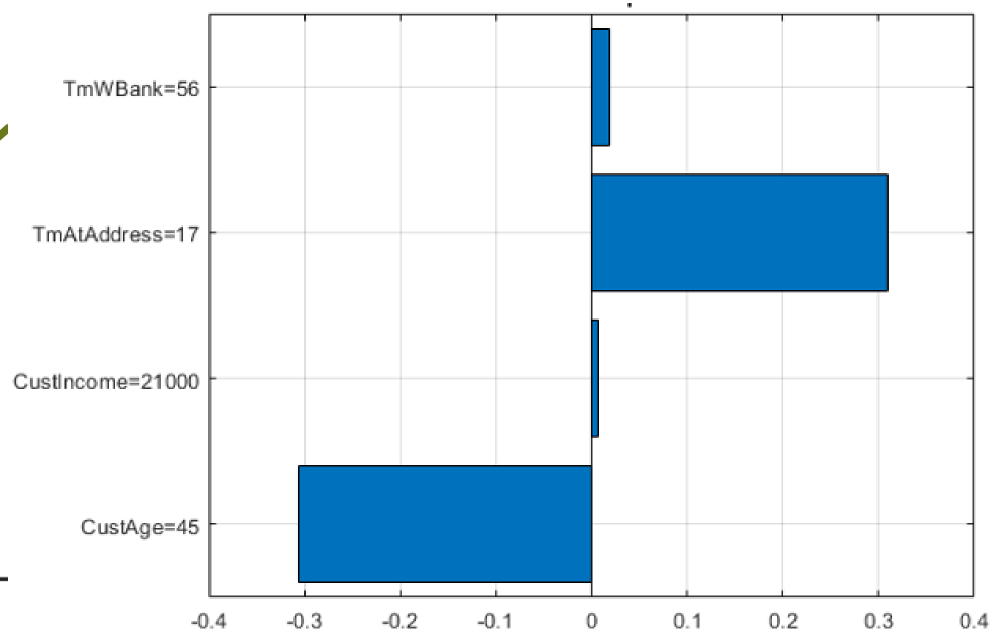


LIME = Local Interpretable Model-Agnostic Explanations

1 Approximate complex model near Point of Interest with simple model

2 “Explain” using weights of simple model






```
lObj = lime(mdl,'QueryPoint',queryPt,'NumImportantPredictors’,4);
plot(lObj);
```

Agenda

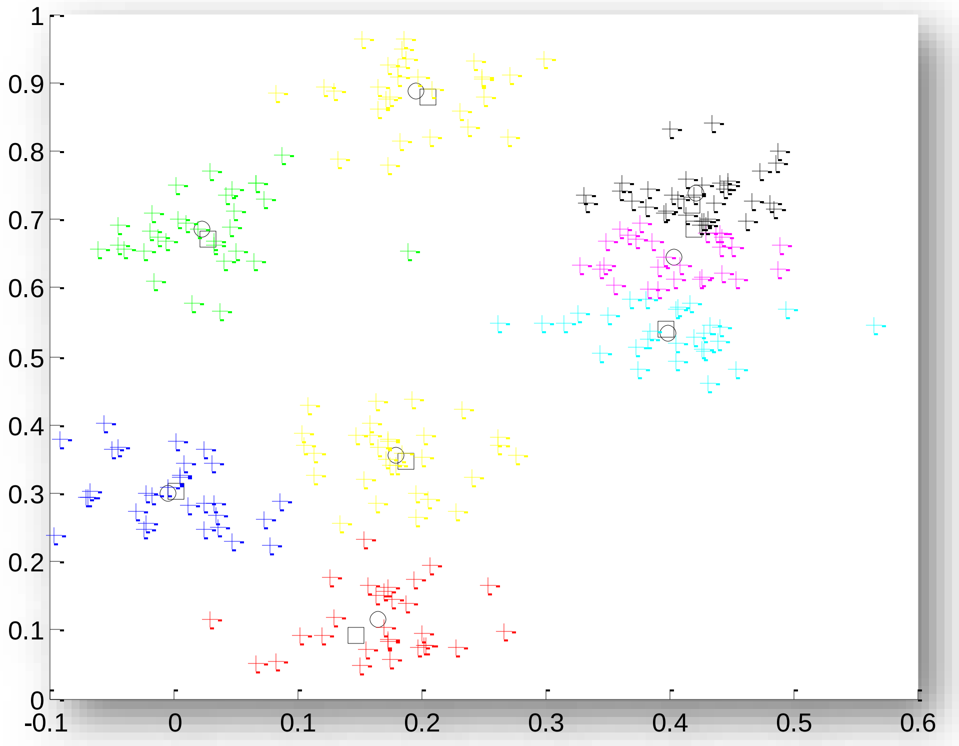
- Machine learning introduction
- Supervised machine learning models
  - Predicting fuel economy (Regression)
  - Human activity learning (Classification)
- Feature extraction and feature selection
- Unsupervised learning (optional)
- Working with big data (optional)
- Deploying Machine Learning Algorithms

Clustering

Why? Discover patterns, identify possible features, check for outliers

What is clustering?

Segment data into groups, based on data similarity



How is clustering done?

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

Exercise 4: Clustering Human Activity

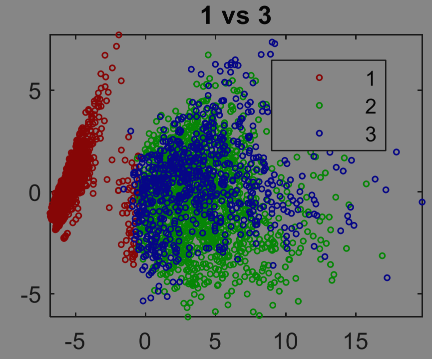
Goal: find additional options to improve
human activity clas

Approach:

– Reduce dimensi
structure of data

– Evaluate differe
identify groups c

1 vs 2



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

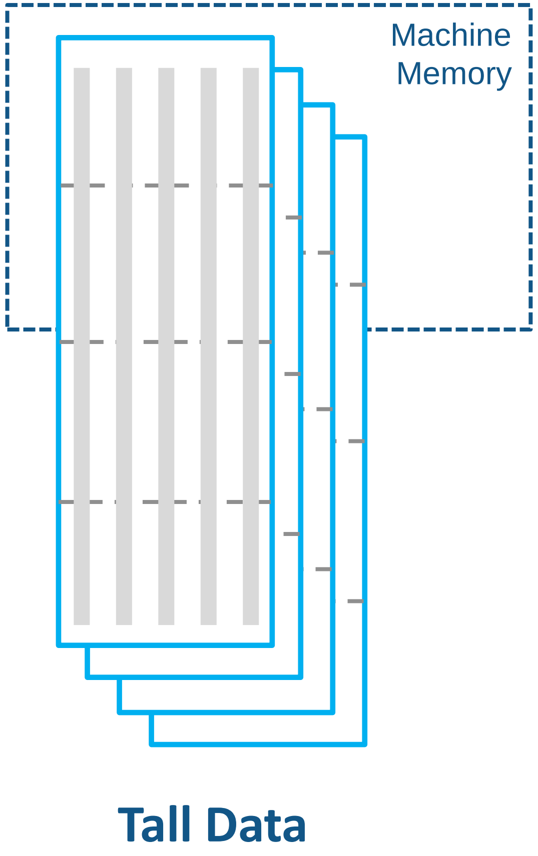
Extends the “array” data type to out-of-memory

- 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



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);
```

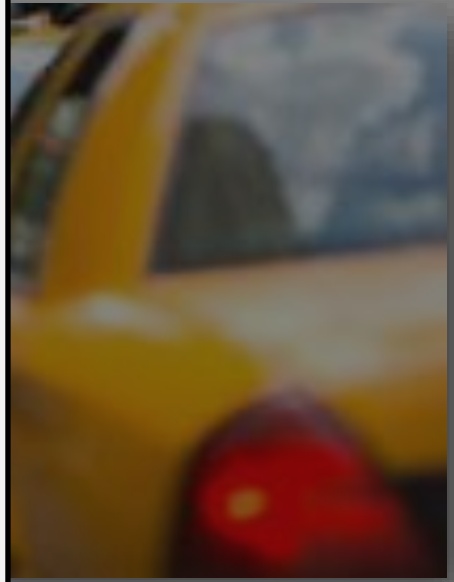
```
[m,s] = gather(m,s);
```

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

MATLAB

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

Embedded Hardware

C, C++

HDL

PLC

CUD

Let's observe!

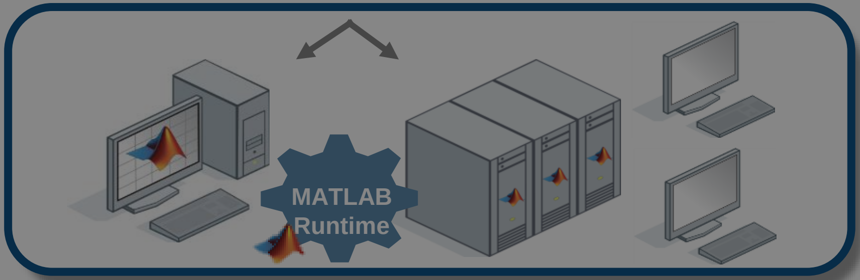
ns

a

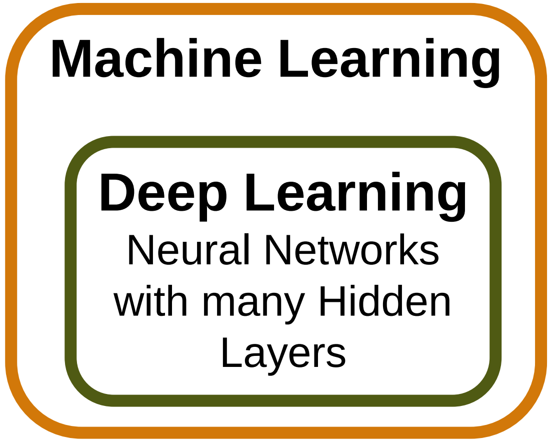
Python

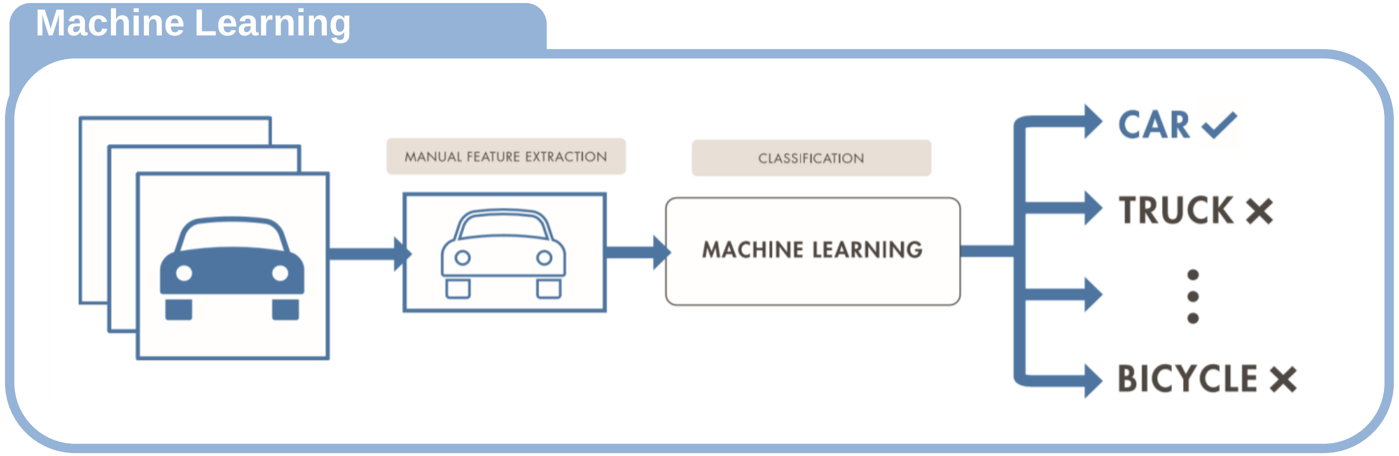
.NET

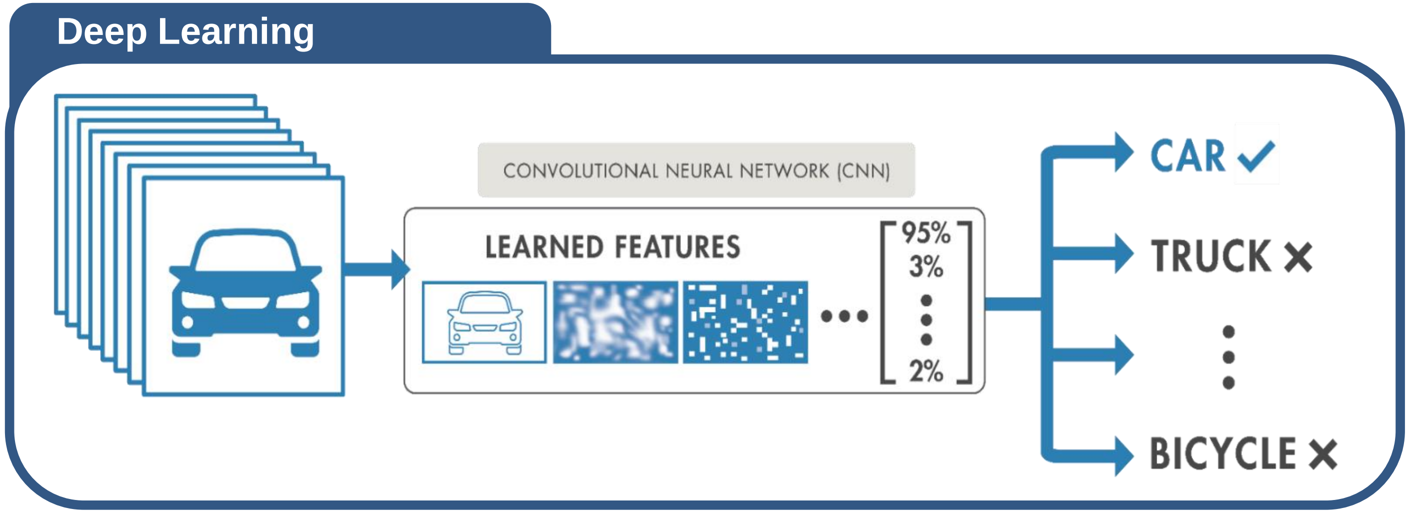
MATLAB Production Server



Beyond traditional Machine Learning: Deep Learning

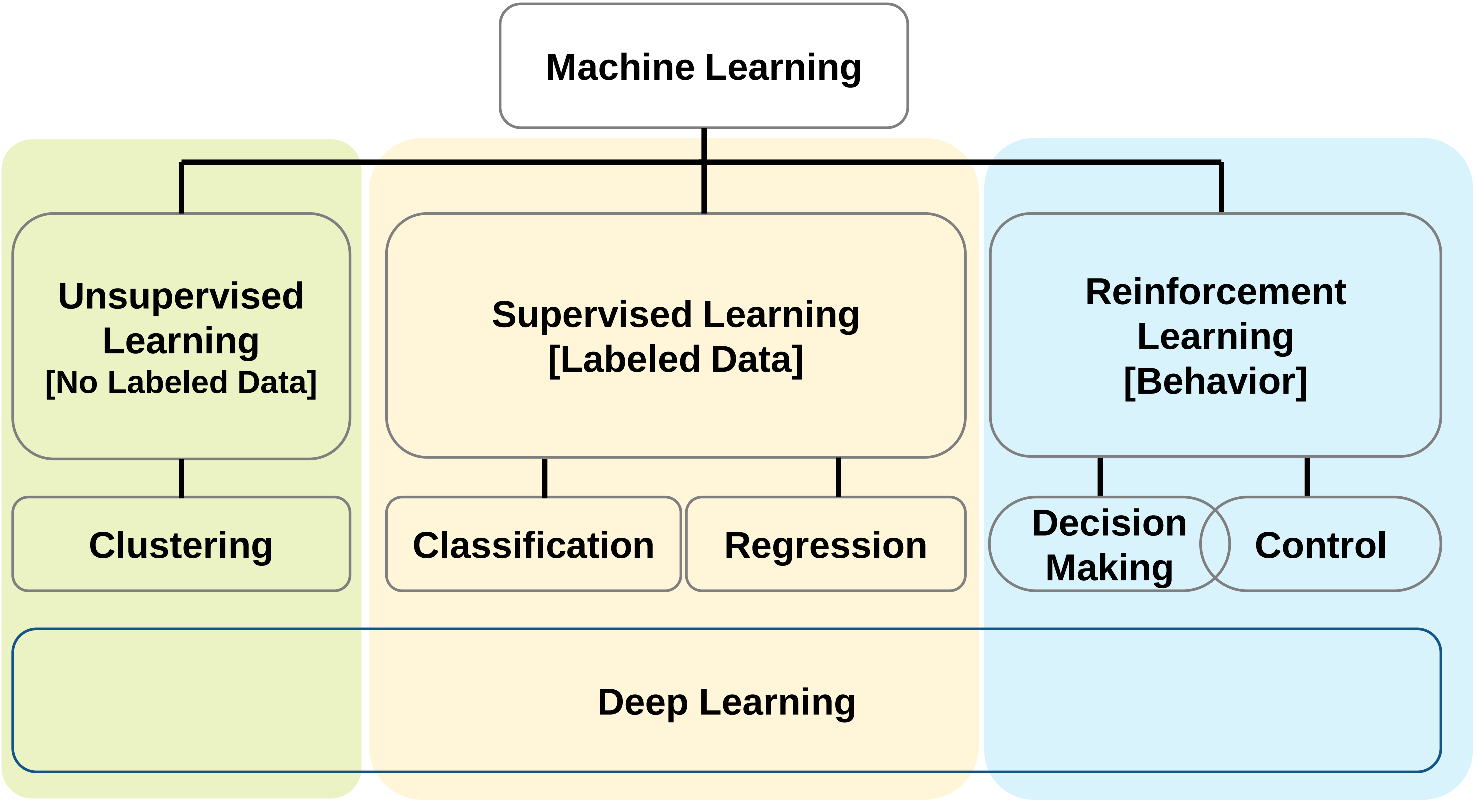






- Learns directly from data
- More Data = better model
- Computationally Intensive
- **Not interpretable**

Beyond Machine Learning: Reinforcement Learning



Reinforcement learning:

Learning through trial & error [*interaction data*]

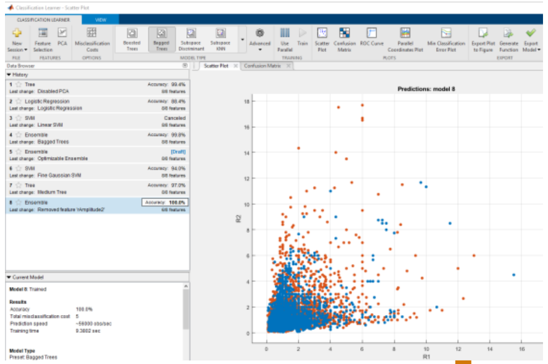
Complex problems typically need deep learning

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

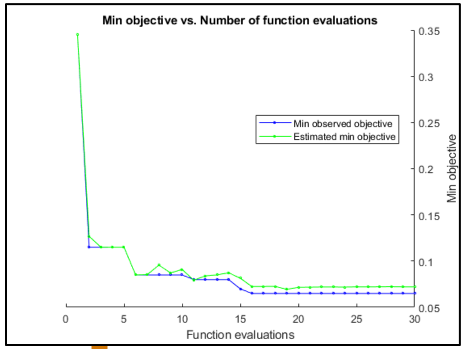
Resources, sessions, workshops and training available

Summary: Tools for building & deploying Machine Learning

Interactive Model Training

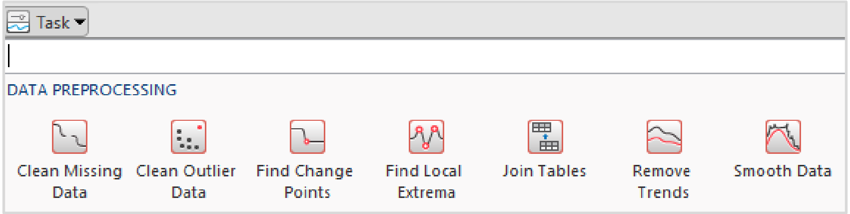


Automated Model Optimization

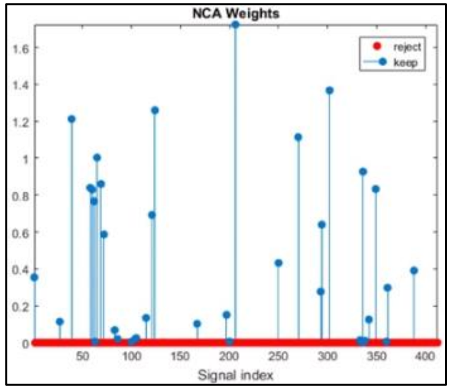


Import Data → Preprocess Data → Engineer Features → Build Model → Deploy & Integrate

Clean Messy Raw Data



Feature Selection

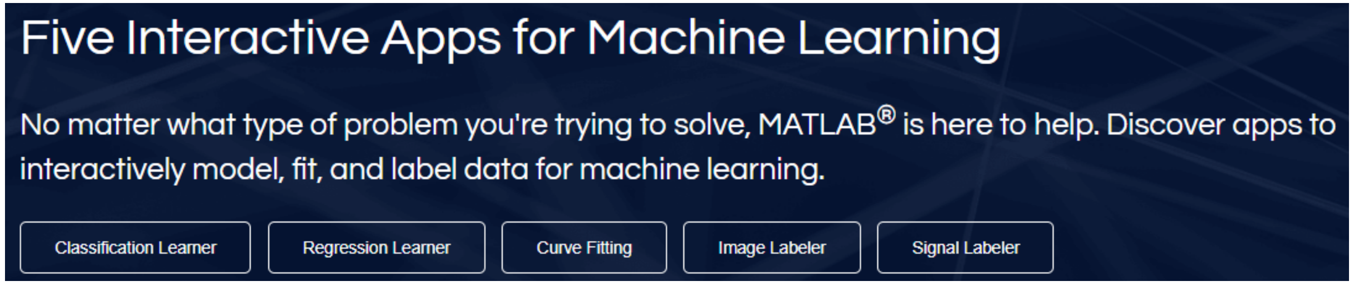


Automated Code Generation

Resources

Machine Learning Onramp (2 hr online introduction)

Practical Data Science with MATLAB (4 course Specialization)



Machine Learning with MATLAB:

- Overview, Cheat sheet
- Machine Learning Intro (Tech talks)
- Machine Learning with MATLAB Introduction (eBook)
- Mastering Machine Learning (eBook)
- Applied Machine Learning (Tech Talk videos)

Machine and Deep Learning

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

MathWorks can help you apply Machine Learning

Training

Guided Evaluations

Onsite Workshops

Consulting

Technical Support

Additional Material

- Working with Data in MATLAB
- Reinforcement Learning Example
- Explain different accuracies
- Alternative for Wavelet Scattering, drawing comparison to CNN filters

Working with Data

Business and Transactional Data

Repositories
- Databases (SQL/NoSQL)
- Hadoop

File I/O
- Text
- Spreadsheet

Web Sources
- RESTful/SOAP
- JSON
- HTML/XML
- Mapping
- Financial datafeeds
- FTP

Recent Additions

File I/O
- PDF
- Microsoft Word
- Parquet
- Vector BLF
- STL (Stereolithography)

Web Sources
- Amazon Web Services
- Azure Blob Storage

Internet of Things (IOT)
- ThingSpeak

Engineering, Scientific and Field Data

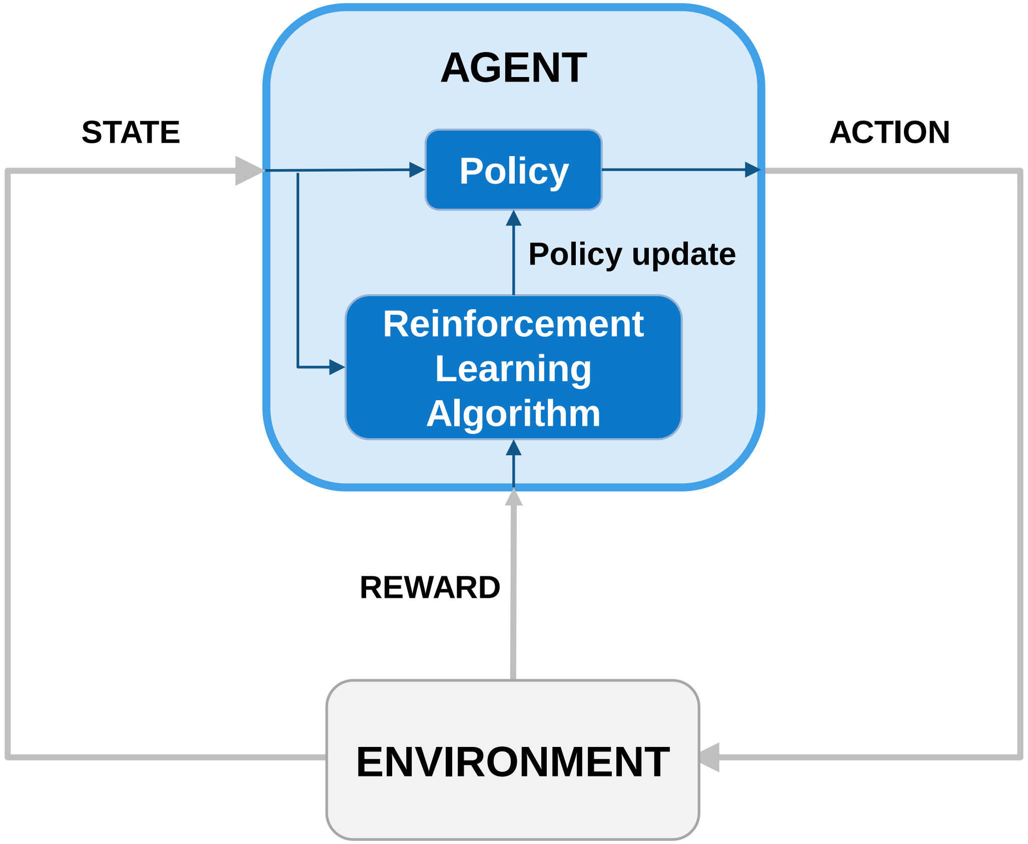
File I/O
- CDF/HDF
- Audio/Image/ Video
- Geospatial
- Microarrays
- CAD Models
- MDF

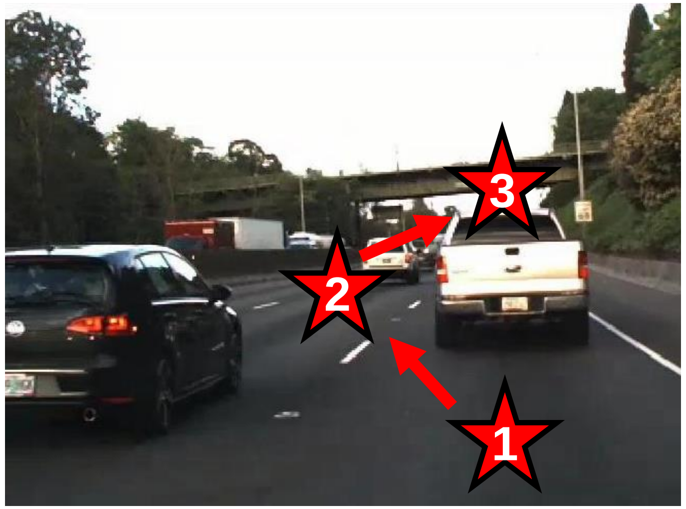
Communication Protocols
- CAN (Controller Area Network)
- DDS (Data Distribution Service)
- OPC (OLE for Process Control) (e.g. PI)
- XCP (eXplicit Control Protocol)
- TCP/IP
- Serial/Bluetooth/USB

Real-Time Sources
- Sensors/Instrumentation/Cameras
- GPS
- Communication systems
- Machines (embedded systems)
- Robot Operating System (ROS)

The above list is not all-inclusive, but is intended for guidance only

A Practical Example of Reinforcement Learning
Training a Self-Driving Car

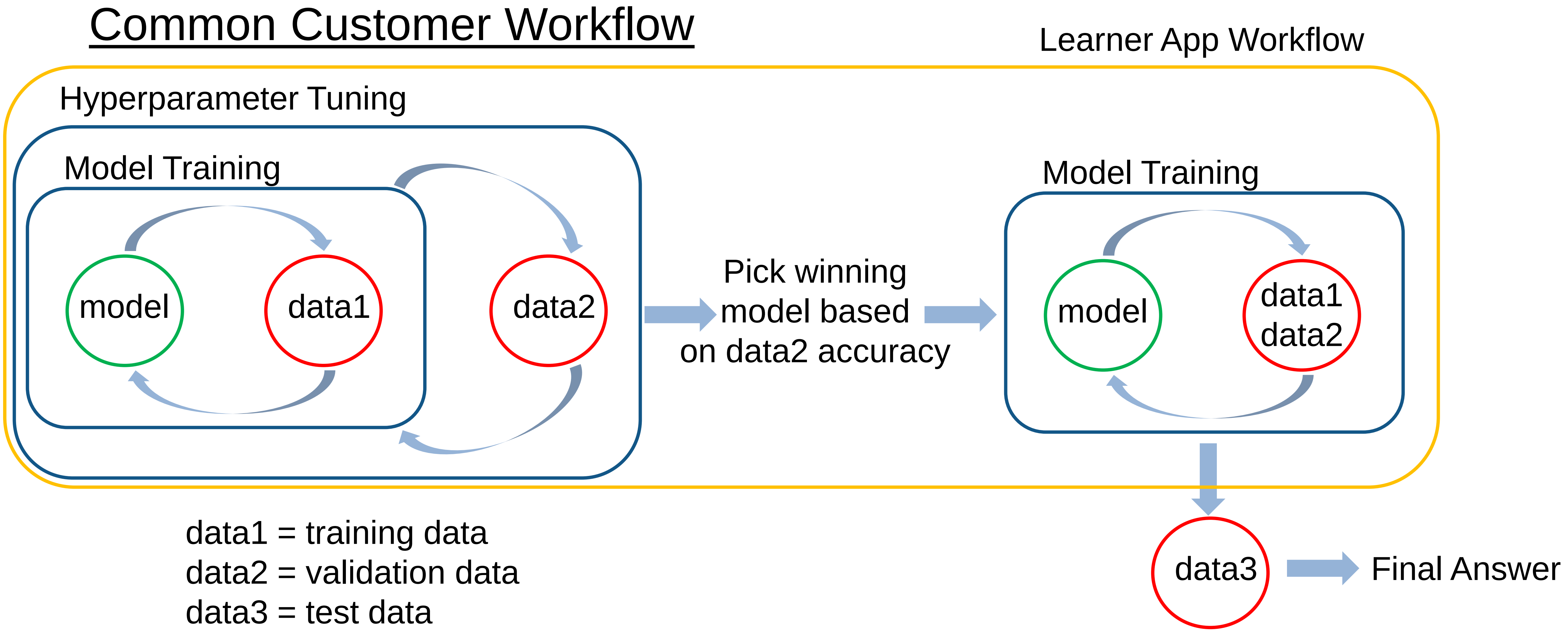




Vehicle's computer learns how to drive… (**agent**)
using sensor readings from LIDAR, cameras,… (**state**)
that represent road conditions, vehicle position,… (**environment**)
by generating steering, braking, throttle commands,… (**action**)
based on an internal state-to-action mapping… (**policy**)
that tries to optimize driver comfort & fuel efficiency… (**reward**).

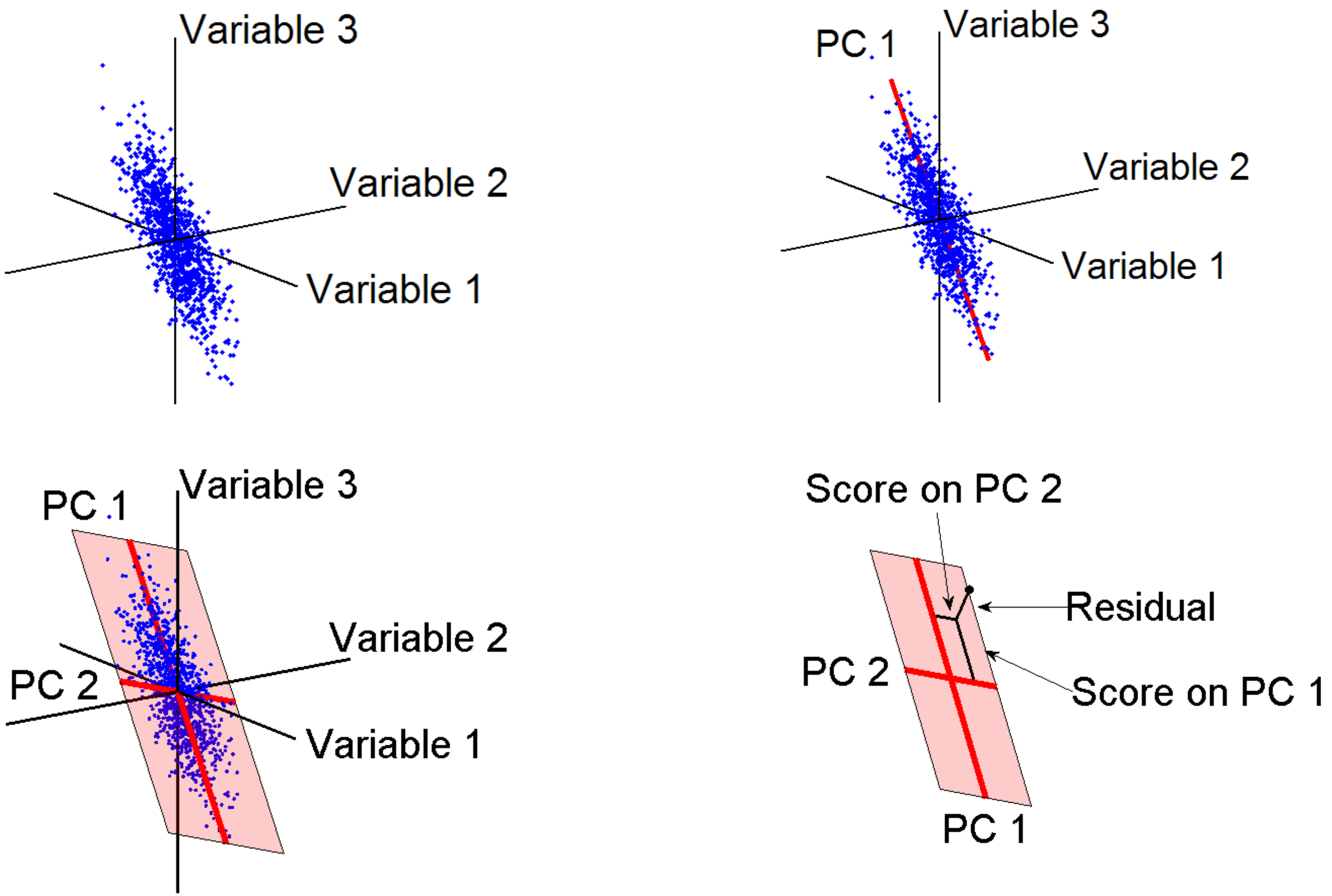
The policy is updated through repeated trial-and-error by a **reinforcement learning algorithm**

Evaluating Model Accuracy, properly.

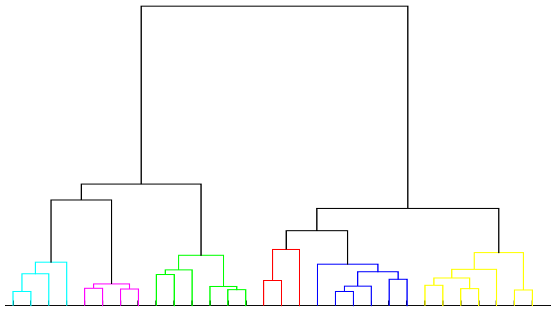


data1 = training data
data2 = validation data
data3 = test data

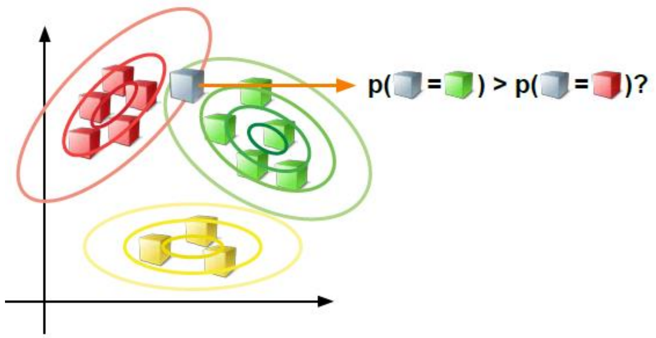
Principal Components Analysis (PCA)



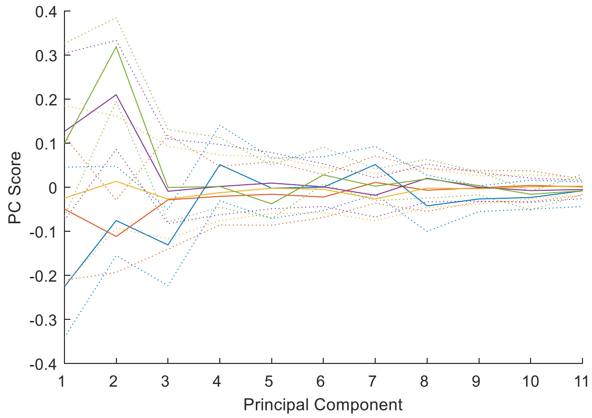
Training: *Machine Learning with MATLAB*



After this 2-day course you will be able to:



- Discover natural patterns in data
- Create predictive models
- Validate the predictions of a model
- Simplify and improve models



http://www.mathworks.com/training-schedule/machine-learning-with-matlab

Training: *Processing Big Data with MATLAB*

In this 1-day course, you will apply your data science code to **big** data

- Overcome physical memory limitations.
- Work with many gigabytes or terabytes of data.
- Access data and process it on the cloud.

Out of memory processing

Datastore to represent data | Tall Array on multiple cores | Tall Array on cluster / cloud

https://www.mathworks.com/training-schedule/processing-big-data-with-matlab.html