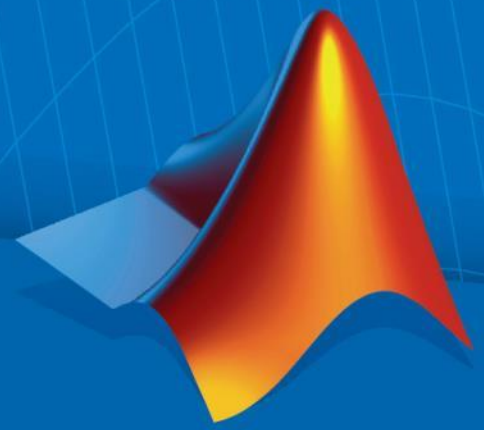
TeraSoft
www.terasoft.com.tw



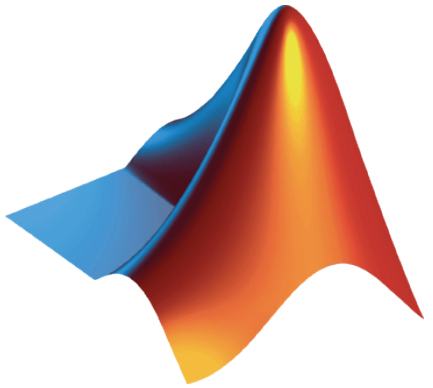
Building Predictive Models for Sensor Data Analytics

Terasoft
Application Engineer
Jeffrey Liu

MathWorks®
Accelerating the pace of engineering and science

Agenda

- Data Analytics Overview
- Machine Learning
  - Introduction
  - Example: Human activity learning
- Deep Learning
  - Introduction
  - Example: Speech Command Recognition
- Wrap-up and Q&A



What is Data Analytics?

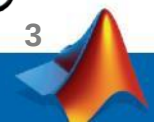
Turn large volumes of complex data into actionable information

Data →

Descriptive	• What happened?
Diagnostic	• Why did it happen?
Predictive	• What will happen?
Prescriptive	• What should be done?

→ **Decisions**

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

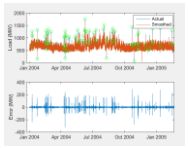


Data Analytics Workflow

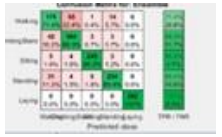
Access and Explore Data

- **Files**
  
- **Databases**
  
- **Sensors**
  

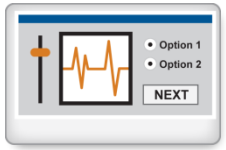
Preprocess Data

- **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**
  MATLAB Excel .NET .exe C/C++ Java .dll
- **Embedded Devices and Hardware**
  

Data Sources

Business and Transactional Data

Repositories
- Databases (SQL)
- NoSQL
- Hadoop

File I/O
- Text
- Spreadsheet
- XML

Web Sources
- RESTful
- JSON
- HTML
- Mapping
- Financial datafeeds

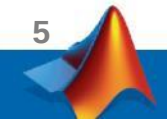
Engineering, Scientific, and Field Data

File I/O
- Text
- Spreadsheet
- XML
- CDF/HDF
- Image
- Audio
- Video
- Geospatial

Communication Protocols
- CAN (Controller Area Network)
- DDS (Data Distribution Service)
- OPC (OLE for Process Control) – Pi Servers
- XCP (eXplicit Control Protocol)

Real-Time Sources
- Sensors
- GPS
- Instrumentation
- Cameras
- Communication systems
- Machines (embedded systems)

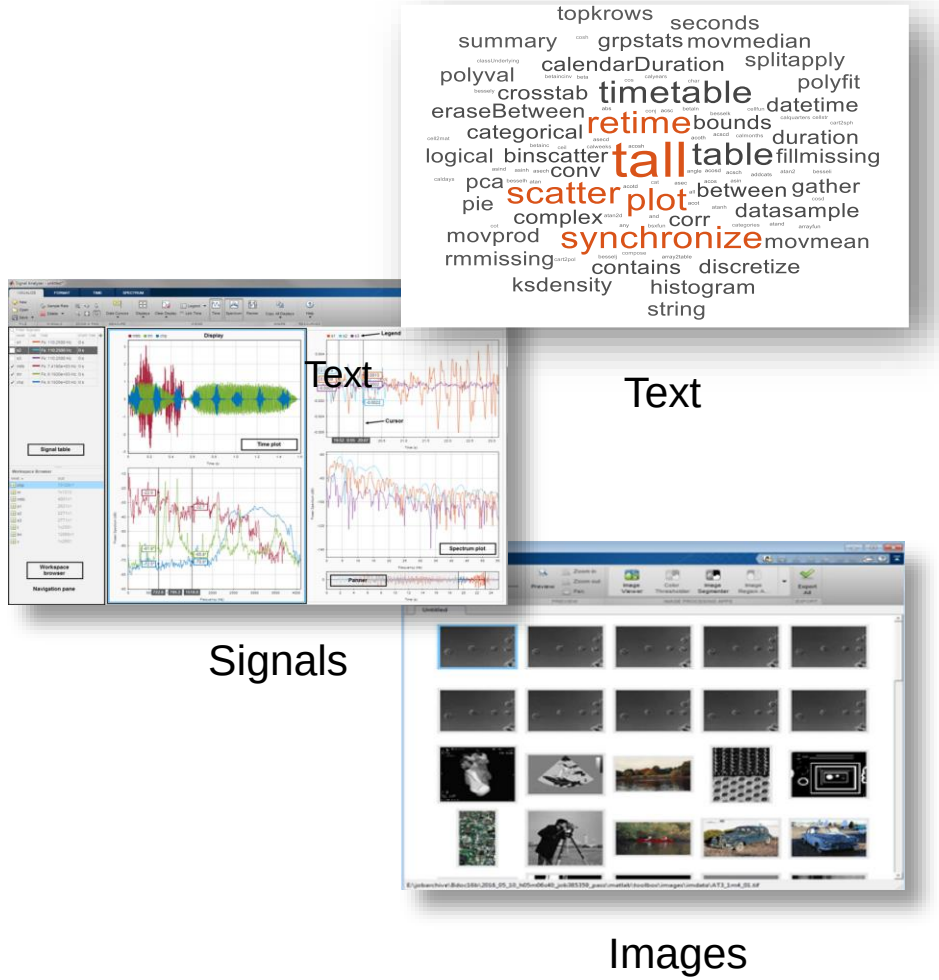
Custom Applications



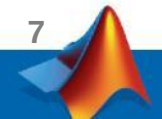
Feature Engineering

- Statistics and Machine Learning Toolbox
- Signal Processing Toolbox
- Computer Vision Toolbox
- Audio System Toolbox
- Wavelet Toolbox
- Text Analytics Toolbox

- … and more!

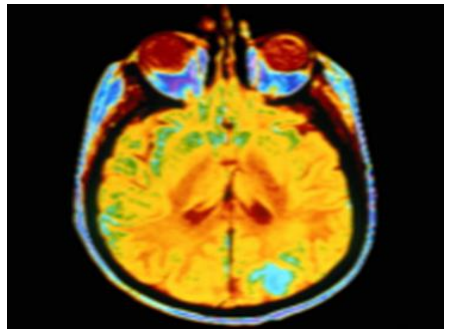


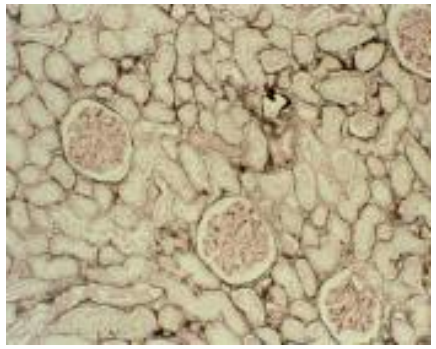
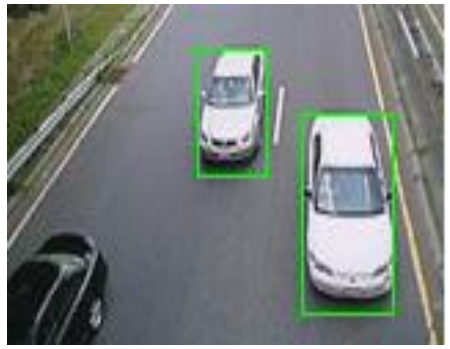
Machine Learning



Machine Learning is Everywhere

- Image Recognition
- Speech Recognition
- Stock Prediction
- Medical Diagnosis
- Data Analytics
- Robotics
- and more…

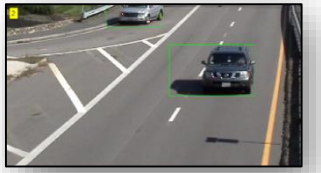


What is Machine Learning?

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

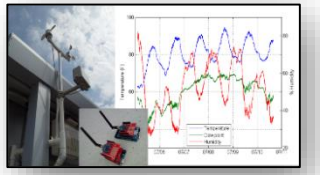

Weather Forecasting


Energy Load Forecasting

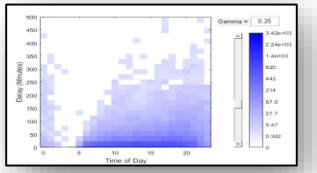

Stock Market Prediction

update as more data becomes available

Solution needs to scale


IoT Analytics

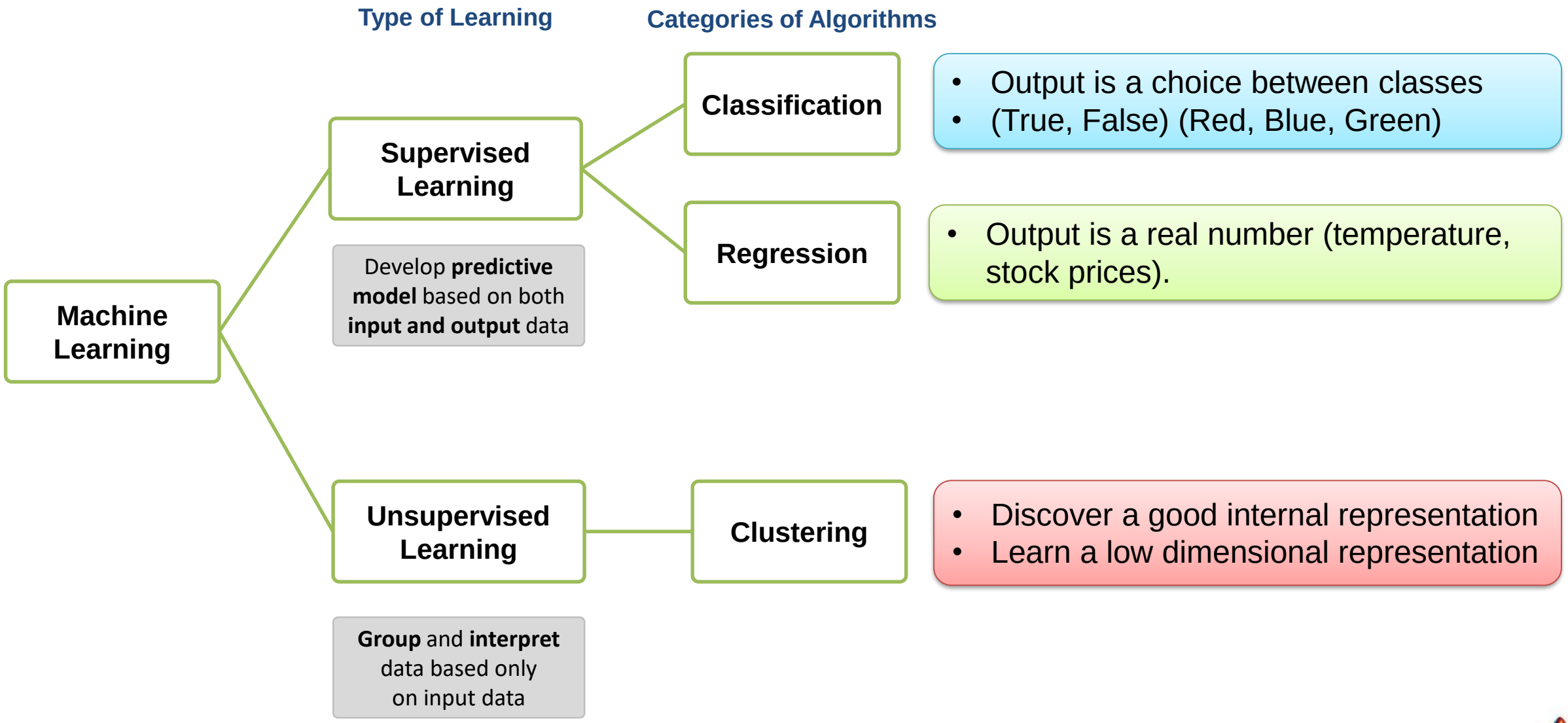
Taxi Availability



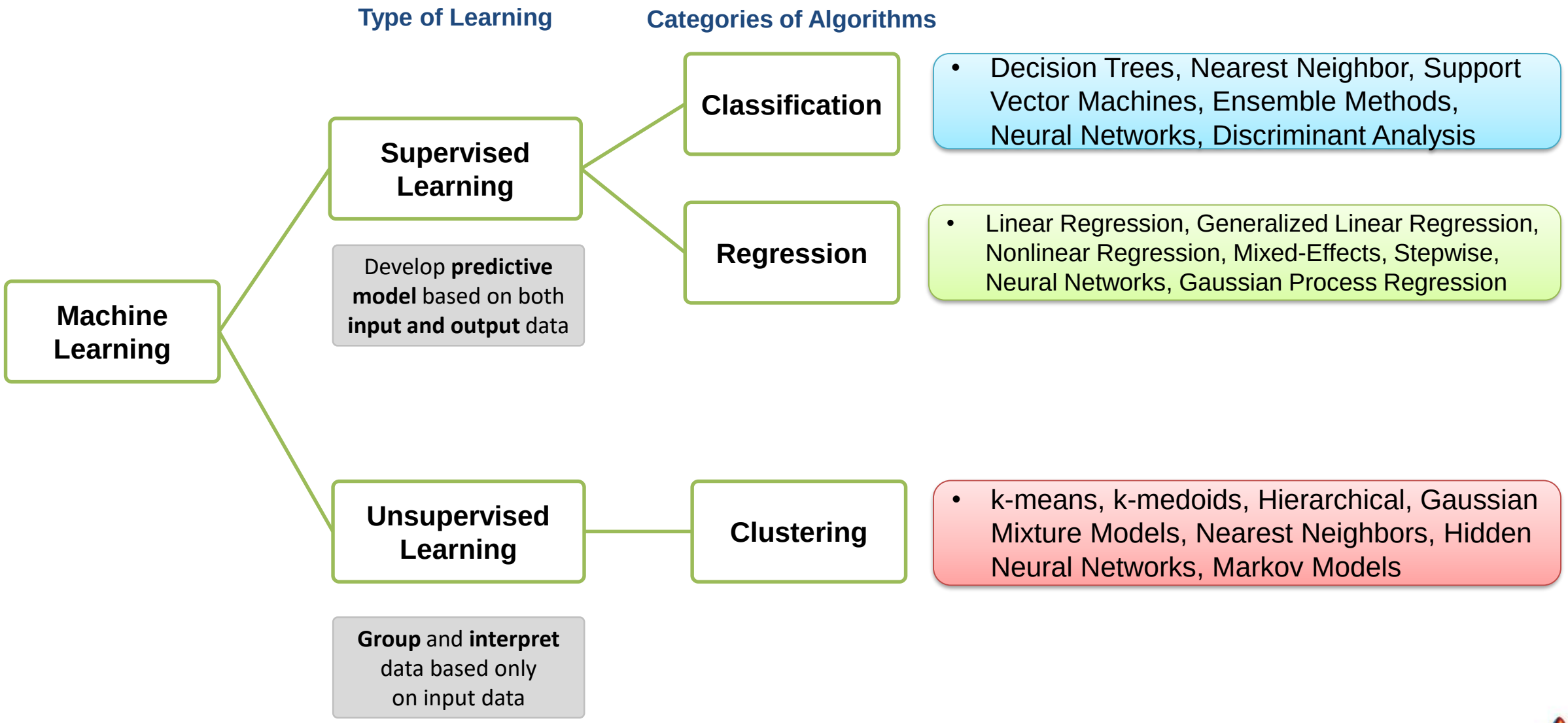
Airline Flight Delays

learn efficiently from very large data sets

Types of Machine Learning

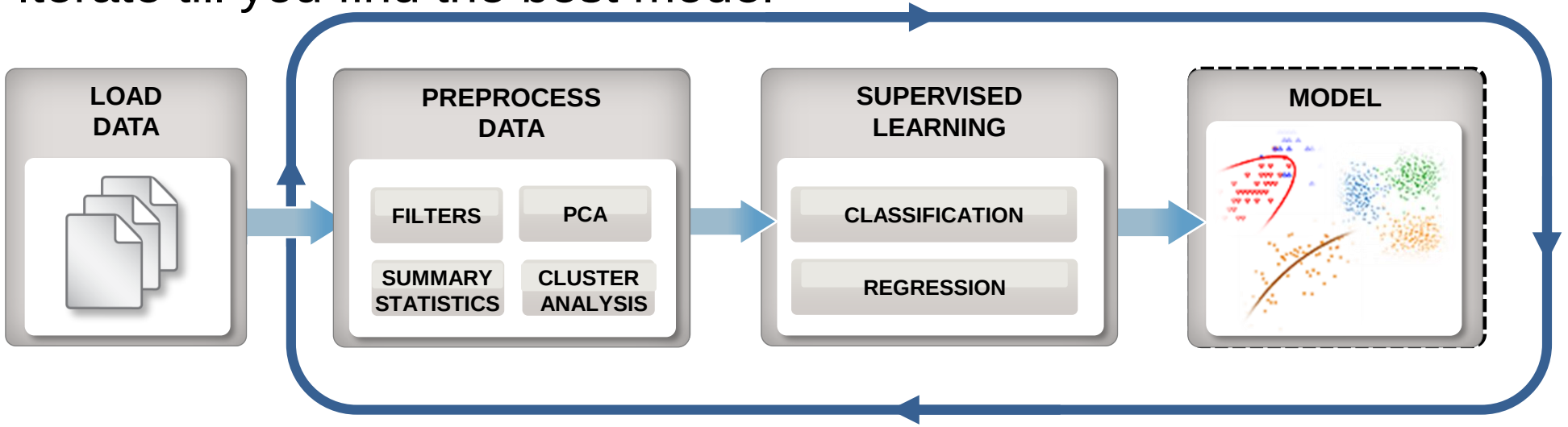


Types of Machine Learning

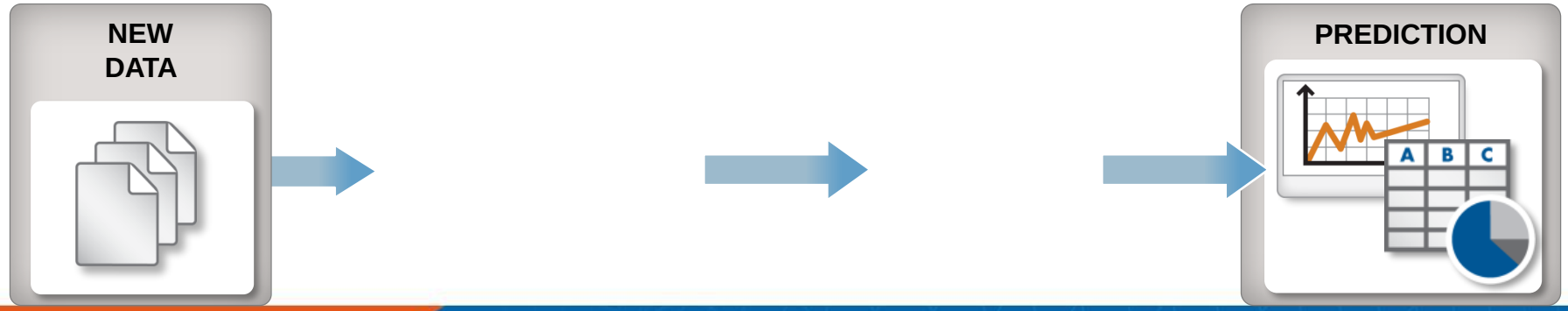


Machine Learning Workflow

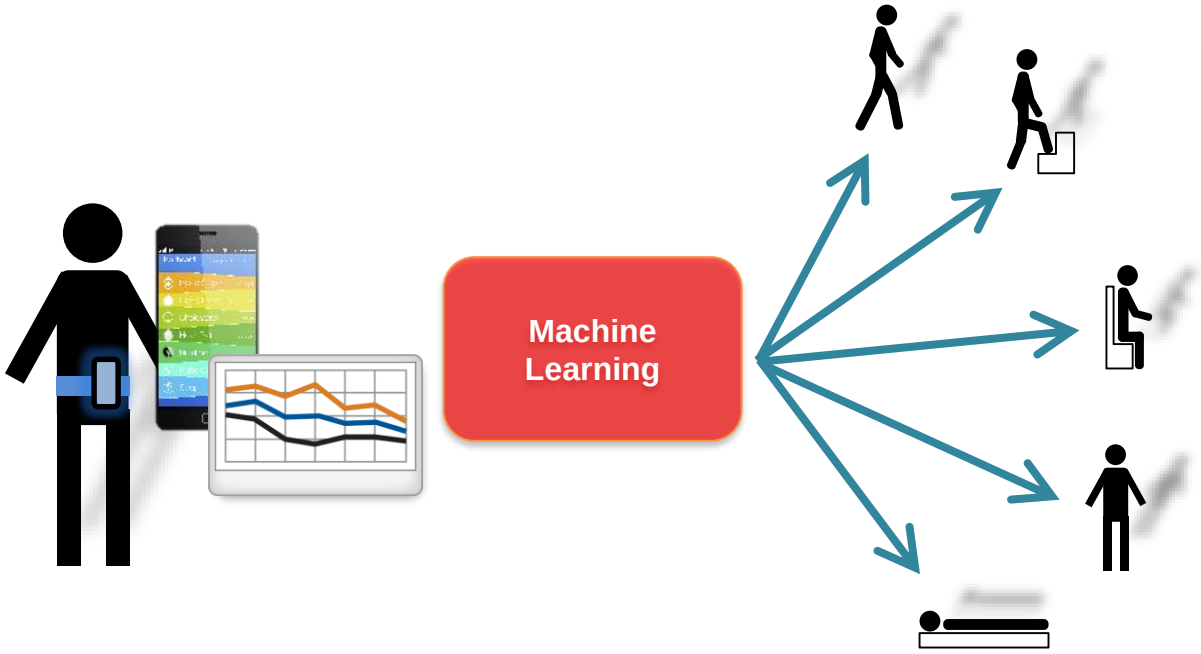
Train: Iterate till you find the best model

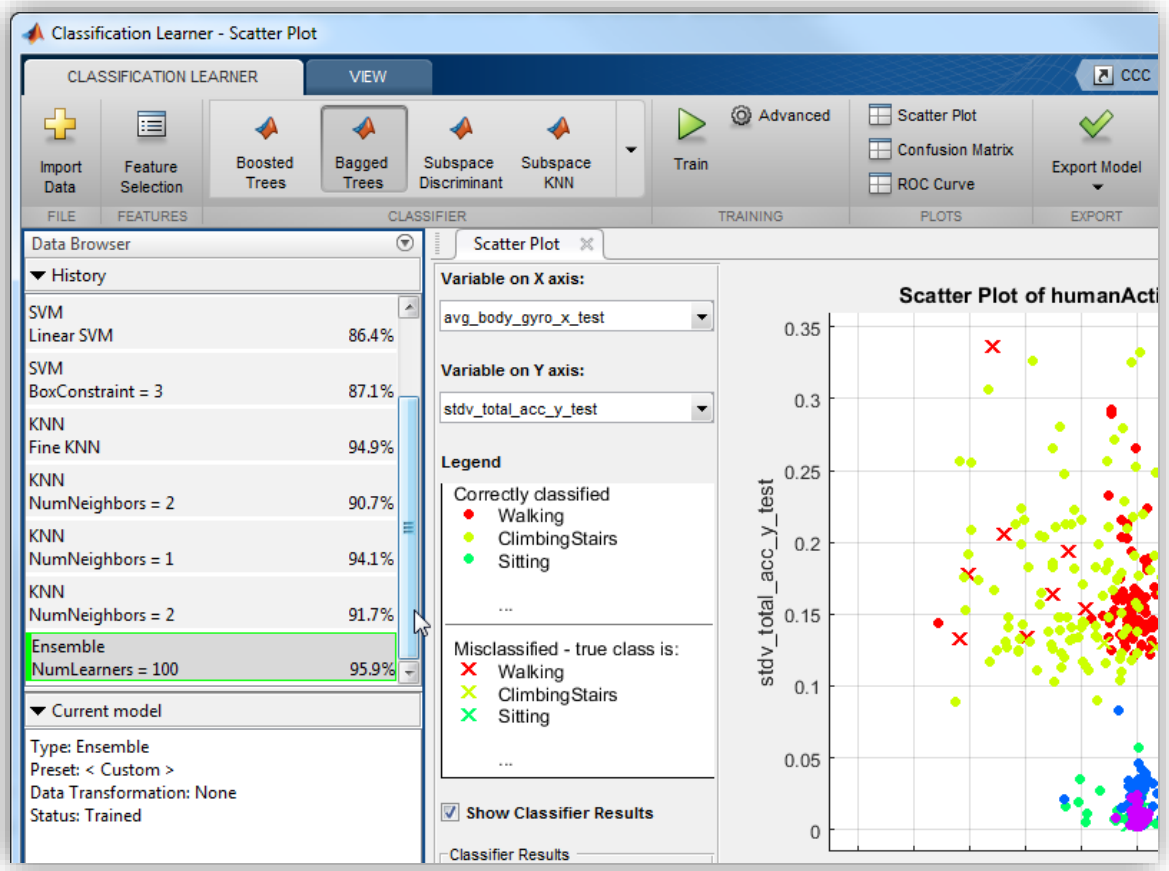


Predict: Integrate trained models into applications



Example: Human Activity Learning Using Mobile Phone Data





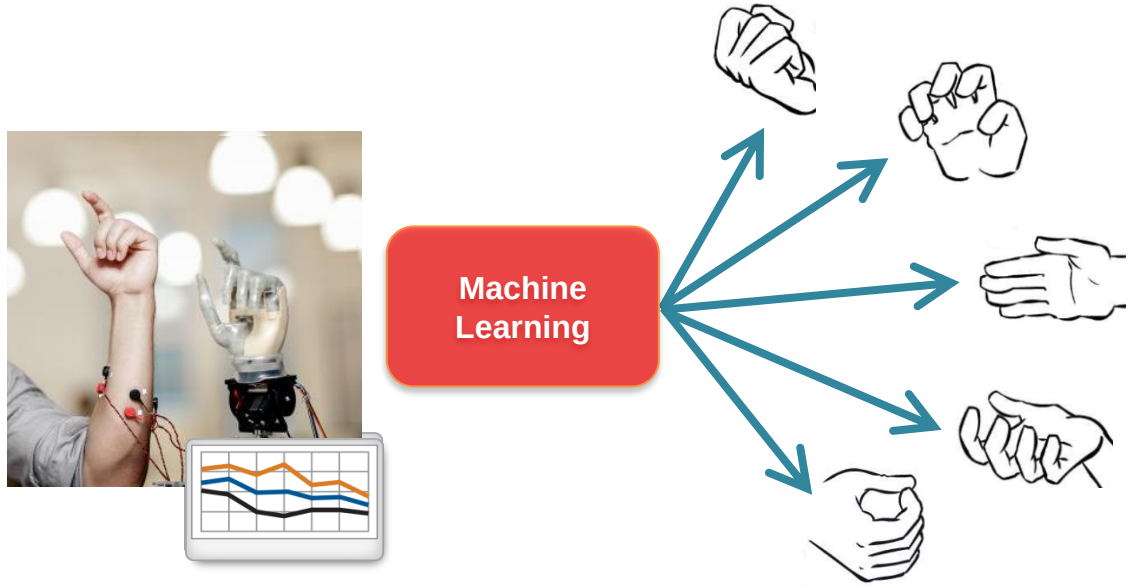
Data:

- 3-axial Accelerometer data
- 3-axial Gyroscope data

Example: Classify Limb Movements using EMG Data

Machine Learning Workflow

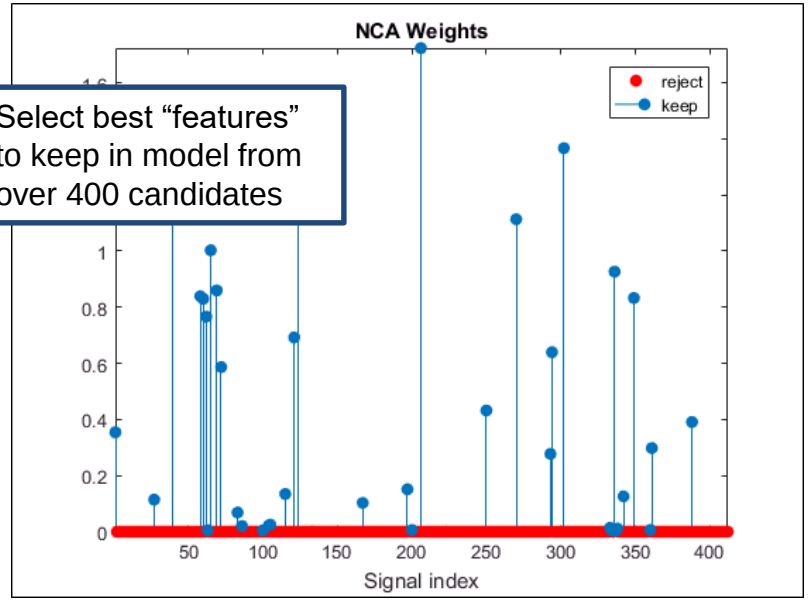
- Goal:
  - Develop a classifier to identify type of movement from EMG signal for Prosthetics Development
- Approach:
  - Access multi-dimensional data from sensors
  - Interactively visualize and explore trends
  - Create a model
  - Document results



Tuning Machine Learning Models

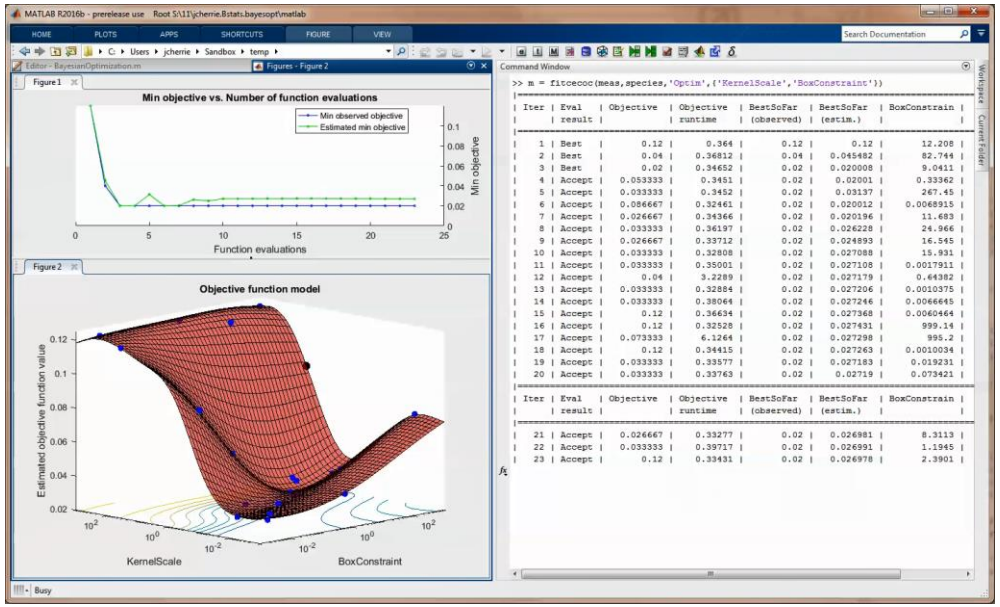
Get more accurate models in less time

Automatically select best machine leaning “features”



NCA: Neighborhood Component Analysis

Automatically fine-tune machine learning parameters



Hyperparameter Tuning

Text Analytics

- Sources of Text Data
  - Maintenance Logs
  - News/Social Media
  - Customer Surveys
  - Field Reports

Split Text into Individual Words

```
documents = tokenizedDocument(repairNotes)
```

Create a Bag-of-Words Model

```
bag = bagOfWords(documents)
```

Fit a Topic Model with 4 Topics

```
numTopics = 4;
mdl = fitlda(bag,numTopics)
```

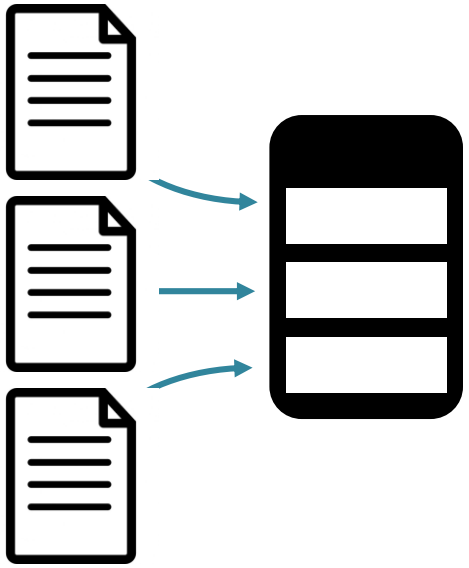
```
  (36,1) coolant leak spinner light outwip
  (37,1) strob lights not workinghyd leak

bag =
bagOfWords with 913 words and 617 documents

    preventative    maintenance    service
               1              1          1
               0              0          0
               …

mdl =
  ldaModel with properties:

            NumTopics: 4
    TopicConcentration: 1.3262
    TopicProbabilities: [4×1 double]
     WordConcentration: 1
            Vocabulary: [1×913 string]
```

- Applications
  - Sentiment Analysis: Determine if news about a product is positive/negative
  - Maintenance: Identify hidden groups of issues in maintenance logs
  - Document Classification: Tag unread documents (eg. for triaging, routing, etc.)

Text Analytics Toolbox

Access and Explore Data → **Preprocess Data** → **Develop Predictive Models**

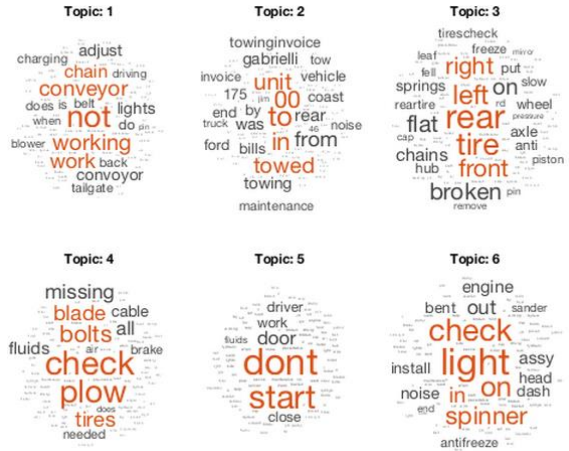
Clean-up Text → **Convert to Numeric**



Media reported two trees blown down along I-40 in the Old Fort area.

media report two tree blown down i40 old fort area

	cat	dog	run	two
doc1	1	0	1	0
doc2	1	1	0	1



- Word Docs
- PDF's
- Text Files

- Stop Words
- Stemming
- Tokenization

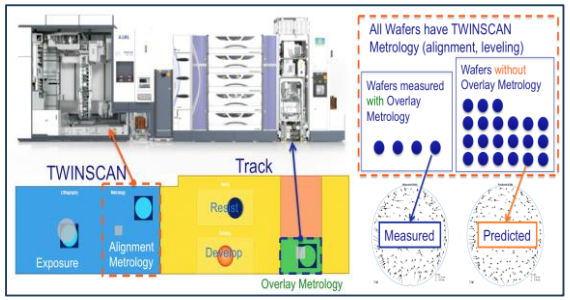
- Bag of Words
- TF-IDF
- Word Embeddings

- Latent Dirichlet Allocation
- Latent Semantic Analysis

ASML Develops Virtual Metrology Technology for Semiconductor Manufacturing with Machine Learning

Challenge

Apply machine learning techniques to improve overlay metrology in semiconductor manufacturing



Cutaway of a TWINSCAN and Track as wafers receive alignment and overlay metrology

Solution

Use MATLAB to create and train a neural network that predicts overlay metrology from alignment metrology

Results

- Industry leadership established
- Potential manufacturing improvements identified
- Maintenance overhead minimized

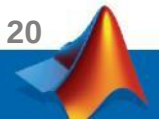
"As a process engineer I had no experience with neural networks or machine learning. I worked through the MATLAB examples to find the best machine learning functions for generating virtual metrology. I couldn't have done this in C or Python—it would've taken too long to find, validate, and integrate the right packages."

Emil Schmitt-Weaver

ASML

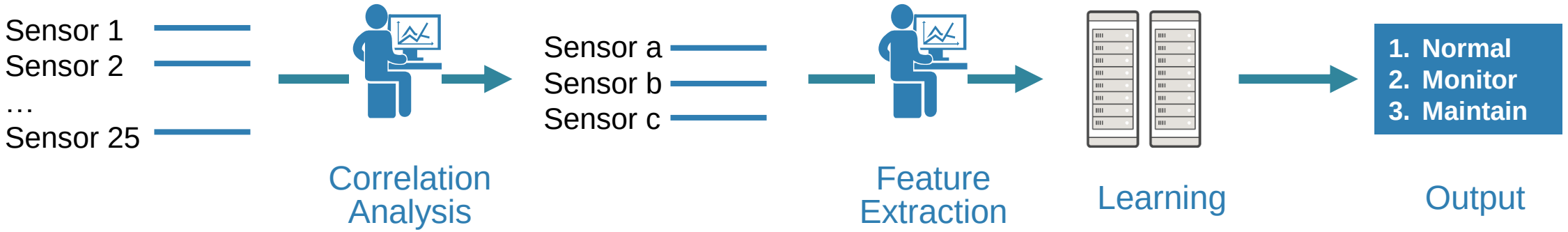
Link to user story

Deep Learning

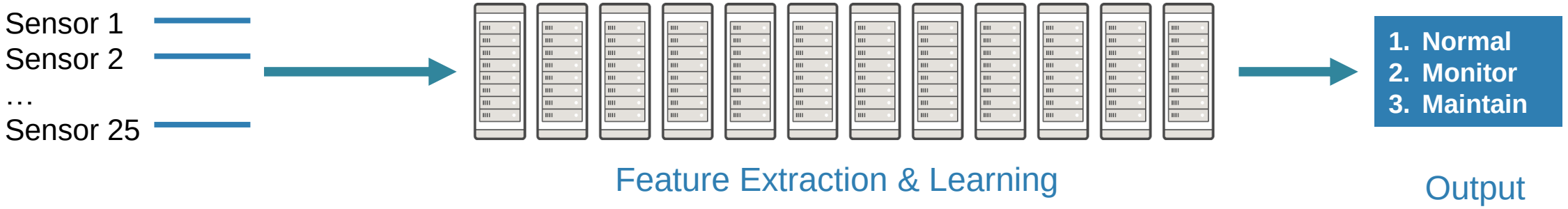


Machine Learning or Deep Learning?

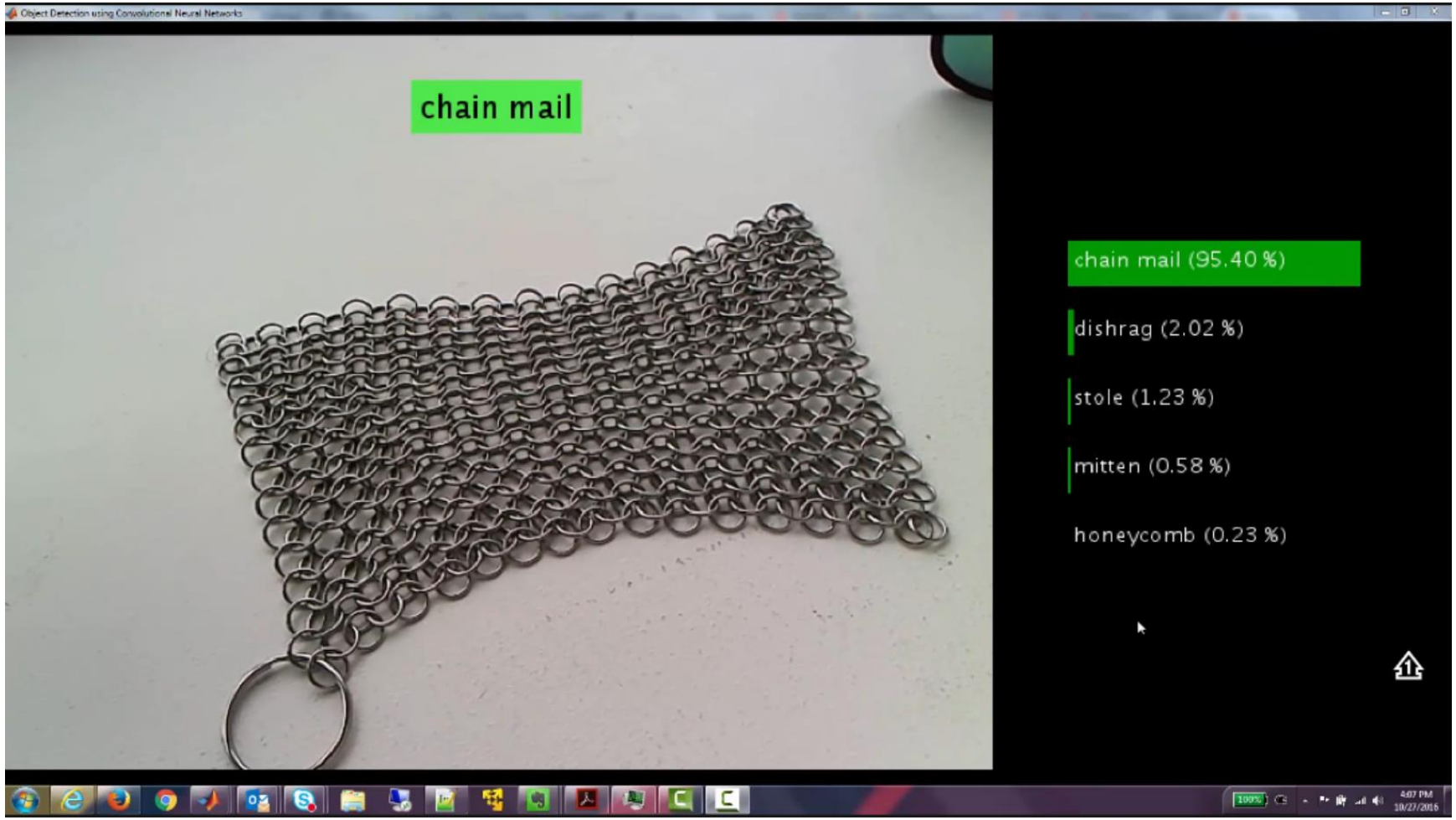
Machine Learning Approach



Deep Learning Approach

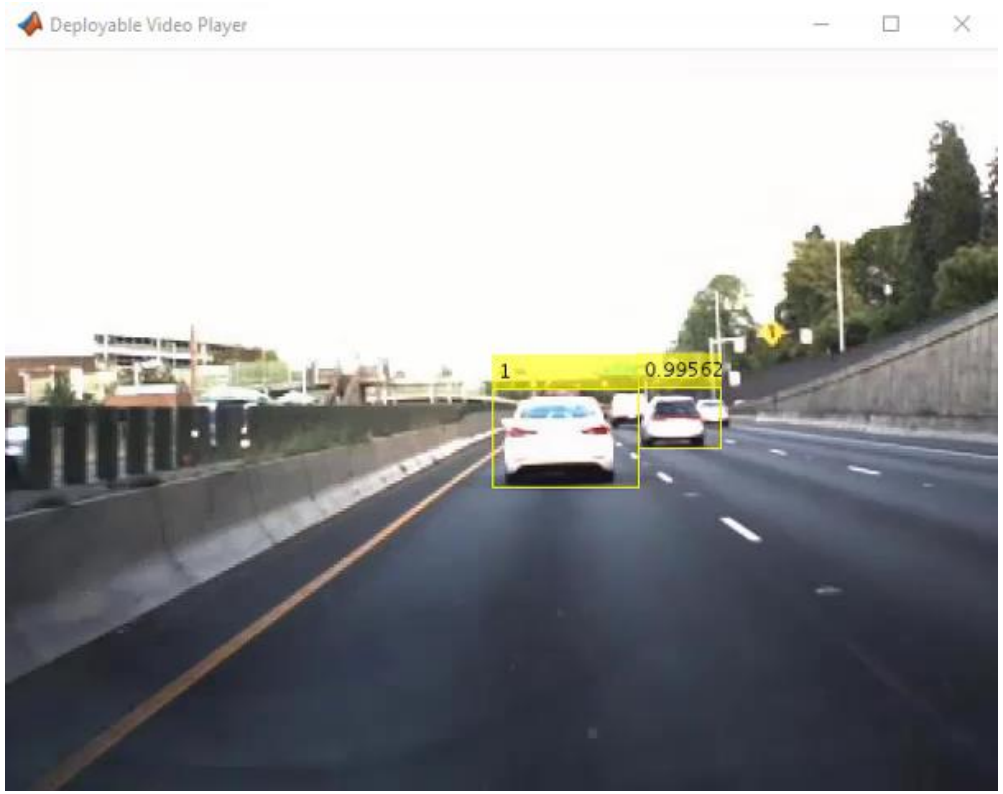


Object recognition using deep learning



Training (GPU)	Millions of images from 1000 different categories
Prediction	Real-time object recognition using a webcam connected to a laptop

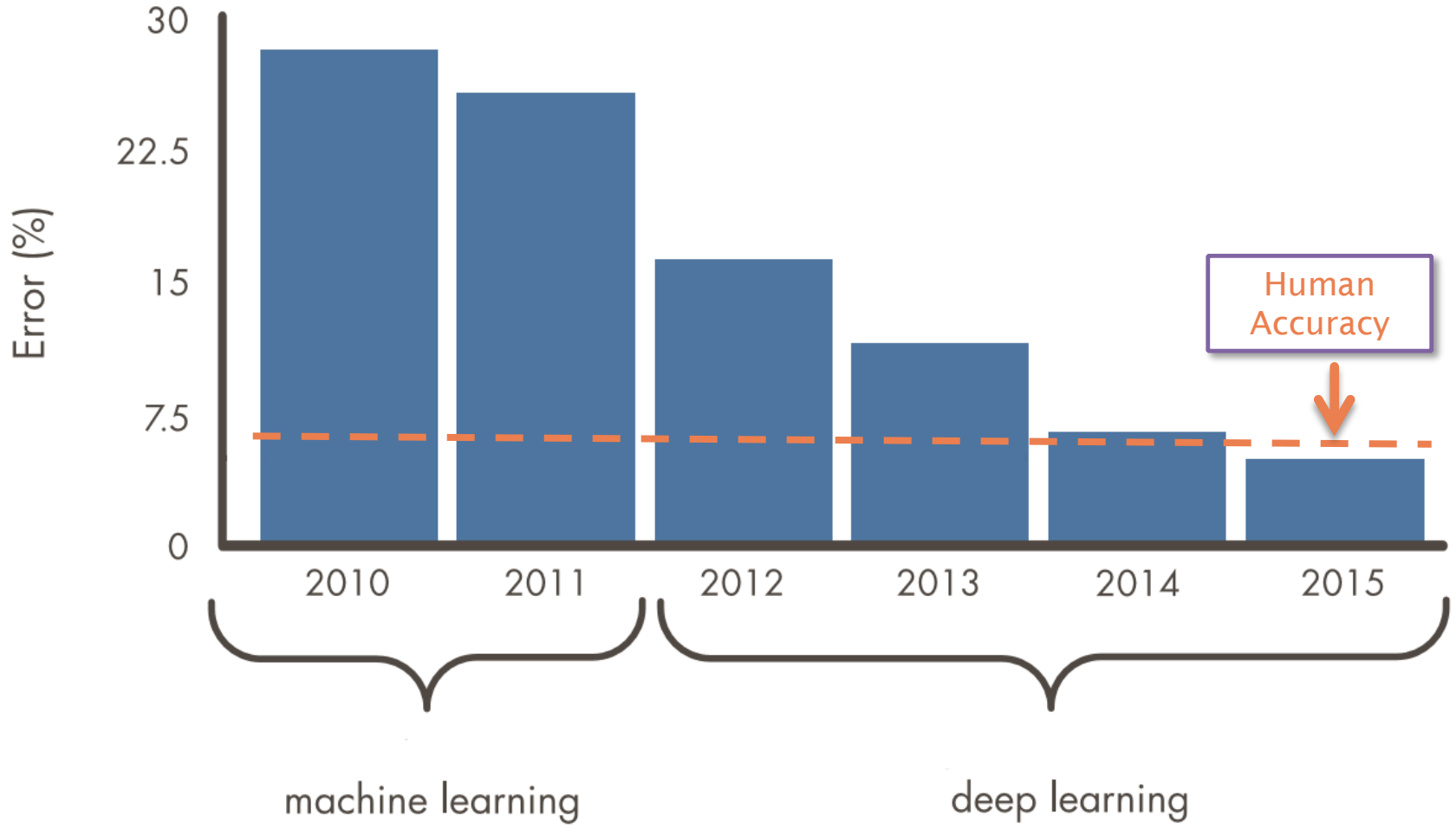
Detection and localization using deep learning



Regions with Convolutional Neural Network Features (R-CNN)

Semantic Segmentation using SegNet

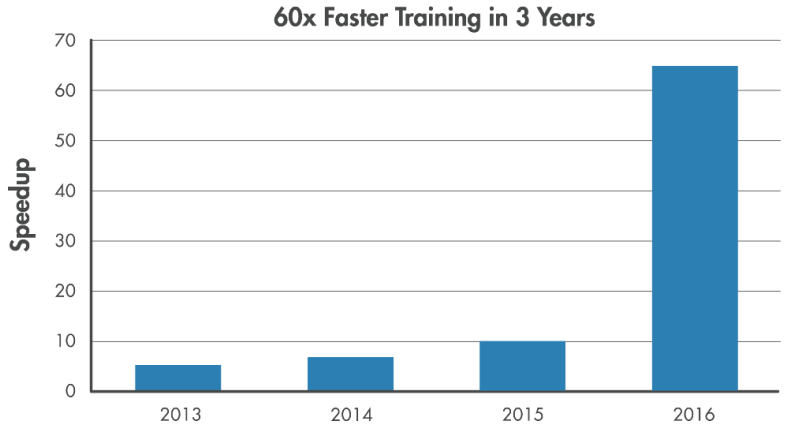
Why is Deep Learning So Popular Now?



Source: ILSVRC Top-5 Error on ImageNet

Deep Learning Enablers

- Increased GPU acceleration



- World-class models

AlexNet
PRETRAINED MODEL

VGG-16
PRETRAINED MODEL

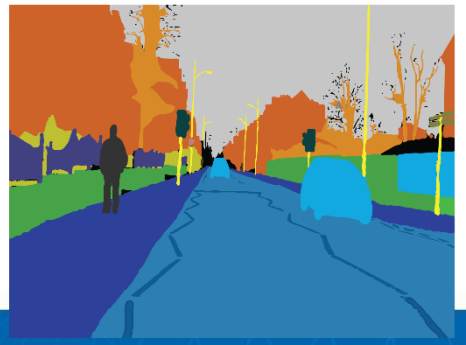
ResNet
PRETRAINED MODEL

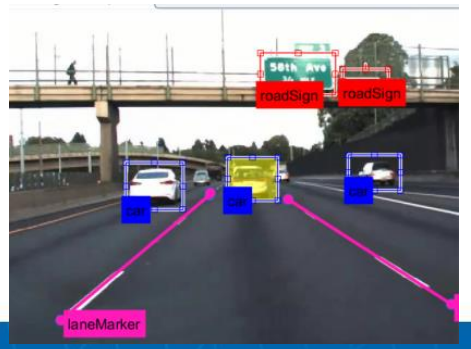
Caffe
MODELS

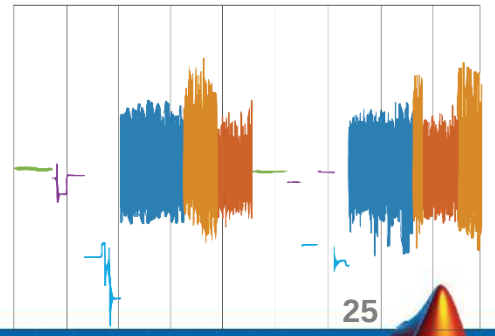
GoogLeNet
PRETRAINED MODEL

TensorFlow/Keras
MODELS

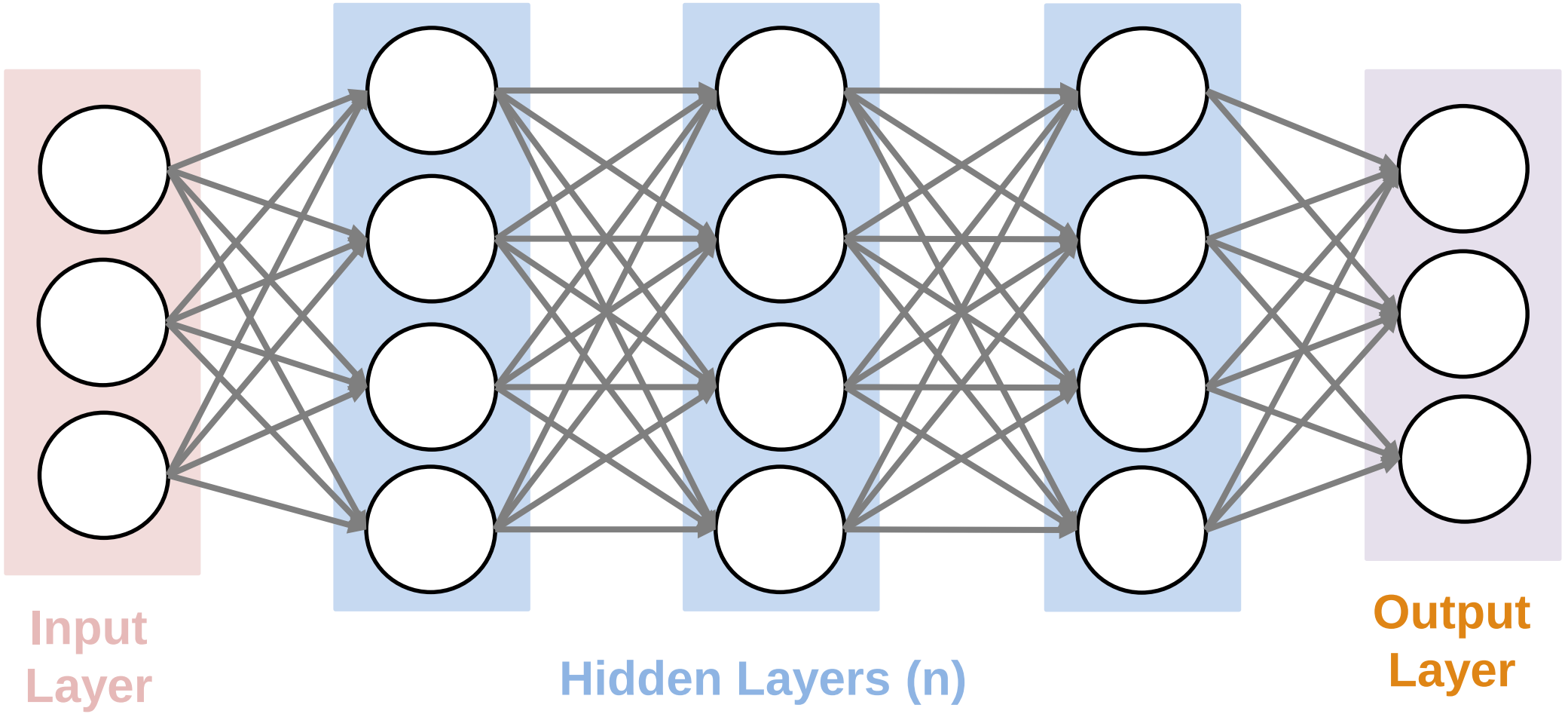
- Labeled public datasets





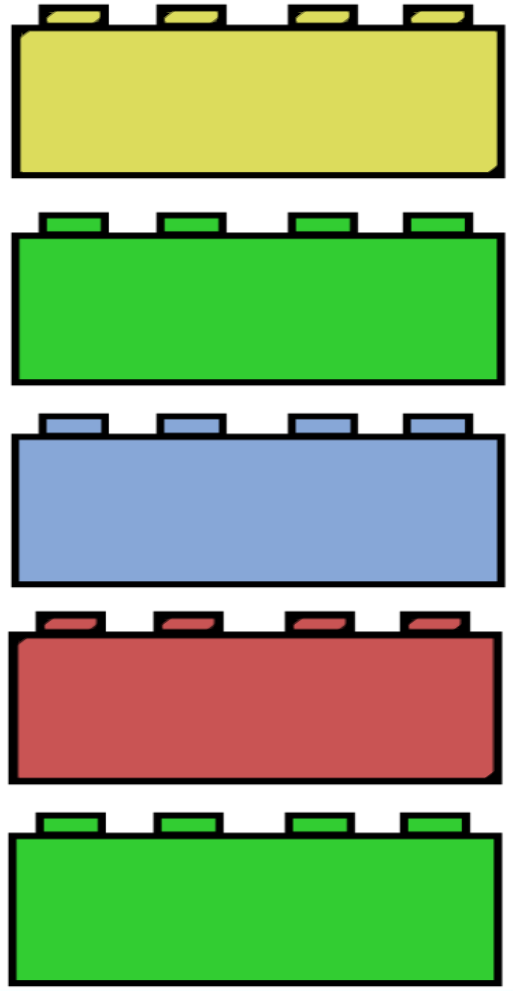


Deep Learning Uses a Neural Network Architecture



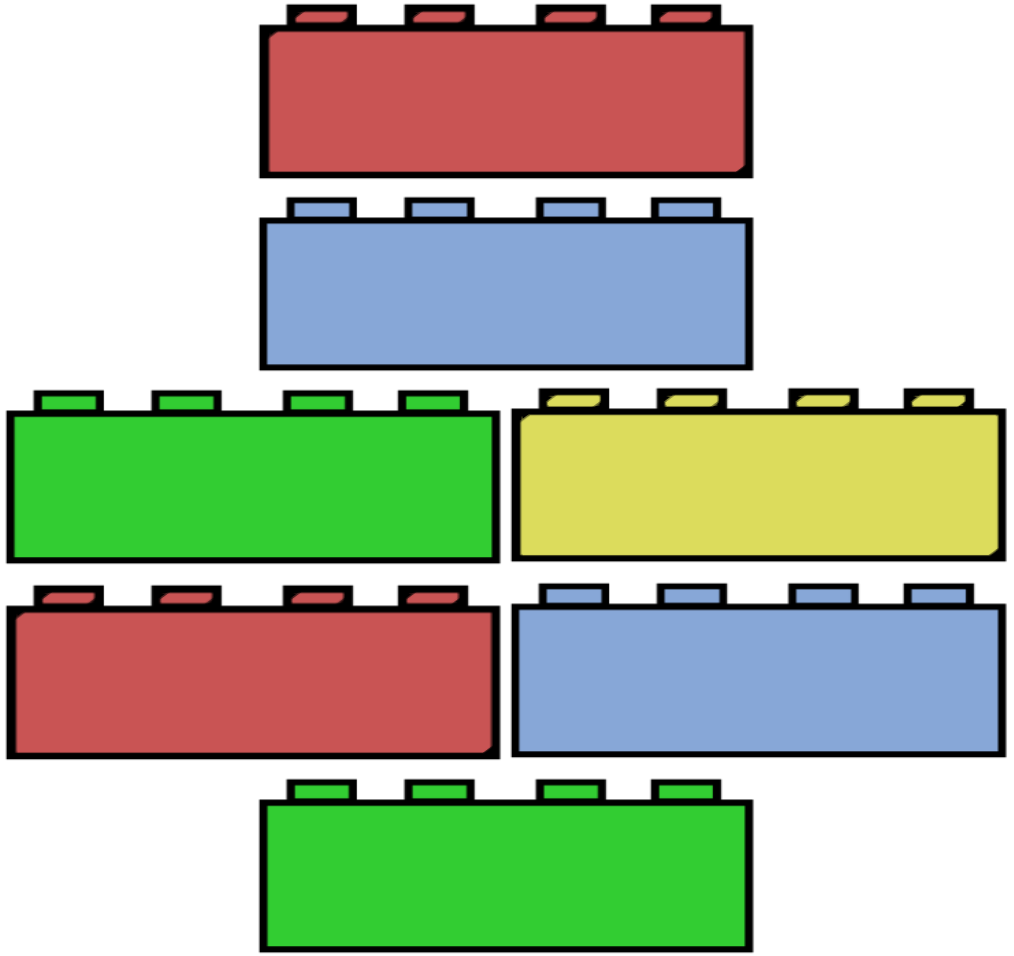
Thinking about Layers

- Layers are like blocks
  - Stack on top of each other
  - Replace one block with a different one
- Each hidden layer processes the information from the previous layer



Thinking about Layers

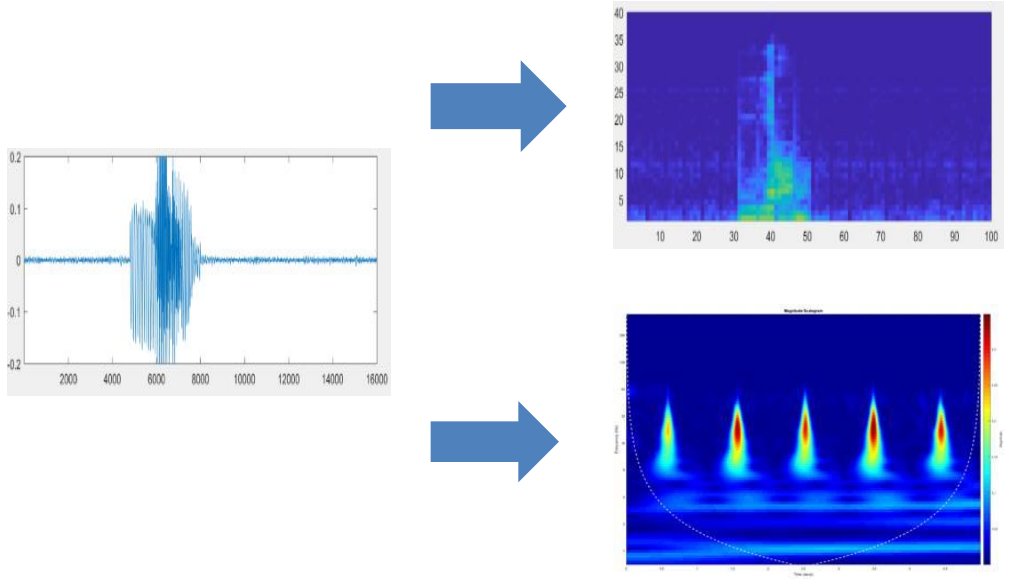
- Layers are like blocks
  - Stack on top of each other
  - Replace one block with a different one
- Each hidden layer processes the information from the previous layer
- Layers can be ordered in different ways



Deep Learning for Signals (Non Image)

Convolutional Neural Networks (CNN)

- Convert the time series into a spectrogram/scalogram and use it as an input to a CNN.



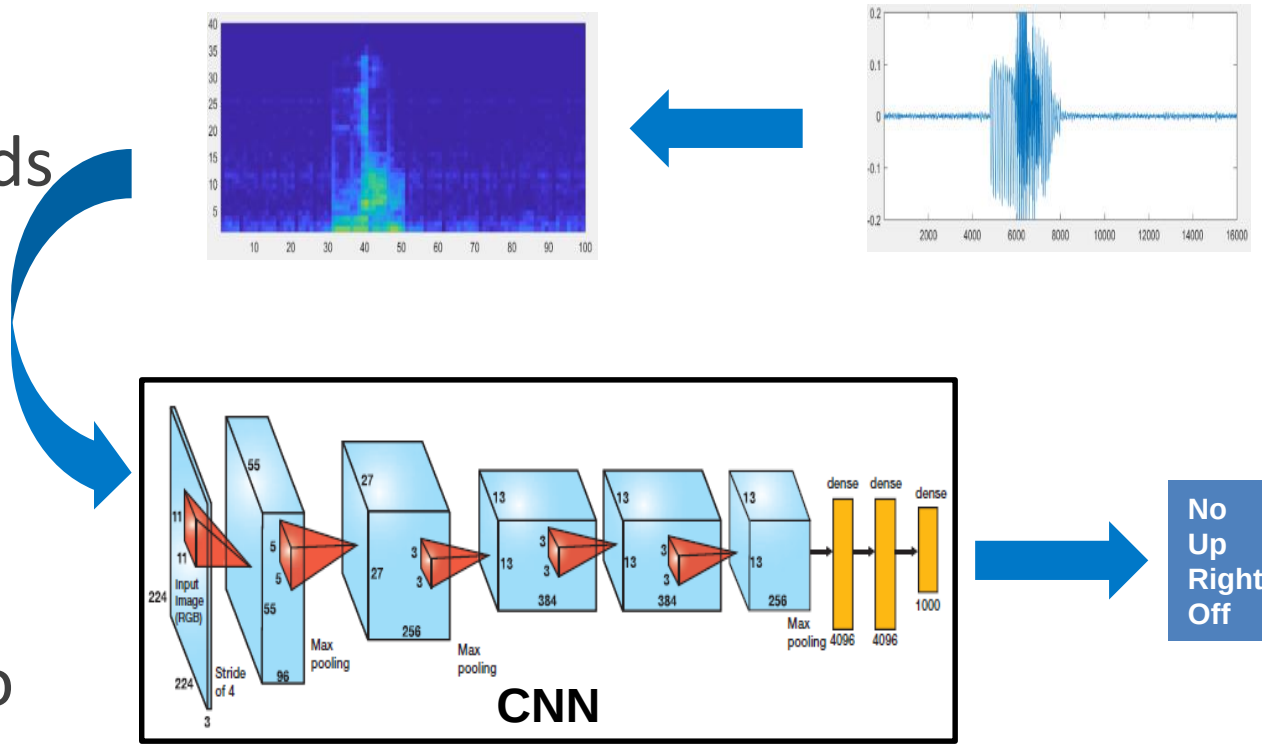
Long Short-Term Memory (LSTM)

- Neural networks which **connect through time.**

Demo: Speech Command Recognition

Goal:

- Use deep learning techniques to recognize a given set of commands in audio signals.

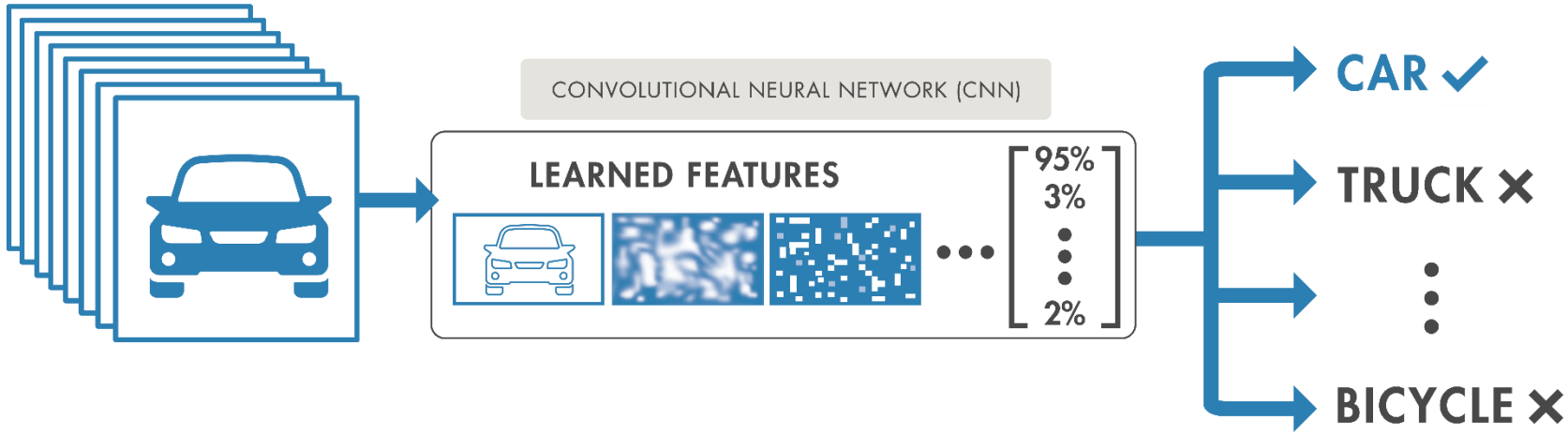


Approach

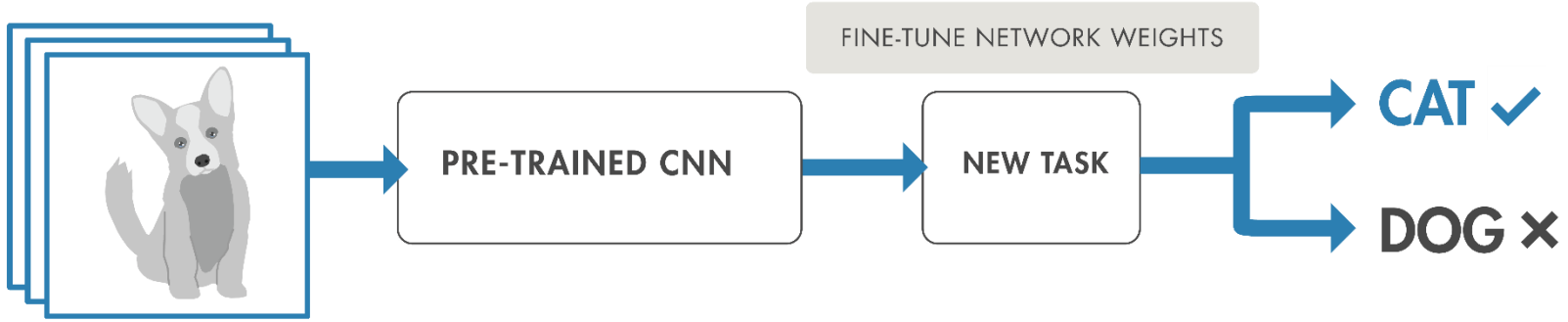
- Load audio signals
- Convert the speech waveforms to log-bark auditory spectrograms.
- Train a CNN to detect presence of speech commands in audio

Two Approaches for Deep Learning

1. Train a Deep Neural Network from Scratch



2. Fine-tune a pre-trained model (transfer learning)



Import the Latest Models for Transfer Learning

Pretrained Models*

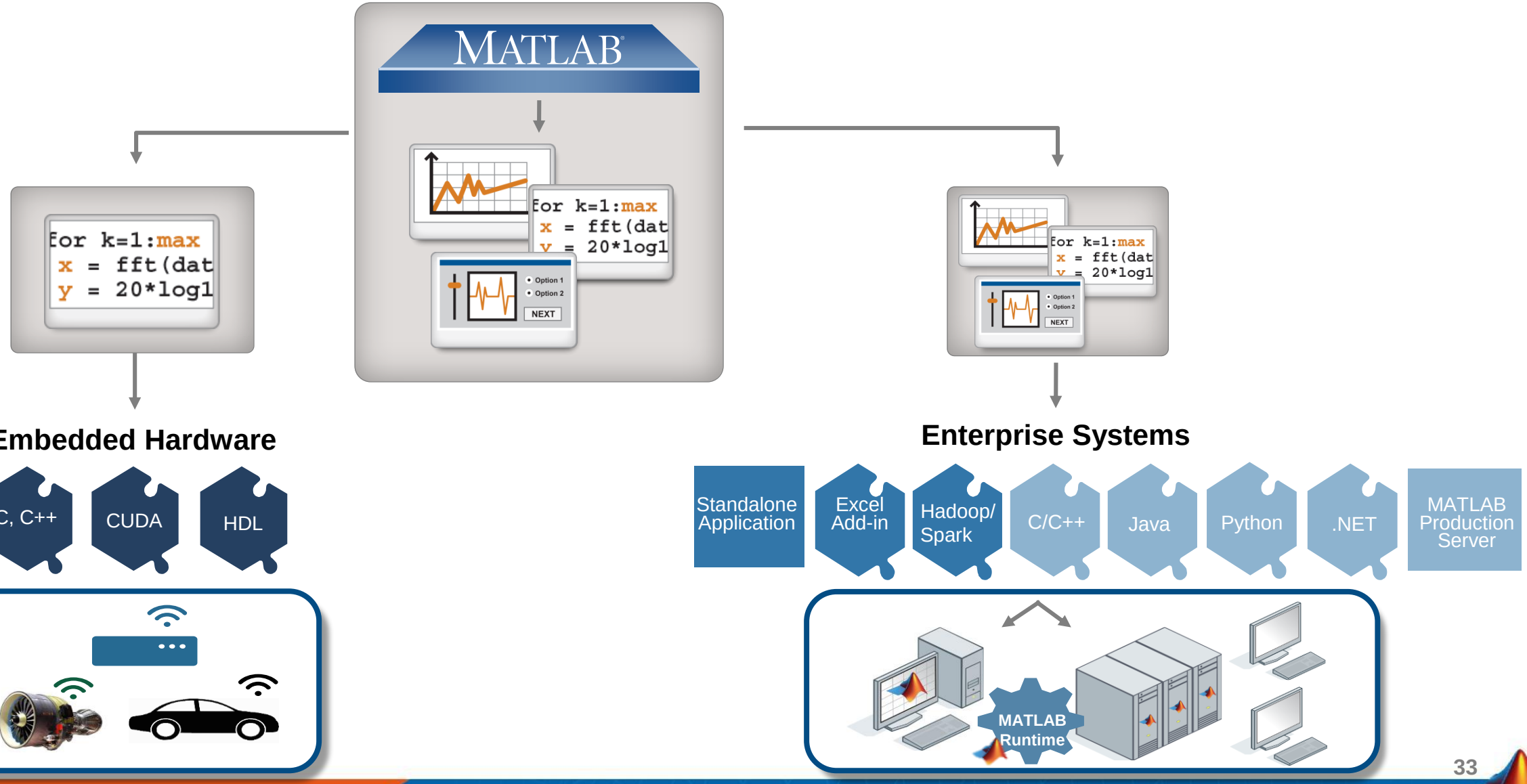
- AlexNet
- VGG-16
- VGG-19
- GoogLeNet
- InceptionV3 (coming soon)
- Resnet50

* single line of code to access model

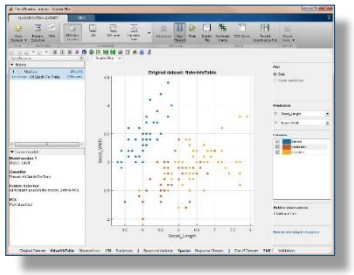
Import Models from Frameworks

- Caffe Model Importer
- TensorFlow/Keras Model Importer

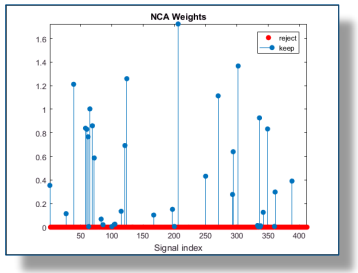
Integrate analytics with systems



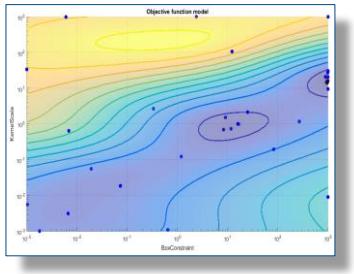
Key Takeaways



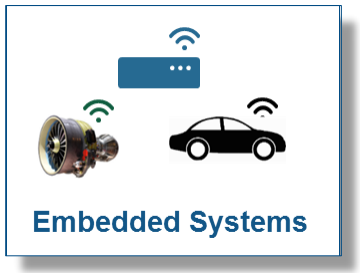
Use apps in your workflow to quickly compare and select candidate algorithms



Use feature selection to get rid of unnecessary features and prevent overfitting



Use programmatic workflows for fine-tuning model parameters to achieve robust performance



Use automatic code generation to rapidly deploy your analytics to embedded targets

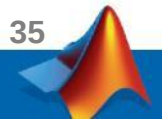
Next Steps – How can we help?

Resources

- Seminars and technical deep dives
- Data Analytics
- Machine Learning
- Deep Learning
- Big Data

More options

- Technical support
- Advanced customer support
- Installation, enterprise, and cloud deployment
- Training courses
- Consulting services



Thank you!

"No matter what industry our client is in, and no matter what data they ask us to analyze—text, audio, images, or video—MATLAB enables us to provide clear results faster."

Dr. G Subrahmanya VRK Rao, Cognizant