

MathWorks®

# Introducing Machine Learning

```
%% Classification Using Nearest Neighbors
knn = ClassificationKNN.fit(Xtrain,Ytrain, ...
    'Distance','euclidean');

%% Ensemble Learning: TreeBagger
opts = statset('UseParallel',true);

tb = TreeBagger(150,Xtrain,Ytrain,'method','classification',...
    'Options',opts,'OOBVarImp','on','cost',[0 1; 5 0]);
```

```
%% Generalized Linear Model - Logistic Regression
glm = GeneralizedLinearModel.fit(Xtrain,double(Ytrain));
    'linear','Distribution','binomial','link','logit');

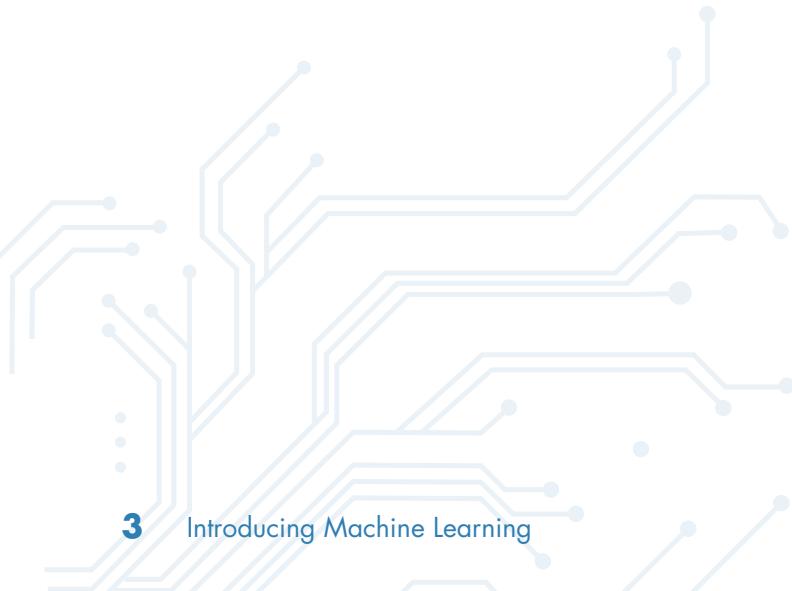
%% Discriminant Analysis
da = ClassificationDiscriminant.fit(Xtrain,Ytrain);
    'discrimType','quadratic');
```

# What is Machine Learning?

Machine learning teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases.

# More Data, More Questions, Better Answers

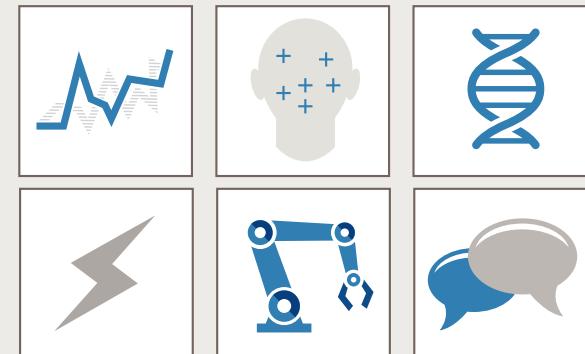
Machine learning algorithms find natural patterns in data that generate insight and help you make better decisions and predictions. They are used every day to make critical decisions in medical diagnosis, stock trading, energy load forecasting, and more. Media sites rely on machine learning to sift through millions of options to give you song or movie recommendations. Retailers use it to gain insight into their customers' purchasing behavior.



## Real-World Applications

With the rise in big data, machine learning has become particularly important for solving problems in areas like these:

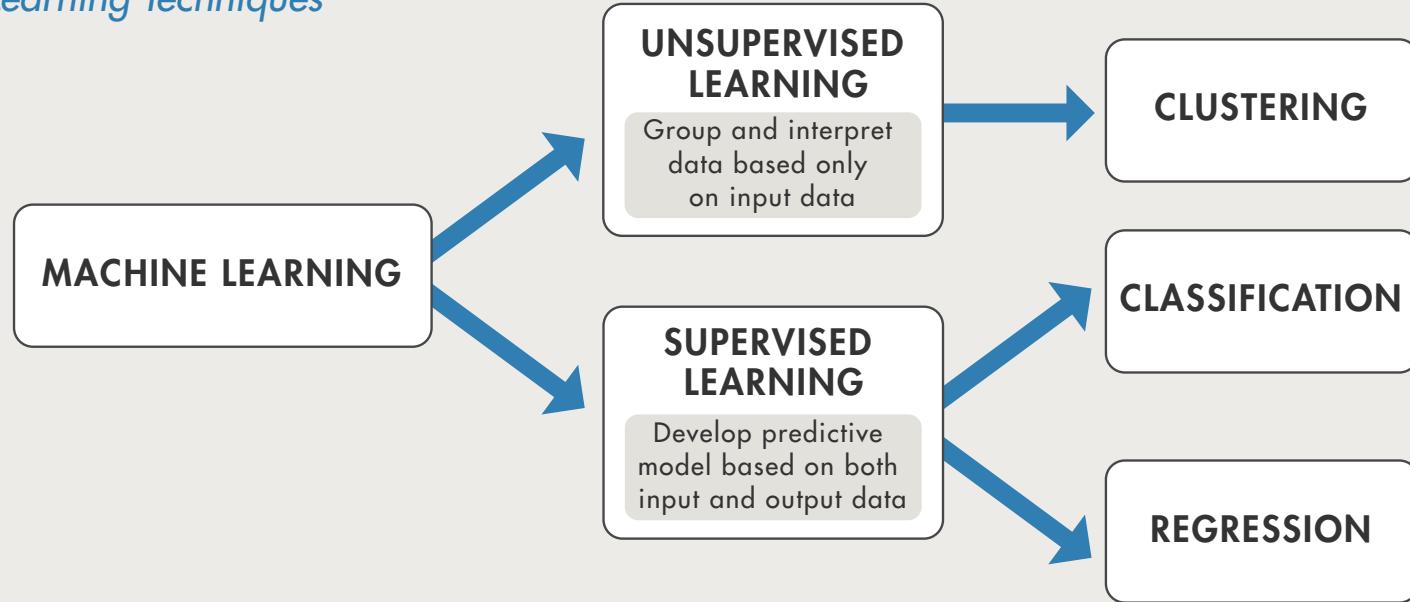
- Computational finance, for credit scoring and algorithmic trading
- Image processing and computer vision, for face recognition, motion detection, and object detection
- Computational biology, for tumor detection, drug discovery, and DNA sequencing
- Energy production, for price and load forecasting
- Automotive, aerospace, and manufacturing, for predictive maintenance
- Natural language processing



# How Machine Learning Works

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data.

## Machine Learning Techniques



# Supervised Learning

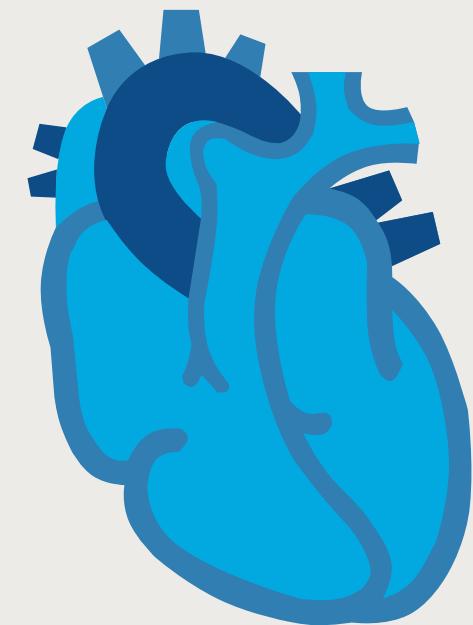
The aim of supervised machine learning is to build a model that makes predictions based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.

Supervised learning uses classification and regression techniques to develop predictive models.

- **Classification techniques** predict discrete responses—for example, whether an email is genuine or spam, or whether a tumor is cancerous or benign. Classification models classify input data into categories. Typical applications include medical imaging, speech recognition, and credit scoring.
- **Regression techniques** predict continuous responses—for example, changes in temperature or fluctuations in power demand. Typical applications include electricity load forecasting and algorithmic trading.

## *Using Supervised Learning to Predict Heart Attacks*

Suppose clinicians want to predict whether someone will have a heart attack within a year. They have data on previous patients, including age, weight, height, and blood pressure. They know whether the previous patients had heart attacks within a year. So the problem is combining the existing data into a model that can predict whether a new person will have a heart attack within a year.

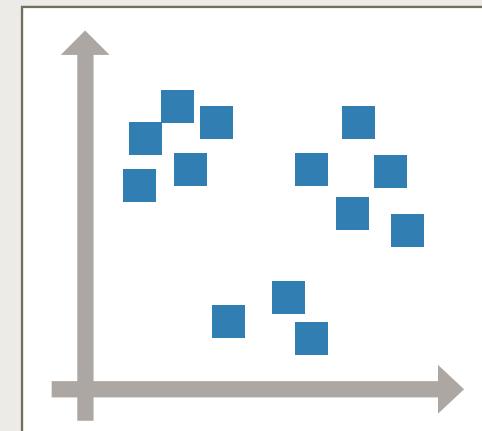


# Unsupervised Learning

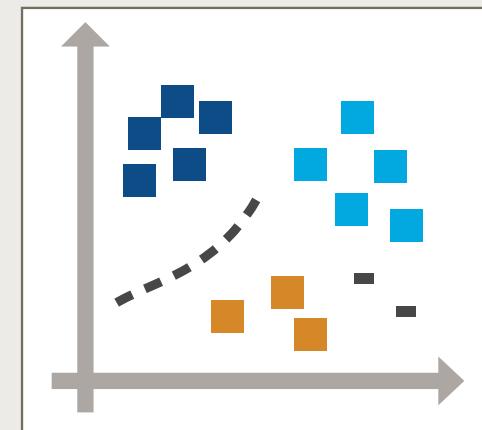
Unsupervised learning finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses.

**Clustering** is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data.

Applications for clustering include gene sequence analysis, market research, and object recognition.



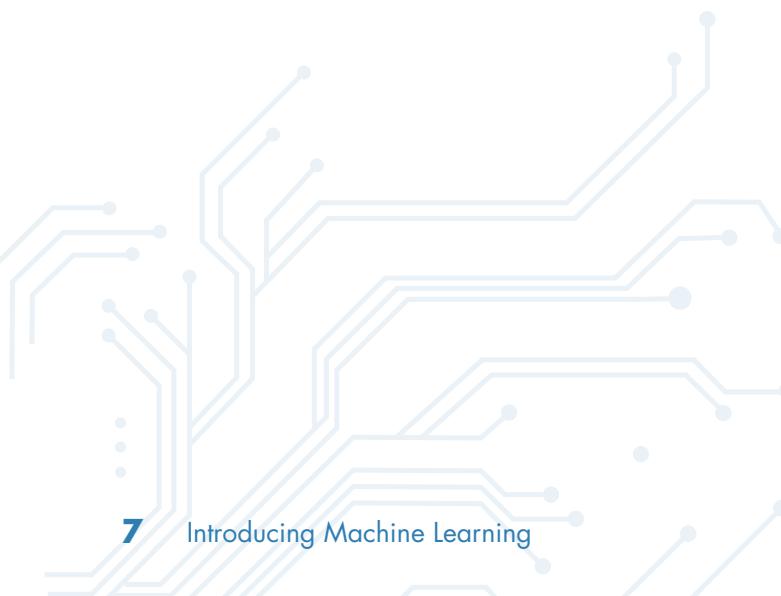
Clustering Patterns in the Data



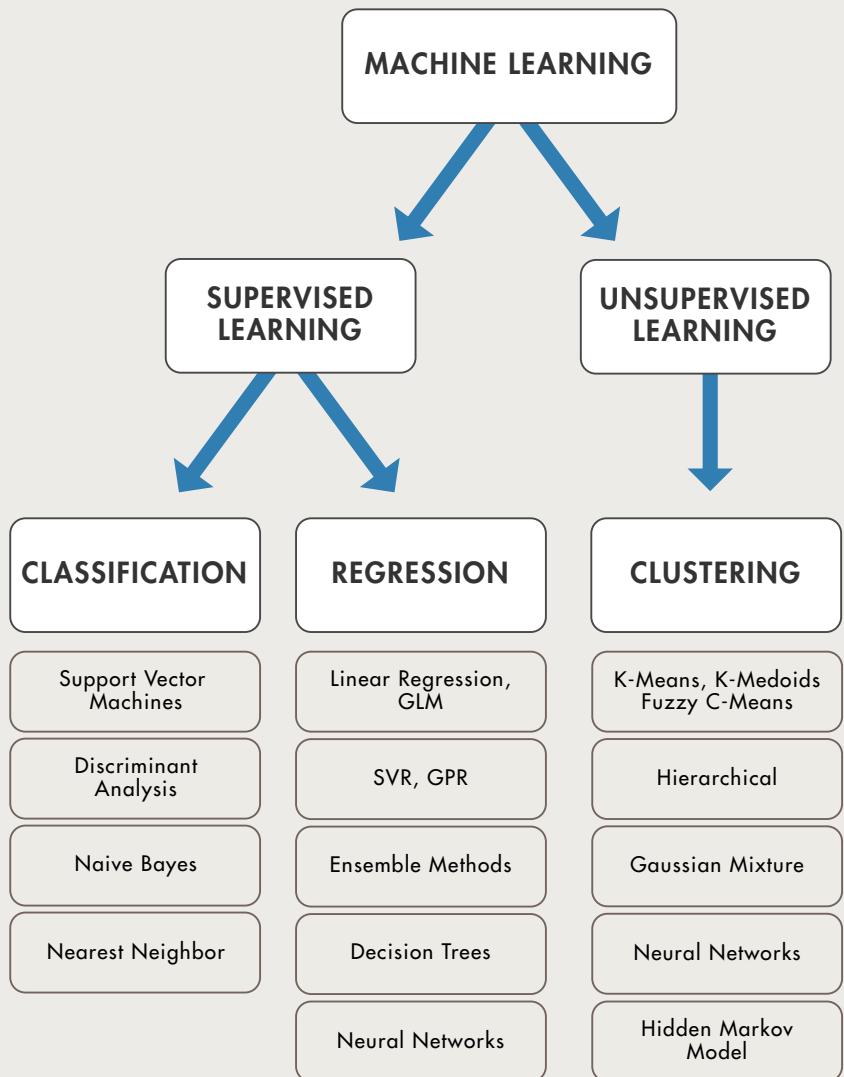
# How Do You Decide Which Algorithm to Use?

Choosing the right algorithm can seem overwhelming—there are dozens of supervised and unsupervised machine learning algorithms, and each takes a different approach to learning.

There is no best method or one size fits all. Finding the right algorithm is partly just trial and error—even highly experienced data scientists can't tell whether an algorithm will work without trying it out. But algorithm selection also depends on the size and type of data you're working with, the insights you want to get from the data, and how those insights will be used.



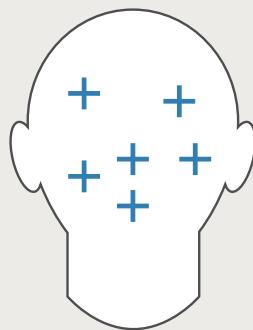
## Selecting an Algorithm



# When Should You Use Machine Learning?

Consider using machine learning when you have a complex task or problem involving a large amount of data and lots of variables, but no existing formula or equation. For example, machine learning is a good option if you need to handle situations like these:

*Hand-written rules and equations are too complex—as in face recognition and speech recognition.*



*The rules of a task are constantly changing—as in fraud detection from transaction records.*



*The nature of the data keeps changing, and the program needs to adapt—as in automated trading, energy demand forecasting, and predicting shopping trends.*



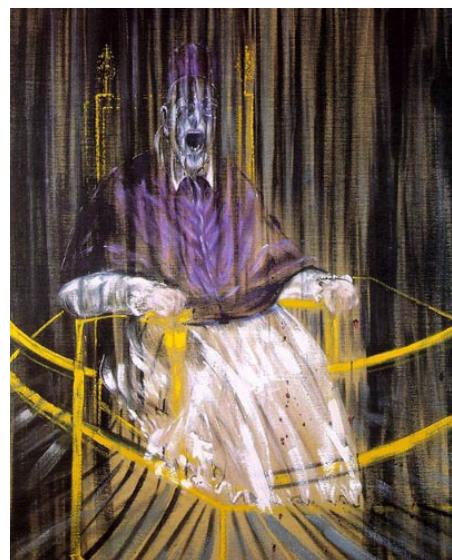
# Real-World Examples

## *Creating Algorithms that Can Analyze Works of Art*

Researchers at the Art and Artificial Intelligence Laboratory at Rutgers University wanted to see whether a computer algorithm could classify paintings by style, genre, and artist as easily as a human. They began by identifying visual features for classifying a painting's style. The algorithms they developed classified the styles of paintings in the database with 60% accuracy, outperforming typical non-expert humans.

The researchers hypothesized that visual features useful for style classification (a supervised learning problem) could also be used to determine artistic influences (an unsupervised problem).

They used classification algorithms trained on Google images to identify specific objects. They tested the algorithms on more than 1,700 paintings from 66 different artists working over a span of 550 years. The algorithm readily identified connected works, including the influence of Diego Velazquez's "Portrait of Pope Innocent X" on Francis Bacon's "Study After Velazquez's Portrait of Pope Innocent X."



# Real-World Examples

## *Optimizing HVAC Energy Usage in Large Buildings*

The heating, ventilation, and air-conditioning (HVAC) systems in office buildings, hospitals, and other large-scale commercial buildings are often inefficient because they do not take into account changing weather patterns, variable energy costs, or the building's thermal properties.

Building IQ's cloud-based software platform addresses this problem. The platform uses advanced algorithms and machine learning methods to continuously process gigabytes of information from power meters, thermometers, and HVAC pressure sensors, as well as weather and energy cost. In particular, machine learning is used to segment data and determine the relative contributions of gas, electric, steam, and solar power to heating and cooling processes. The building IQ platform reduces HVAC energy consumption in large-scale commercial buildings by 10% - 25% during normal operation.

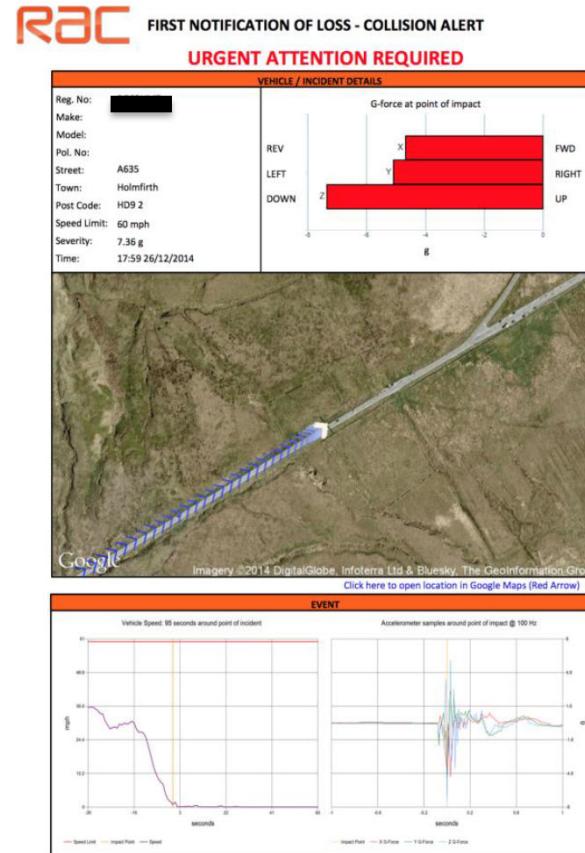


# Real-World Examples

## Detecting Low-Speed Car Crashes

With more than 8 million members, the RAC is one of the UK's largest motoring organizations, providing roadside assistance, insurance, and other services to private and business motorists.

To enable rapid response to roadside incidents, reduce crashes, and mitigate insurance costs, the RAC developed an onboard crash sensing system that uses advanced machine learning algorithms to detect low-speed collisions and distinguish these events from more common driving events, such as driving over speed bumps or potholes. Independent tests showed the RAC system to be 92% accurate in detecting test crashes.



# Learn More

*Ready for a deeper dive? Explore these resources to learn more about machine learning methods, examples, and tools.*

## ▶ Watch

[Machine Learning Made Easy](#) 34:34

[Signal Processing and Machine Learning Techniques for Sensor Data Analytics](#) 42:45

## 📄 Read

[Machine Learning Blog Posts](#): Social Network Analysis, Text Mining, Bayesian Reasoning, and more

[The Netflix Prize and Production Machine Learning Systems: An Insider Look](#)

[Machine Learning Challenges: Choosing the Best Model and Avoiding Overfitting](#)

## 🌐 Explore

[MATLAB Machine Learning Examples](#)

[Machine Learning Solutions](#)

[Classify Data with the Classification Learner App](#)