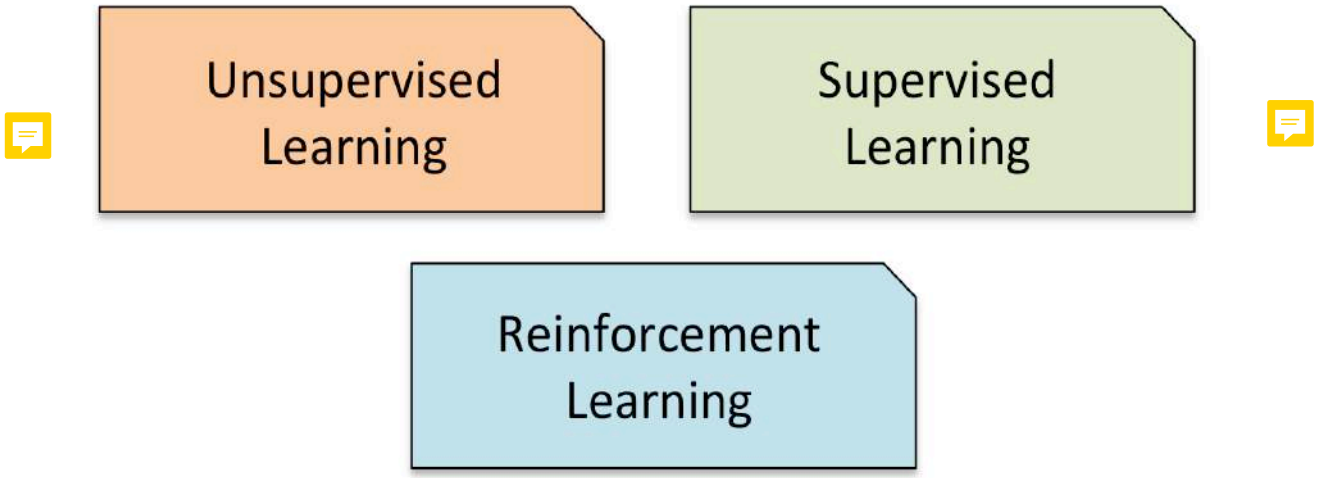
Introduction to Machine Learning

Mehul Motani
Electrical & Computer Engineering
National University of Singapore
motani@nus.edu.sg

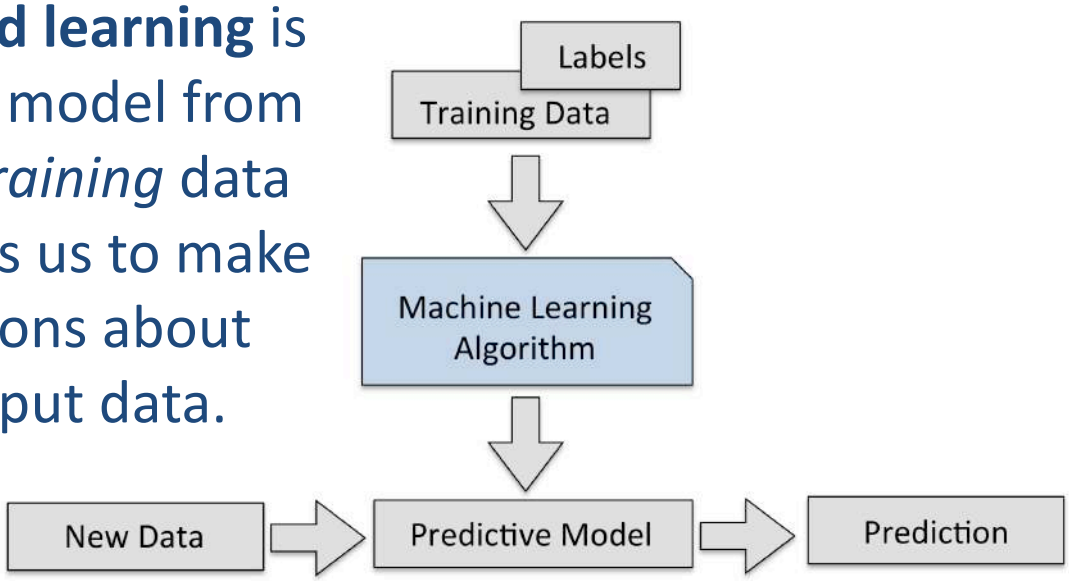
Three approaches to machine learning

Machine learning is the science of getting computers to act without being explicitly programmed. - Andrew Ng (Stanford)

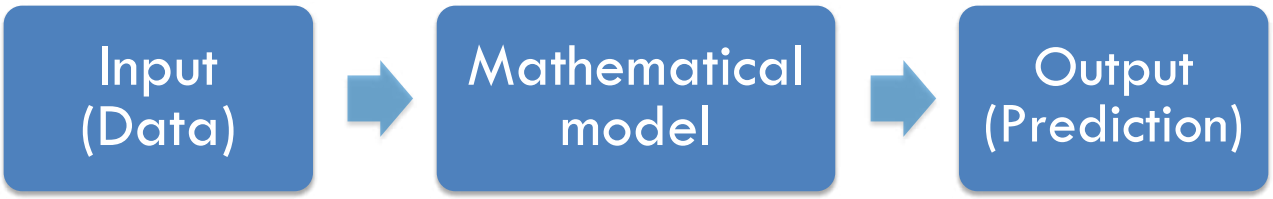


Supervised Learning

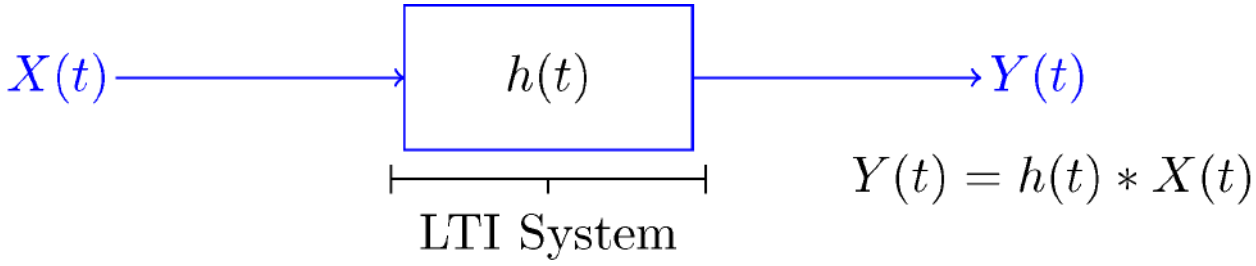
The main goal in **supervised learning** is to learn a model from labeled *training* data that allows us to make predictions about new input data.



Compare this to problem solving in engineering systems



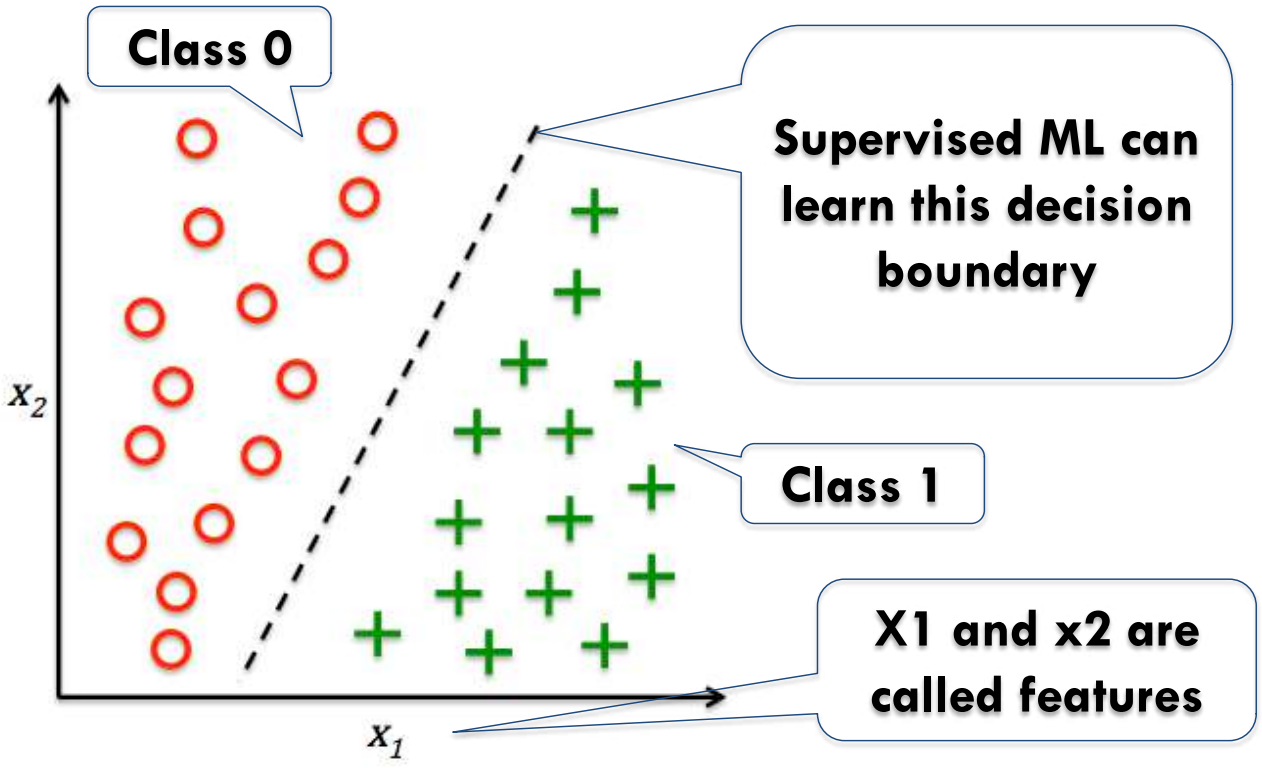
Linear Time Invariant Systems



Classification versus Regression

- **Classification task**
  - A supervised learning task with *discrete class labels*.
  - An example of a *binary classification* task is **email spam filtering**, where the ML algorithm learns a set of rules in order to distinguish between two possible classes: spam and non-spam e-mail.
  - An example of a *multi-class classification* task is **handwritten character recognition**. We first collect a training dataset that consists of multiple handwritten examples of each letter in the alphabet. Given a new handwritten character, the ML algorithm predicts the correct letter in the alphabet with certain accuracy.

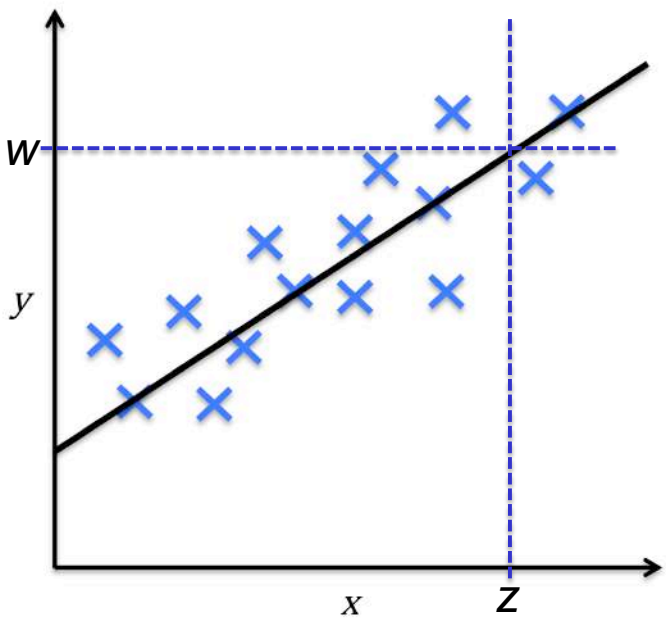
Example: Binary Classification



Classification versus Regression

- **Regression task**
  - A supervised learning task in which the outcome signal takes a *continuous value*.
  - In *regression analysis*, we are given a number of *predictor* (explanatory) variables and a continuous response variable (outcome), and we try to find a relationship between those variables that allows us to predict an outcome.
  - Suppose we are interested in predicting the test scores of students. If there is a relationship between the time spent studying for the test and the final scores, we could use it as training data to learn a model that uses the study time to predict the test scores of future students.

Example: Linear Regression



Given data which consists of (x,y) pairs, where x is the independent or predictor variable and y is the dependent or response variable.

We fit a straight line to this data that minimizes the average squared distance between the sample points and the fitted line.

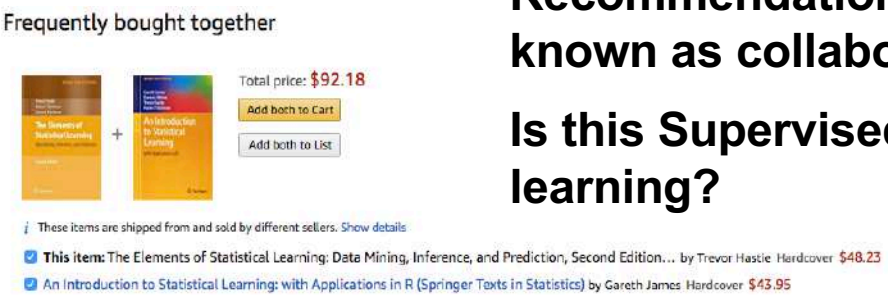
We can now use the learned best fit line to predict the output for new data, as shown here.

For example, z is mapped to w.

Examples of supervised learning

- Email spam filtering
- Facial recognition in Google Photos
- Fingerprint identification
- Siri / Google Assistant / Alexa / Cortana
- Facebook sponsored content
- Stock market prediction
- Recommendation systems **?**

Amazon Recommendation System



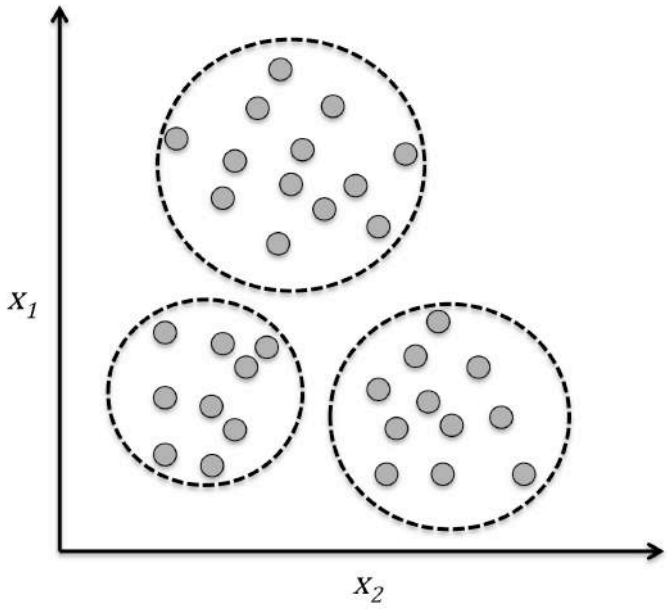
Recommendation systems are also known as collaborative filtering.

Is this Supervised or Unsupervised learning?

Unsupervised learning

- In unsupervised learning, we are dealing with unlabeled data or data of *unknown structure*.
- Using unsupervised learning techniques, we are able to explore the structure of our data to extract meaningful information without the knowledge of an outcome variable (or label).
- Compare that with supervised learning, in which we are dealing with labeled data.
- Examples of unsupervised learning are clustering and dimensionality reduction.

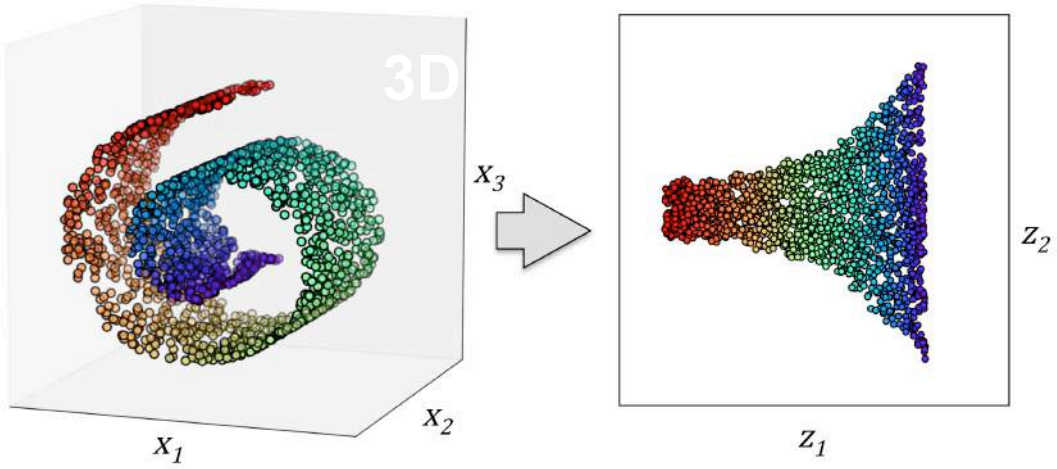
Finding subgroups with clustering



The figure illustrates how clustering can be applied to organizing unlabeled data into three distinct groups based on the similarity of their features *x*1 and *x*2.

For example, clustering allows marketers to discover customer groups based on their interests in order to develop targeted campaigns.

Dimensionality reduction for data compression

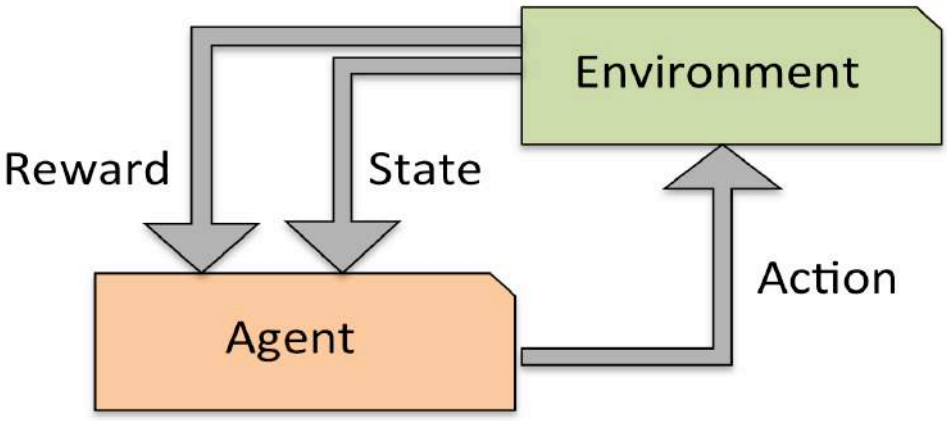


Data can be high-dimensional. Often we can reduce the data dimension by compressing the data without losing relevant information needed for the ML task. A typical algorithm for this case is principal component analysis (PCA).

Reinforcement learning

- In reinforcement learning (RL), the goal is to develop a system (*agent*) that improves its performance based on interactions with the *environment*.
- RL is related to *supervised* learning. However, in RL this feedback is not the correct ground truth label or value, but a measure of how well the action was measured by a *reward* function.
- RL learns a series of actions that maximizes the reward.

Reinforcement learning



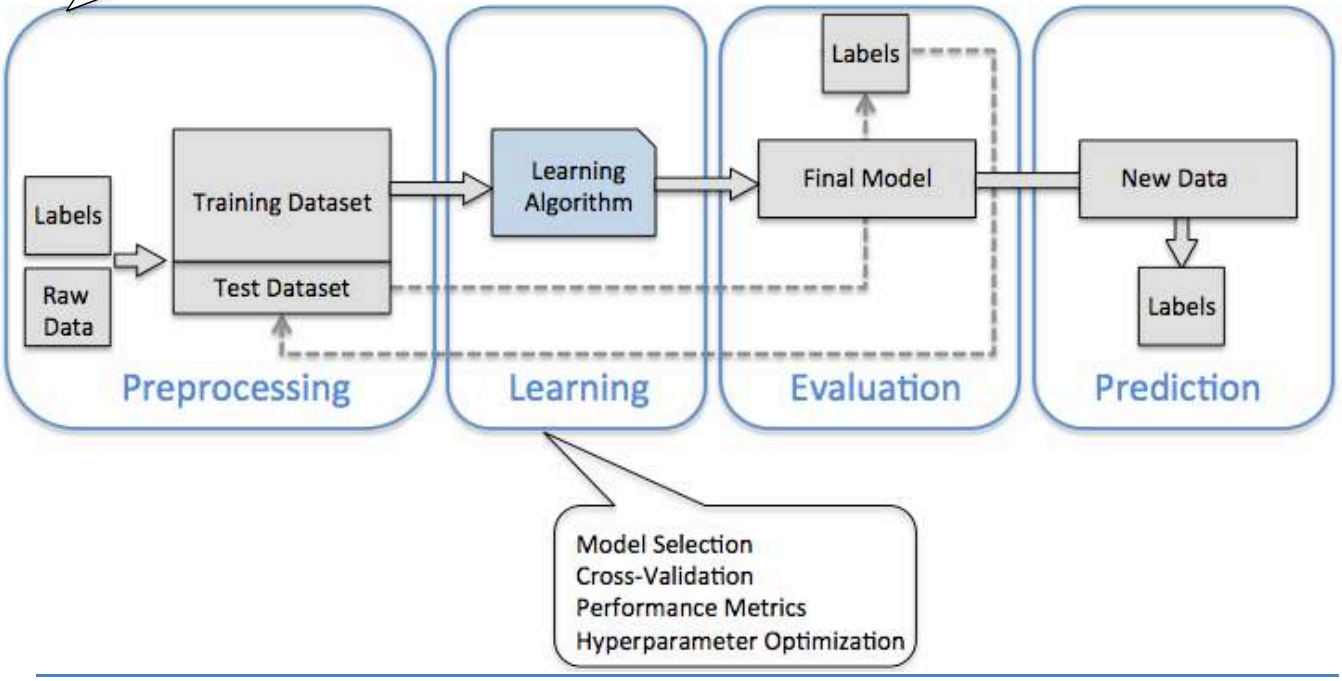
An example of reinforcement learning is DeepMind's AlphaGo engine. Here, the agent decides upon a series of moves depending on the state of the board (the environment), and the reward is defined as *win* or *lose* at the end of the game.

AI Robot that learns to walk https://youtu.be/gn4nRCC9TwQ

Robots that can adapt like animals
https://youtu.be/T-c17RKh3uE

Data cleaning and wrangling
Feature extraction & selection
Dimensionality reduction
Feature scaling
Data interpolation & imputation

Typical machine learning system



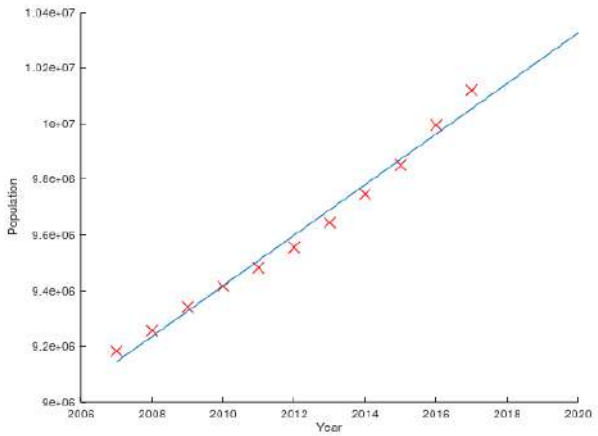
Model Selection
Cross-Validation
Performance Metrics
Hyperparameter Optimization

Popular learning algorithms

- Bayesian classifiers
- Linear models (regression)
- Logistic regression (classification)
- Decision trees
- Random forest (ensemble of decision trees)
- Support vector machine (SVM)
- Artificial neural networks
- Deep learning
- K-nearest neighbors
- K-means clustering

Example – Regression for Population Prediction

- Given historical population figures, predict the population in the future.
- Linear regression – Find the best fit line through a set of data
- Linear regression is supervised machine learning (function approximation)



Data: (x,y) – where x is the year, and y is the population

Data model: y = f(x) + noise, where x = (x1, x2...xn)

Training: Machine learns (fits) f(x) from labelled training set

Test: Machine predicts y from unlabeled test set

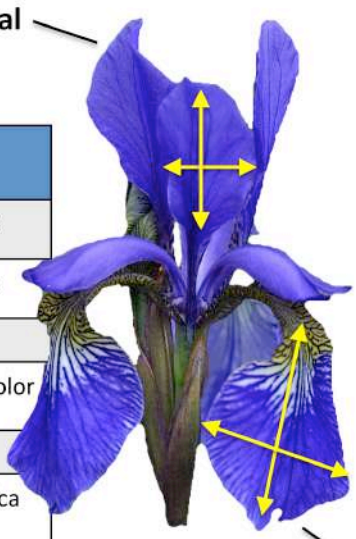
Example – Classification on Iris Dataset

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)



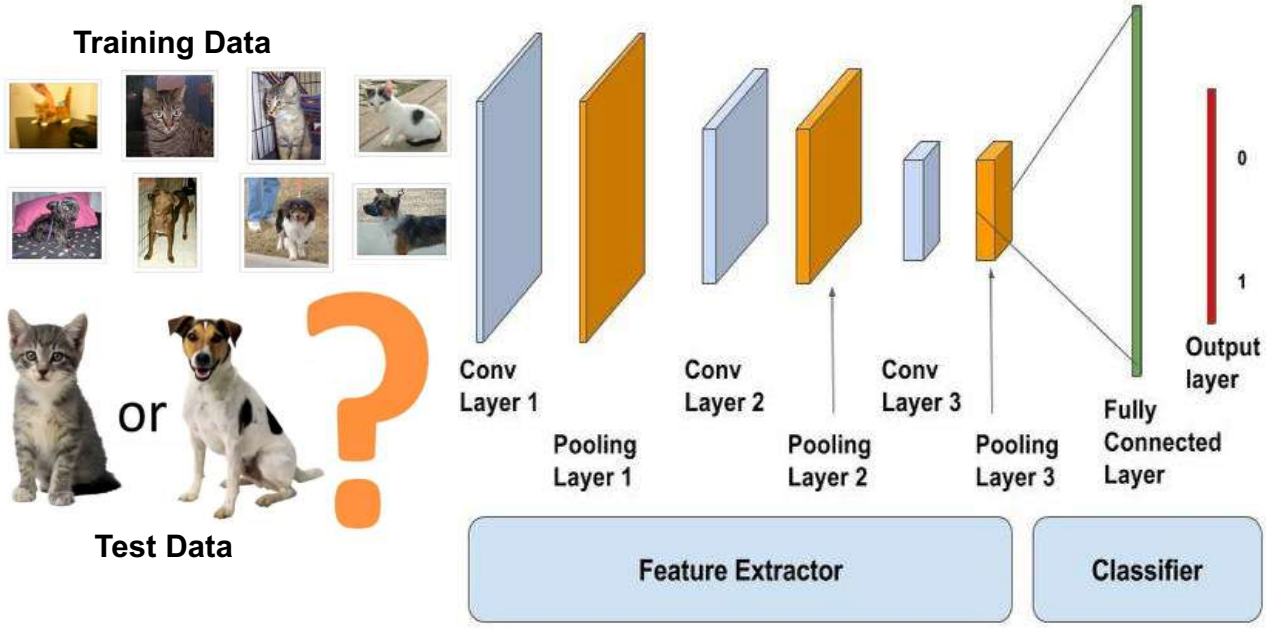
Iris Dataset: introduced by British statistician and biologist Ronald Fisher in his 1936 paper 'The use of multiple measurements in taxonomic problems'

Classification problem: Predict the Iris variety from its features.

Source: Principal Component Analysis by Sebastian Raschka

Example – Image Classification

Deep Learning Image Classification using Convolutional Neural Networks



Source: learnopencv.com

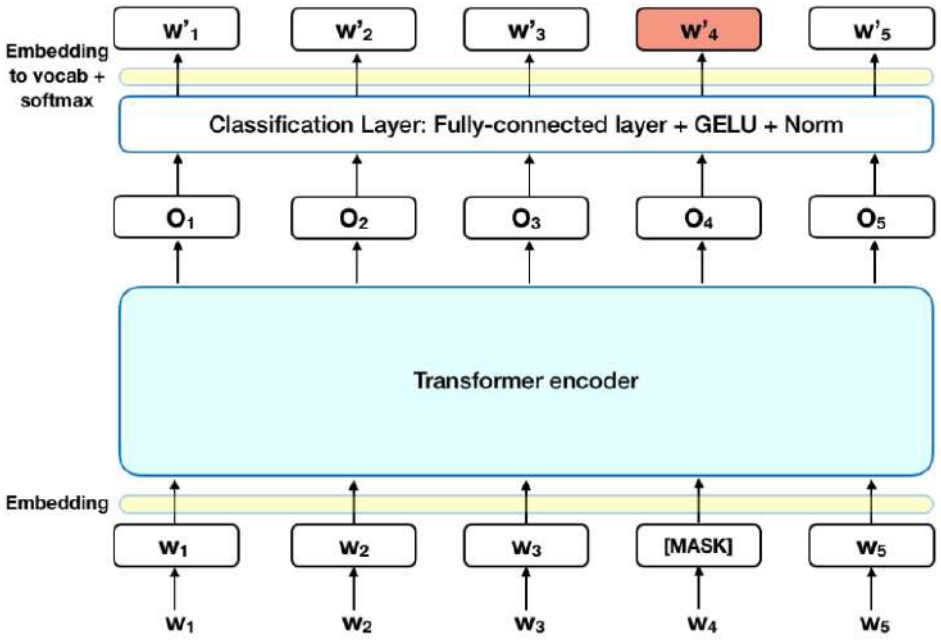
Example – Natural Language Processing

Deep Learning Natural Language Processing using BERT (Bidirectional Encoder Representations from Transformers)

Training Data

Large corpus of text data containing pairs of sentences in a certain language.

BERT is useful in many NLP tasks such as Question-Answering, Natural Language Understanding, and Machine Translation.



Source: towarddatascience.com

Thank you!

- Please send me your feedback and any questions you may have.
- The best way to contact me is via email:

  mehul.motani@gmail.com
- Thanks for listening!