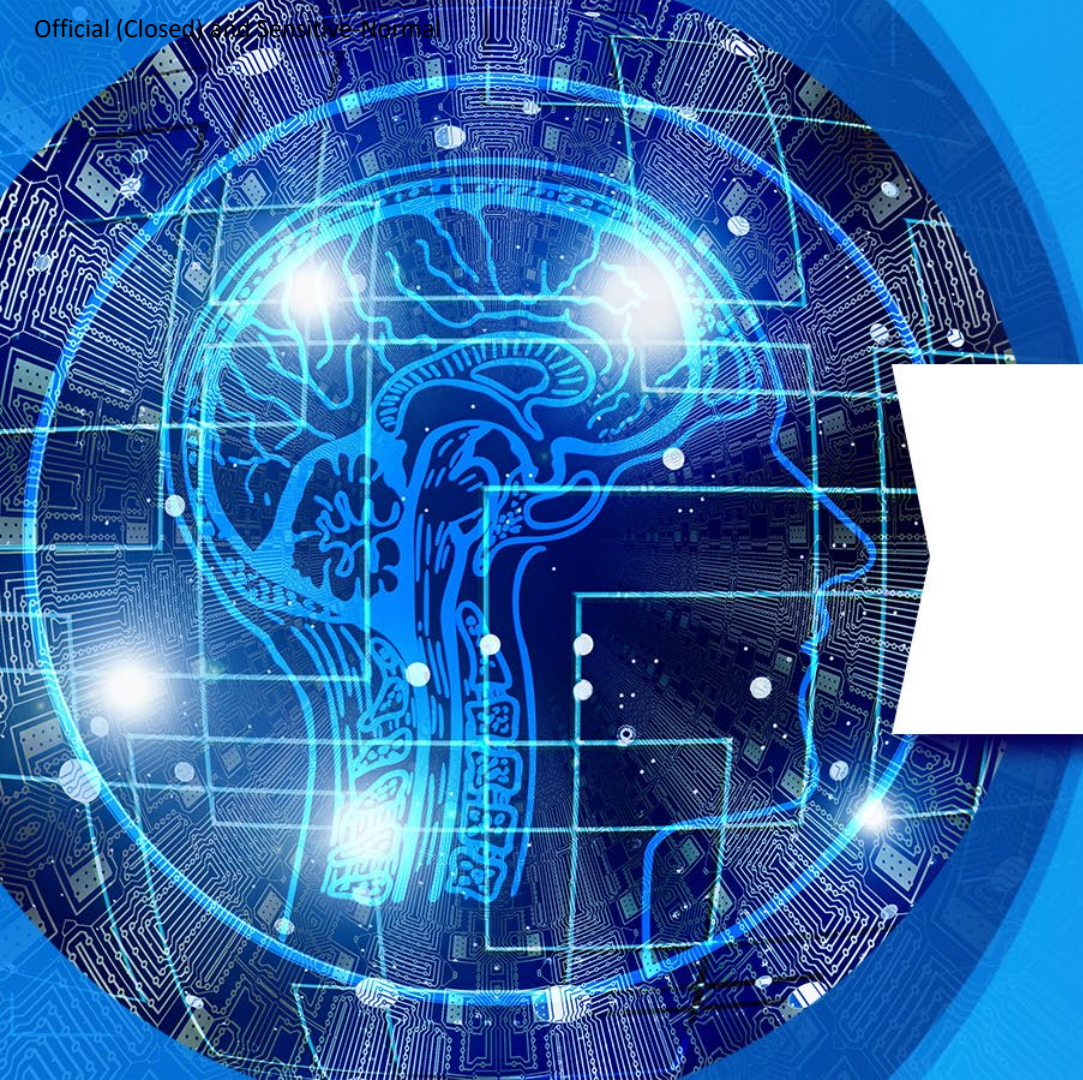


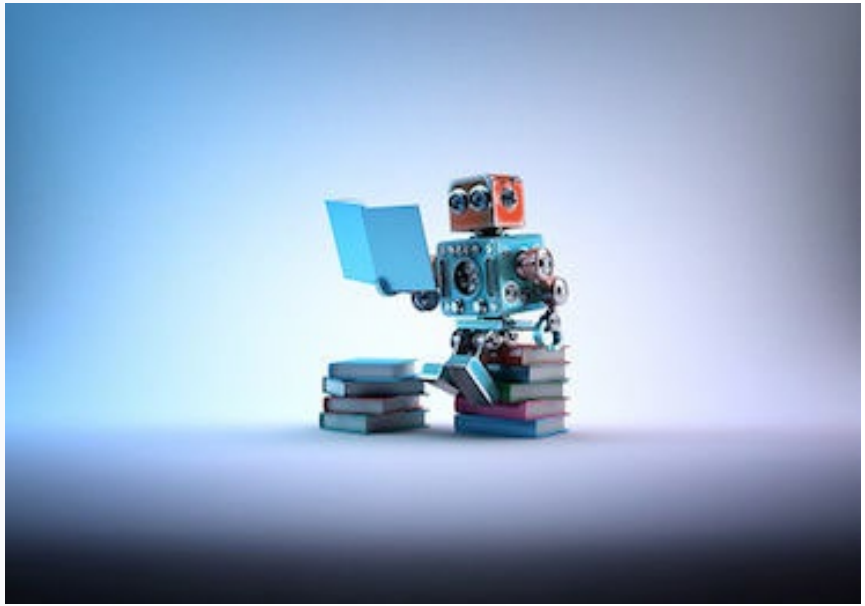
NANYANG
THE INNOVATIVE POLYTECHNIC

Introduction to Artificial Intelligence and Machine Learning

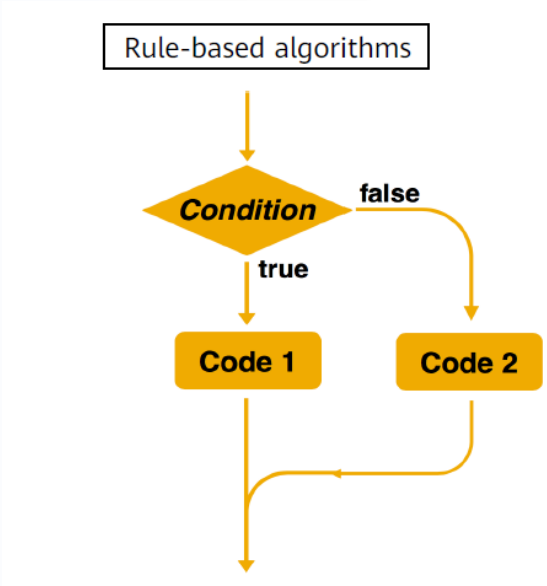
School of Engineering
Nanyang Polytechnic

2 What is Machine Learning

Can a machine learn?



Classical Rule-Based Algorithms

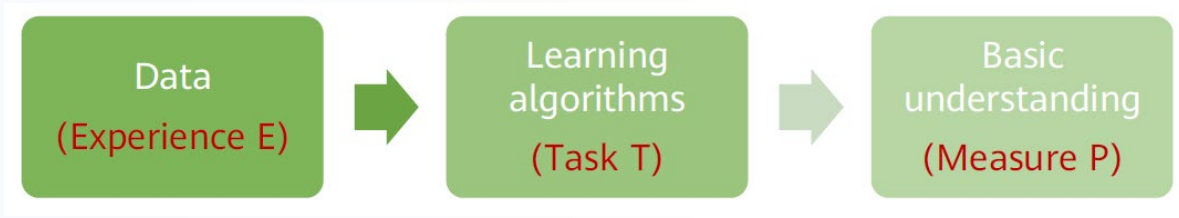


- Explicit programming is used to solve problems
- Rules can be manually specified

Definition of learning

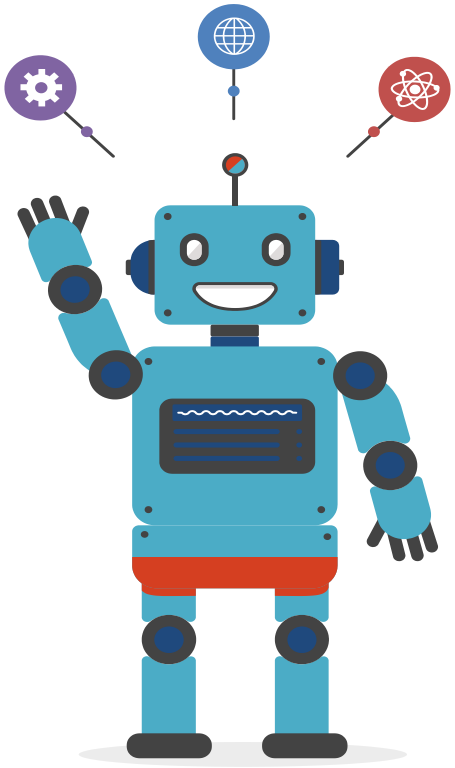
A computer programme is said to **learn** from **experience E** with respect to some class of **tasks T** and **performance measure P,** if its performance at tasks in T, as measured by P, improves with experience E."

– Tom Michell



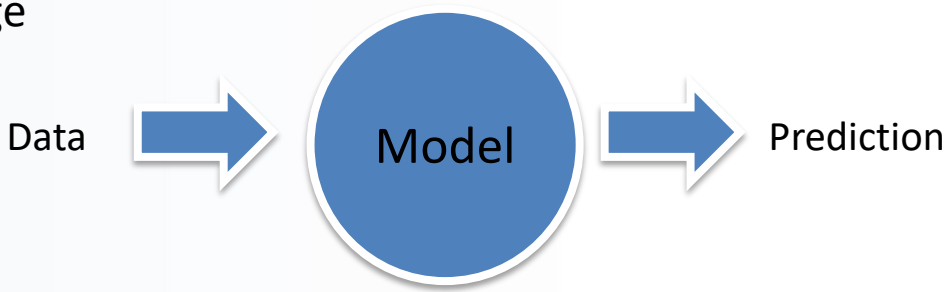
Machine Learning

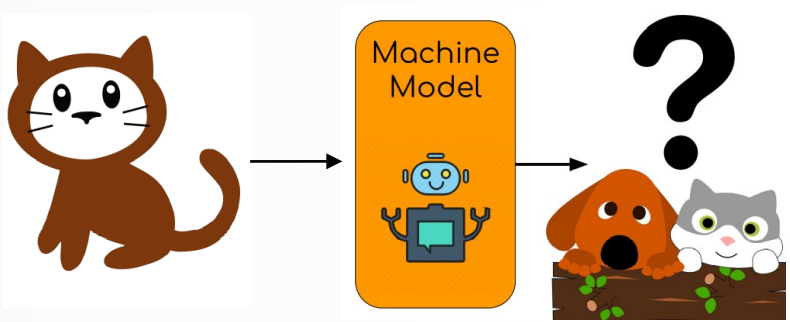
Machine learning allows computers to learn and infer from data



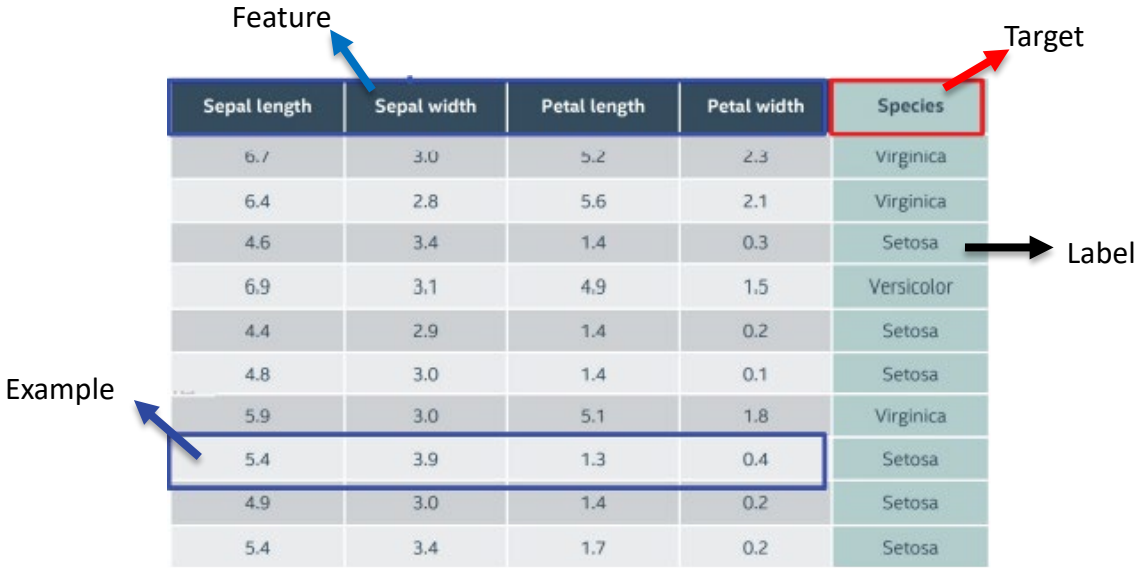
Machine Learning Model

After you trained the machine (called a model) to recognize dog or cat, then machine is able to tell your answer when you give it an image





Machine Learning Vocabulary



Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

- **Target**: Predicted category or value of the data (column to predict)
- **Features**: properties of the data used for prediction (non-target columns)
- **Example**: a single data point within the data (one row)
- **Label**: the target value for a single data point (for category), also known as a response value (for numeric)

Types of Machine Learning

Supervised Learning	Data Points have known outcome
Unsupervised Learning	Data Points have unknown outcome
Reinforcement Learning	Evaluate actions rather than telling the learning system how to perform correct actions

Supervised Learning

Apple

Orange

Banana

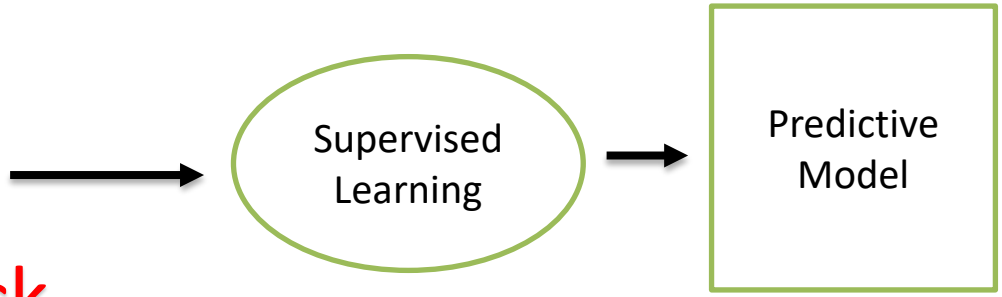
Supervised Learning


Duck

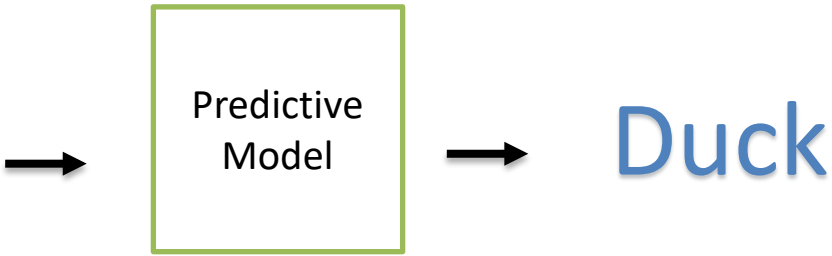

Duck


Not Duck

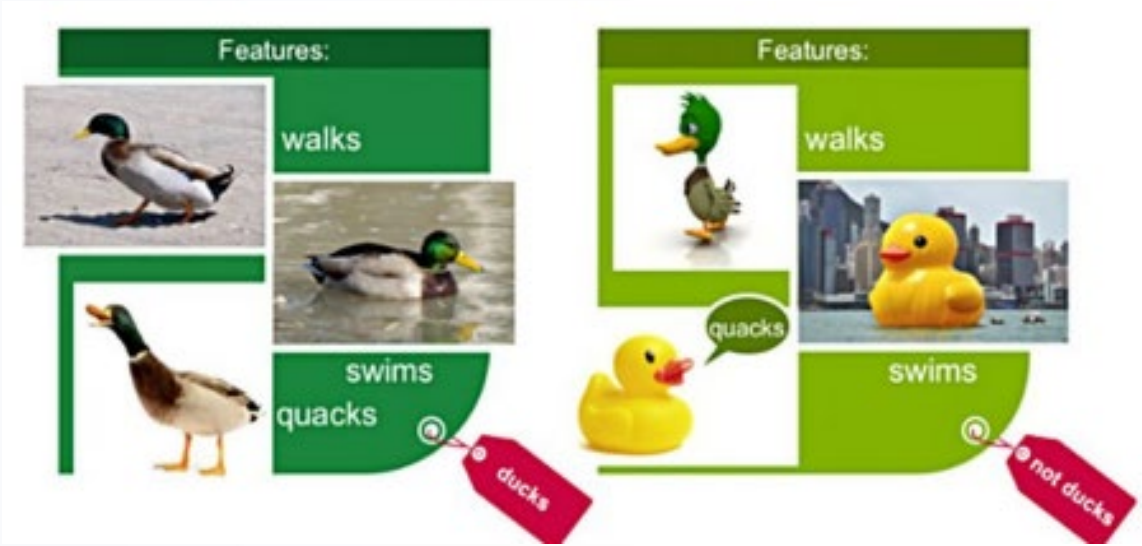

Not Duck



- The algorithm learns patterns from labelled data and makes predictions and try to label new data

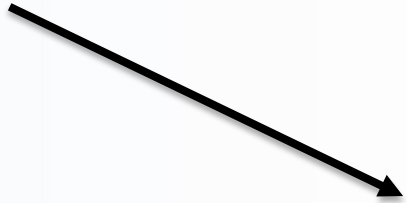



Supervised Learning



- Uses labelled training data to improve programs future actions. Reproduces known knowledge. Learn by example approach.
- Supervised learning needs to be given examples of what is “good” and what is “bad”

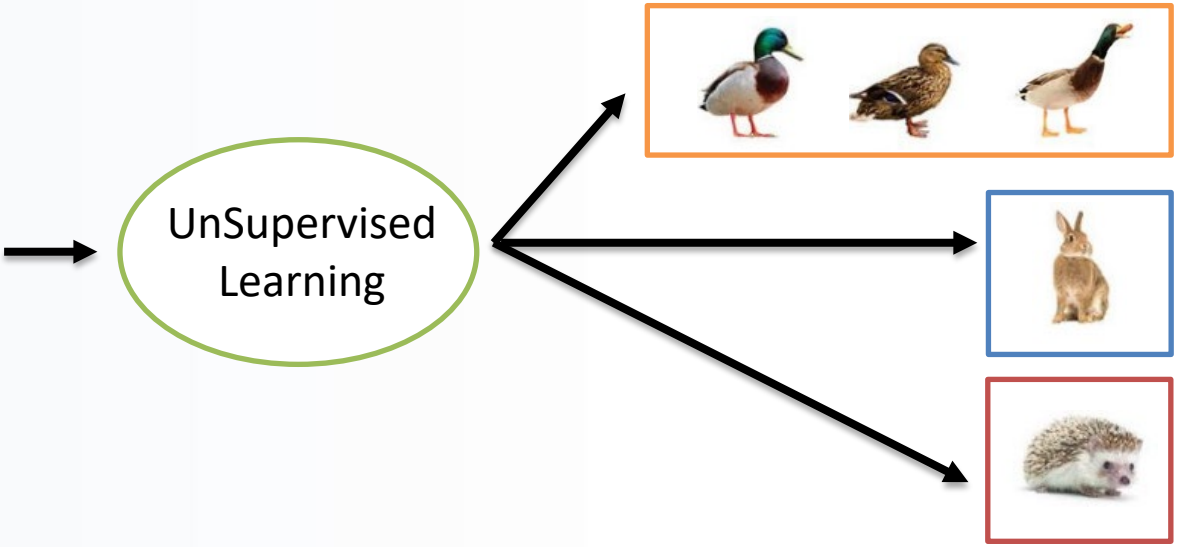
UnSupervised Learning



UnSupervised Learning



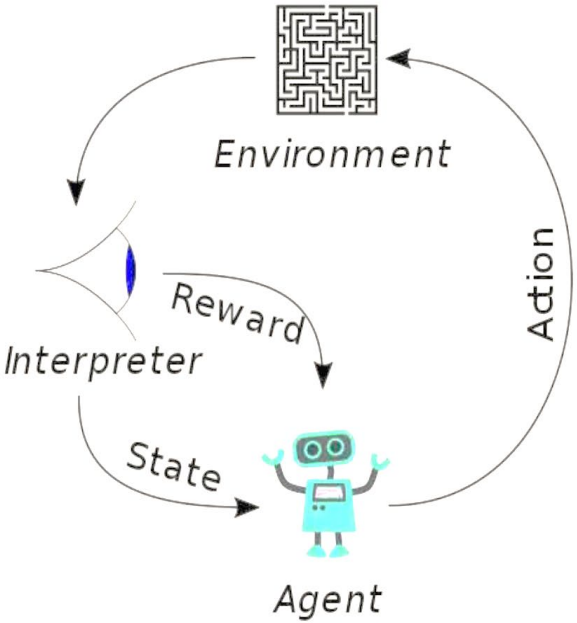




- Uses unlabeled data, correct classes are not known. Interpret and Groups the input data only

Reinforcement Learning

- Under a given environment (eg maze, stage..), an **agent** (eg human, robot…) learns how to choose an optimal action in an **environment**
- Each action (eg move left, right…) will entail a **reward signal** (eg +50, -100 ….) to influence the agent's decision making process.
- The agent learn from **experience** and adjust its behavior (a mapping from states to action) over time to achieve the desired outcome.



Reinforcement Learning - Best Behavior

- Reinforcement learning: always looks for best behaviors. Reinforcement learning is targeted at machines or robots.
  - Autopilot: Should it brake or accelerate when the yellow light starts to flash?
  - Cleaning robot: Should it keep working or go back for charging?

Type of Supervised Learning

Regression

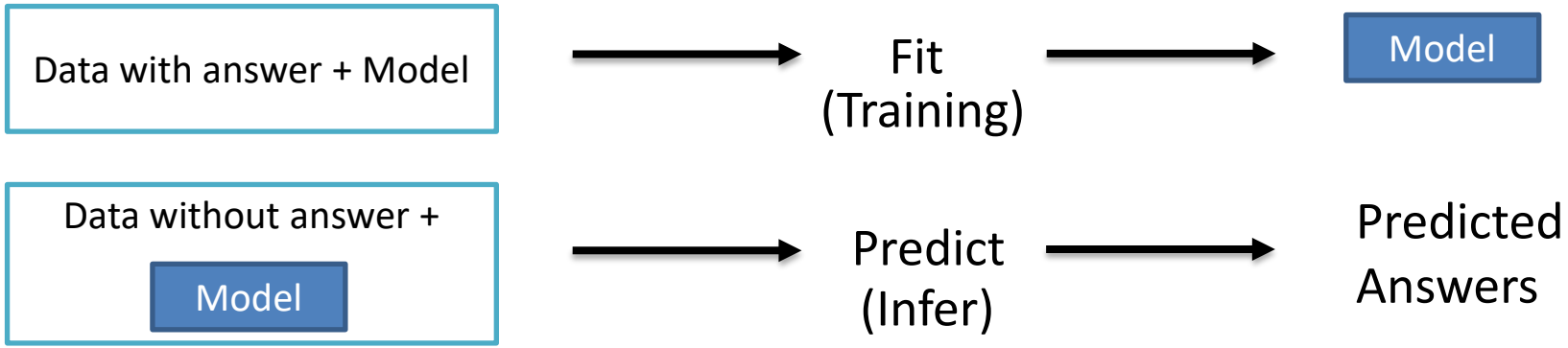
Outcome is continuous (numerical)

Classification

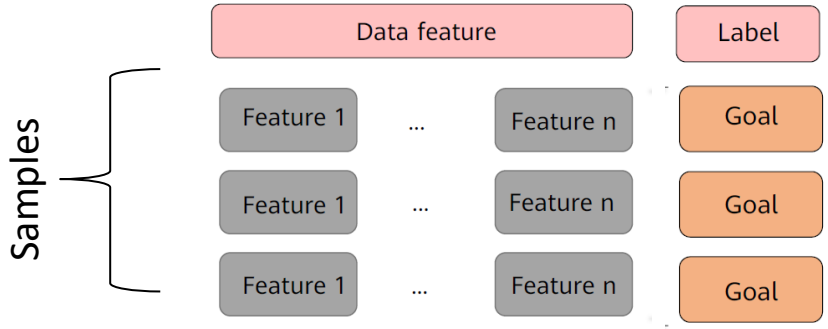
Outcome is a category

- Binary
- Multiclass

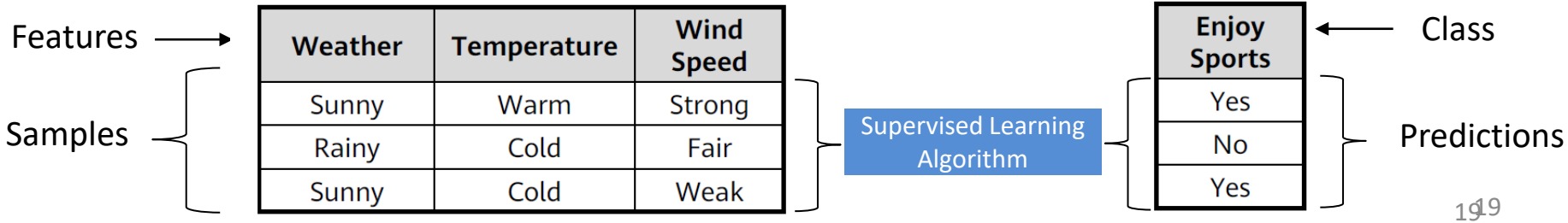
Supervised Learning Overview



Supervised Learning

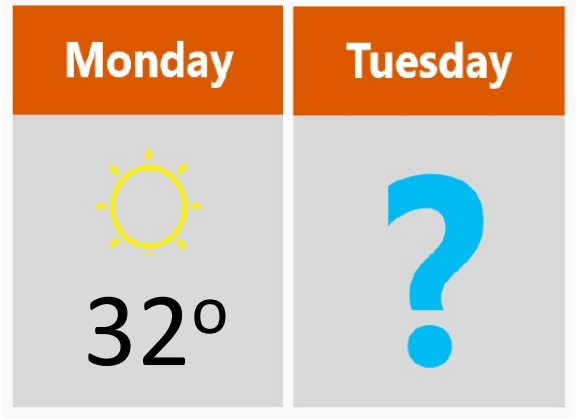


Weather	Temperature	Wind Speed	Enjoy Sports
Sunny	Warm	Strong	Yes
Rainy	Cold	Fair	No
Sunny	Cold	Weak	Yes



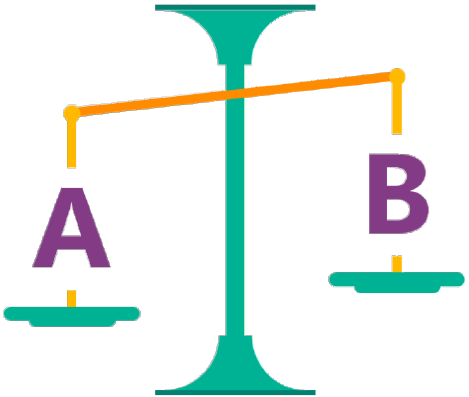
Supervised Learning - Regression

- Regression: To discover the dependencies in the sample dataset by mapping the relationships between the feature values and the response.
- Regression Questions:
  - How much will I benefit from the stock next week?
  - What's the temperature on Tuesday?

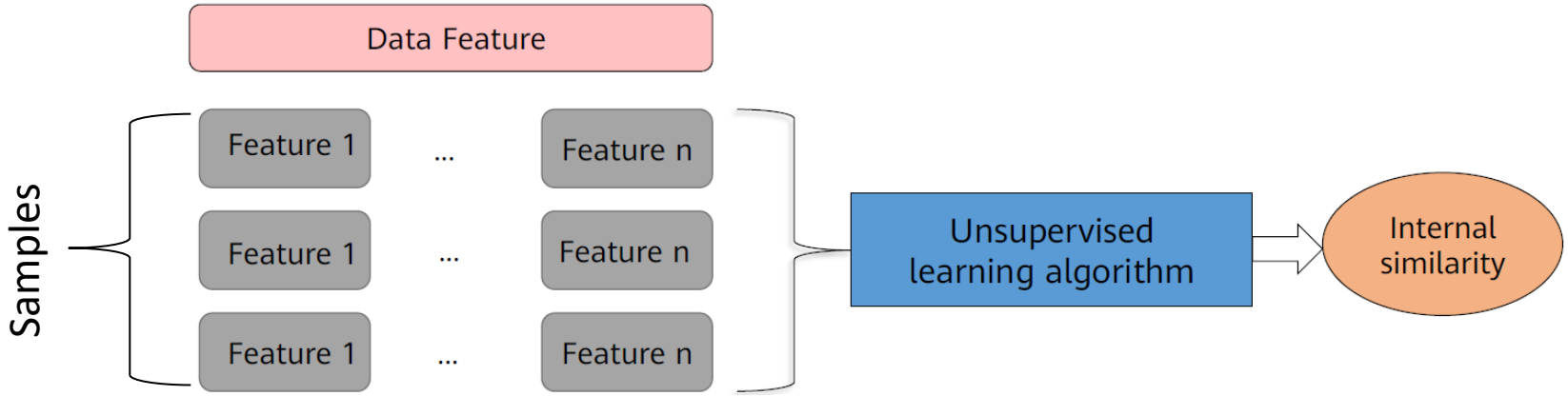


Supervised Learning - Classification

- Classification: maps samples in a sample dataset to a specified category by using a classification model.
- Classification Questions:
  - Will there be a traffic jam on XX road during the morning rush hour tomorrow?
  - Which method is more attractive to customers: $10 voucher or 25% off?



Unsupervised Learning



Features →

Samples

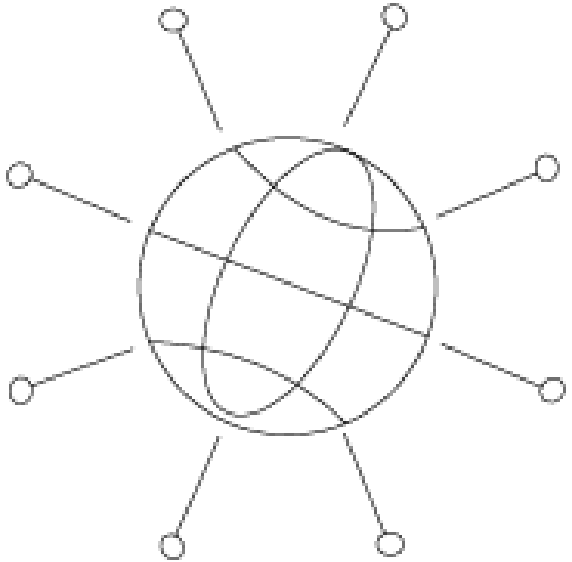
Monthly Consumption	Commodity	Consumption Time
1000–2000	Badminton racket	6:00–12:00
500–1000	Basketball	18:00–24:00
1000–2000	Game console	00:00–6:00

Category
Cluster 1
Cluster 2

22

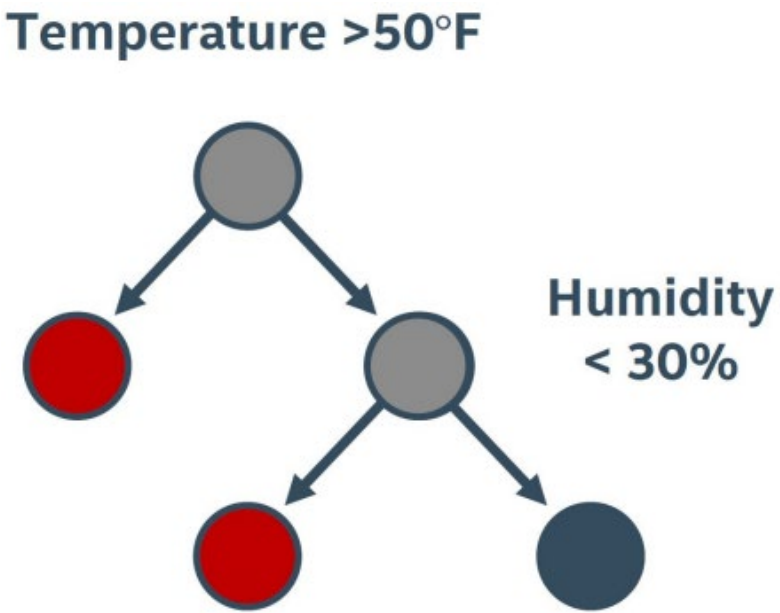
Sample Classification Algorithms

- K Nearest Neighbors
- Logistic Regression
- Support Vector Machine
- Gaussian Naive Bayes
- Decision Tree / Random Forest
- Neural Networks
- Ensemble Methods

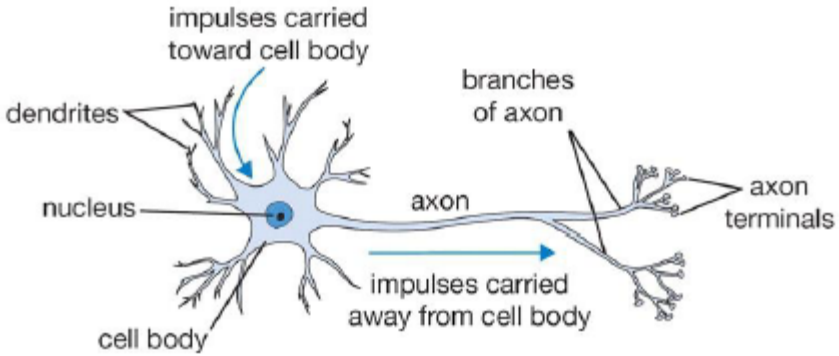


Decision Tree

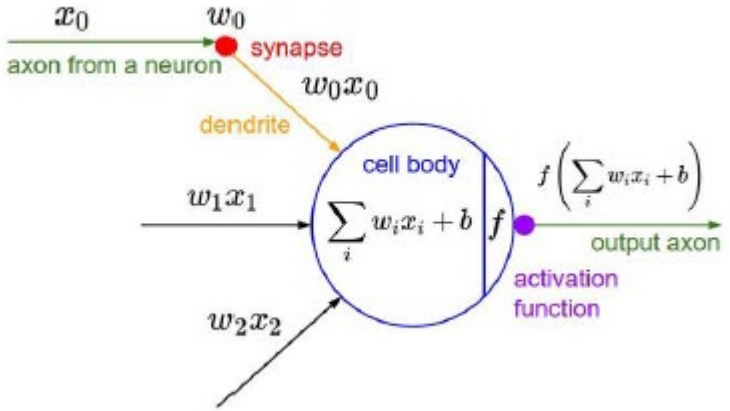
- Decision tree is easy to interpret and implement
- Heterogeneous input data allowed, preprocessing required
- However, decision trees tend to overfit
- Pruning helps reduce variance to a point. Often not significant for model to generalize well



Neural Networks: Biological Inspiration for Computation

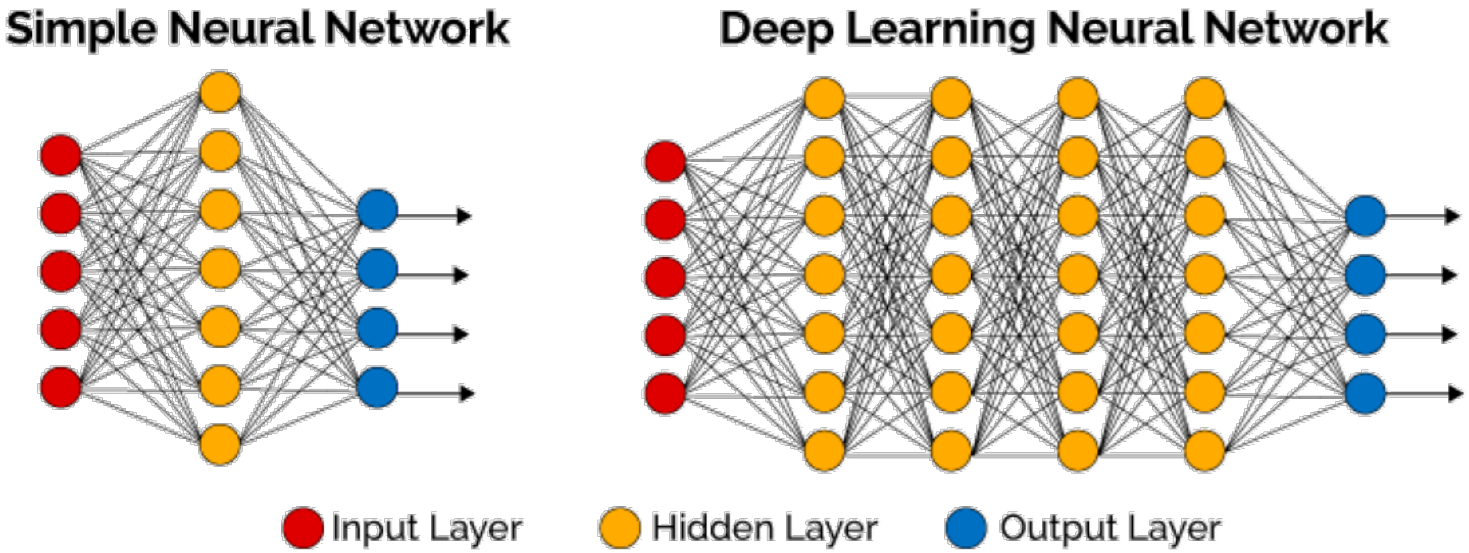


Neuron: computational building block for the brain



(Artificial) Neuron: computational building block for the “neural network”

Combing Neurons in Hidden Layers: Power to Approximate



Universality: For any arbitrary function *f(x)*, there exists a neural network that closely approximate it for any input *x.*

Individual Exercise: Neural Network Playground

- Go to the URL https://playground.tensorflow.org/
- Locate the following hyperparameters and explore different settings.
  - Features
  - Neurons and Hidden Layers
  - Activation Functions
  - Regularization
  - Learning Rate
- Are you able to explain the meaning of the hyperparameters?

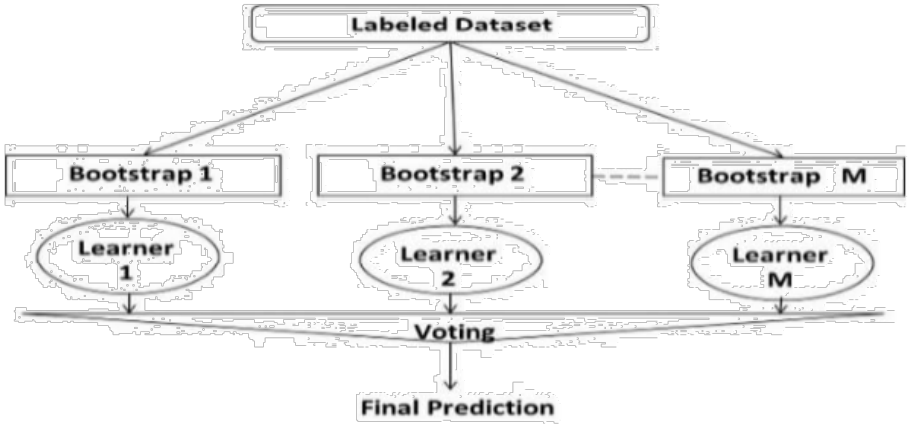
Ensemble Methods

The main principle behind Ensemble methods is that a group of “weak learners” can come together to form a “strong learner”.

There are 3 types of Ensemble methods

1. Bagging (Bootstrap Aggregating)
2. Boosting
3. Stacking

Bagging

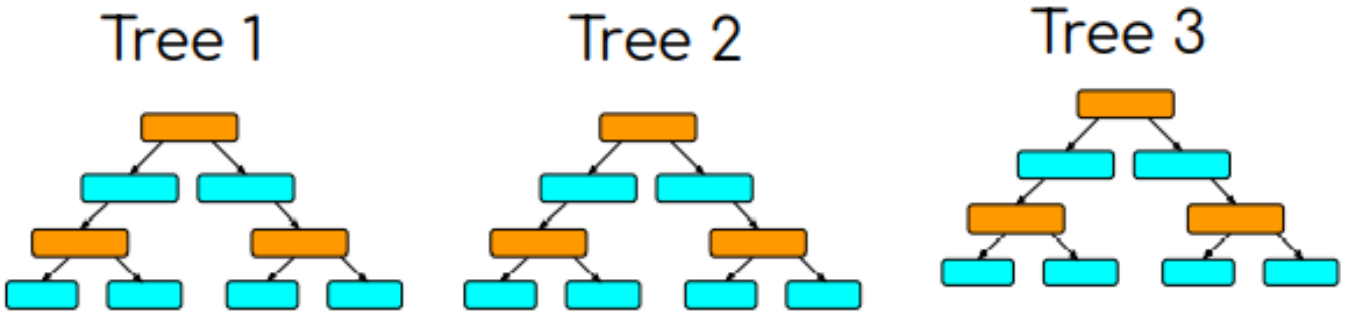


Bagging (Bootstrap Aggregating) creates separate samples of the training dataset and creates a classifier for each sample.

- The results of these multiple classifiers are then combined (such as averaged or majority voting).

Example: Random Forest

Random Forest classifier is a bagging ensemble method based on lots of decision trees with random selection of subsets of training samples.



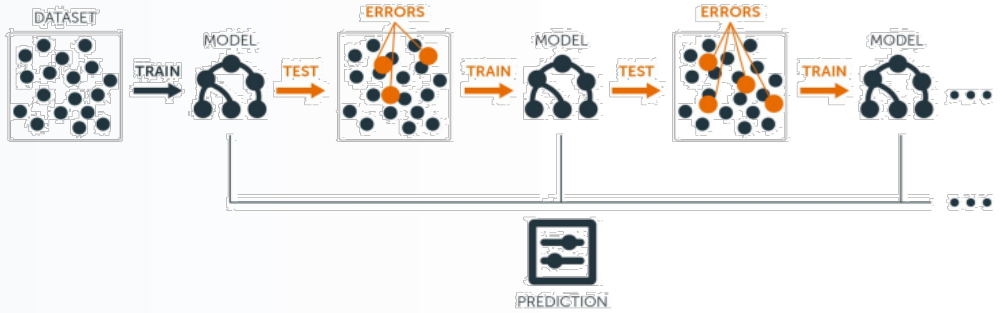
The result is based on the majority votes from all the decision trees

Boosting

- **Boosting** starts out with a base classifier that is prepared on the training data.
- A second classifier is then created behind it to focus on the instances in the training data that the first classifier got wrong.
- The process continues to add classifiers until a limit is reached in the number of models or accuracy

Example: Gradient Boosting

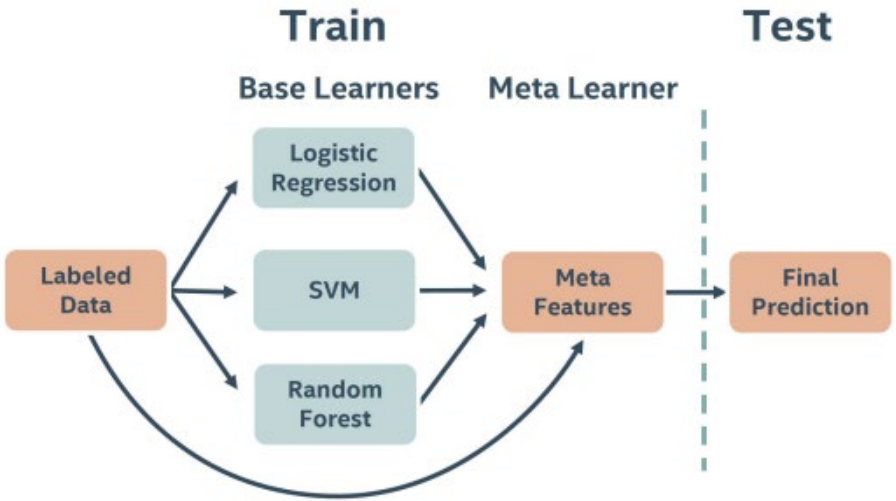
- Gradient Boosting is a ensemble boosting method that "boosting" many weak predictive models into a strong one, in the form of ensemble of weak models.
- Boosting utilizes different loss functions At each stage, the margin is determined for each point. Margin is positive for correctly classified points and negative for misclassifications. Value of loss function is calculated from margin



Stacking

- **Stacking** starts out with a set of base-level classifiers and train a meta-level classifier to combine the outputs of the base-level classifiers

  - Models of any kind combined to create stacked model
  - Output of base learners can be combined via majority vote or weighted
  - Additional hold out data needed if meta learner parameters are used
  - Be aware of increasing model complexity



Binary Classification

- In many predictive analysis, we are interested in YES/NO analysis such as spam/not-spam, health/not-healthy etc.
- In this case, we can form a confusion matrix of 2 columns and 2 rows

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Type II Error

Type I Error

Accuracy

Accuracy is the percentage of correct hits

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

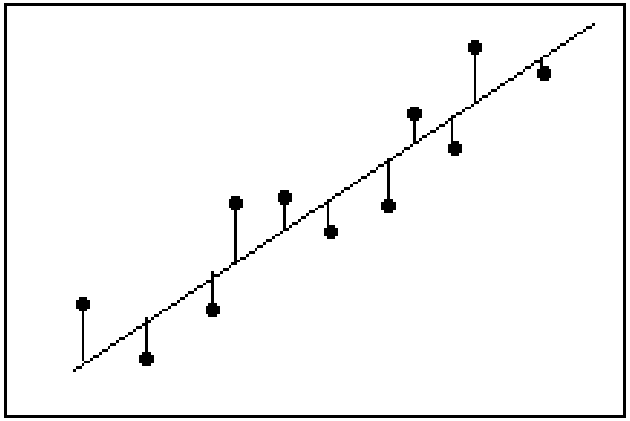
$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

Regression Algorithms

- Linear Regression (Most Common)
- Ridge Regression
- Lasso Regression
- Elastic Net Regression

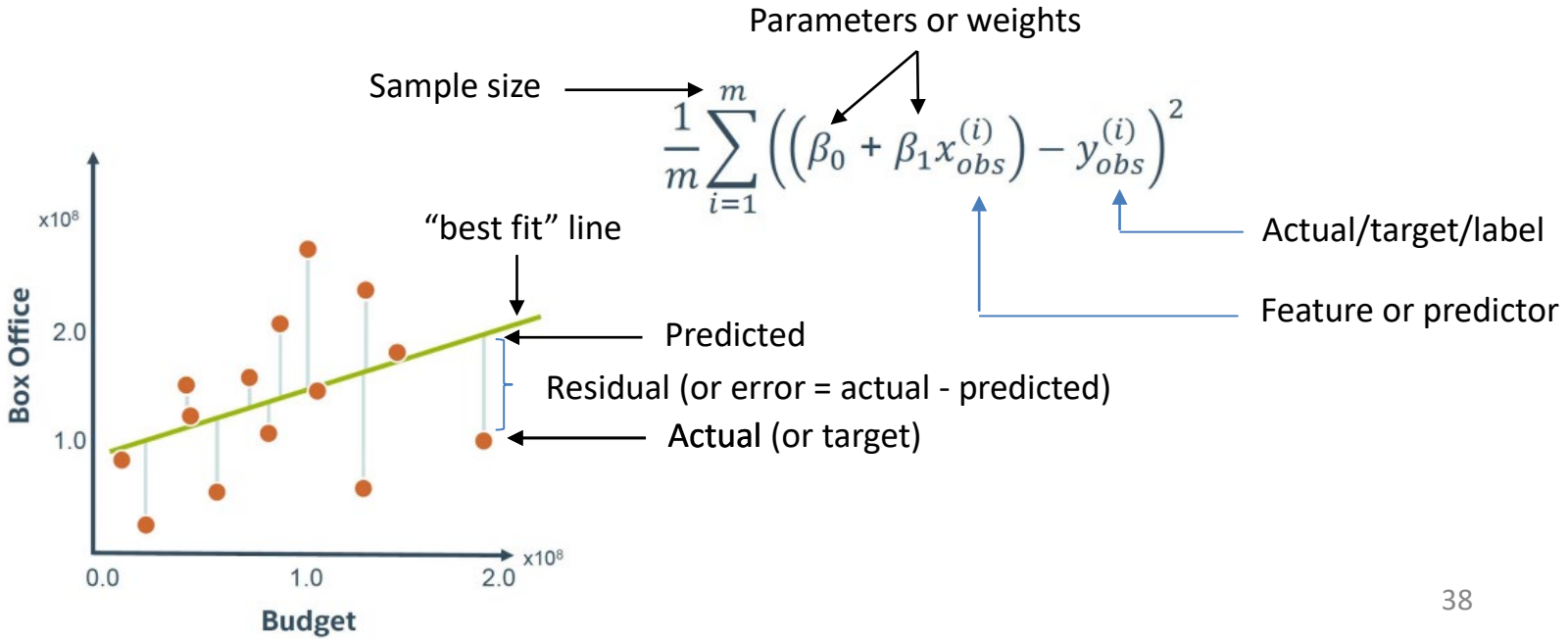
Assessing the Goodness-of-Fit

- After you have fit a linear model using regression analysis, you need to determine how well the model fits the data. There are several common functions used:
  - Mean squared error
  - Mean absolute error
  - R-squared (or Coefficient of Determination)



Mean Square Error

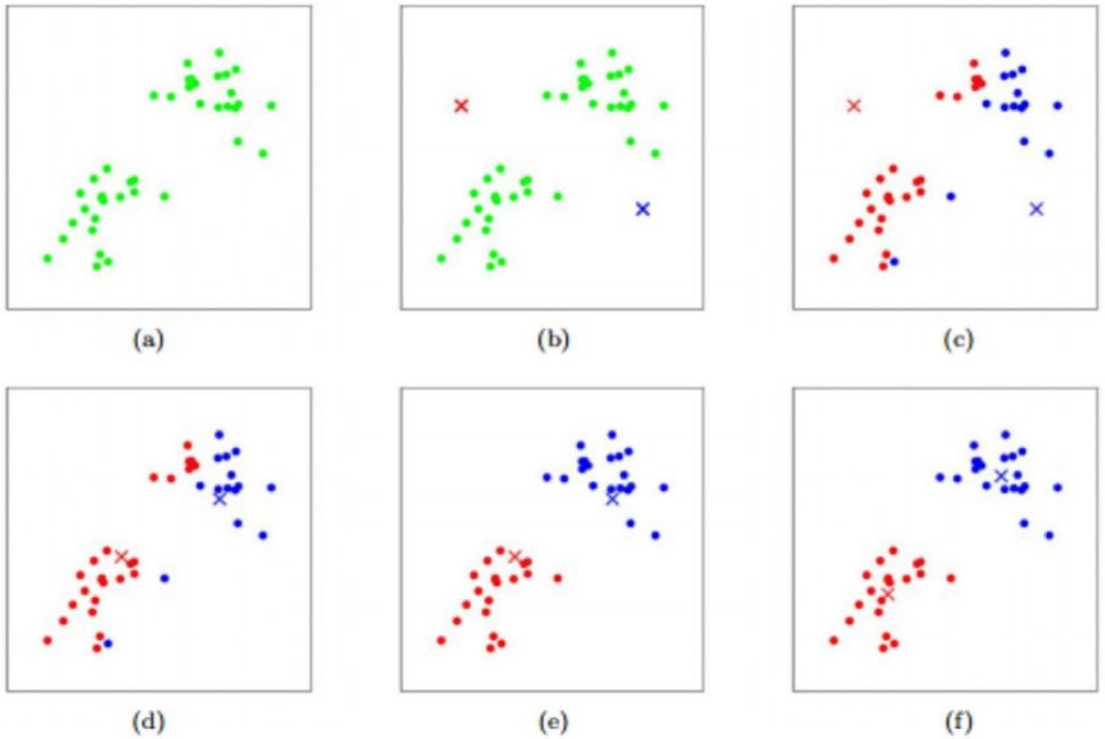
- Mean Square Error (MSE) is the common loss function to measure how good is the linear regression model.
- Machine Learning aims to minimize the MSE to find the best linear regression model.



Clustering Algorithms

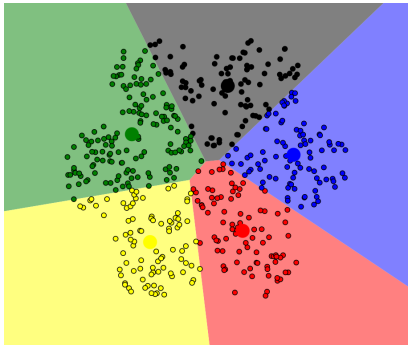
- K-Means Clustering
- Hierarchical Agglomerative Clustering

K-Mean Algorithm



Individual Exercise: K-Mean Algorithm Simulation

- Go to the URL https://www.naftaliharris.com/blog/visualizing-k-means-clustering/
- Click on the box ‘I’ll Choose’
- Pick any problem image
- Decide the number of centroids
- Then randomly place the centroids on the image
- Click ‘Go’ followed by ‘Reassign Points’ until the solution converges
- Can you explain how the algorithm works and what are its strength and weaknesses?



Quiz

What is the main difference between classical computer programming and machine learning?

Classical: Explicit programming is used to solve problems
ML: allows computers to learn and infer from data

Quiz

How many different types/categories of ML are there?

Three types of ML: Supervised Learning, Unsupervised Learning, and Reinforcement Learning

Quiz

Which of the following is most suitable for supervised learning?

A. Identifying birds in an image

B. Grouping people into smaller groups based on buying habits

C. Cleaning robot should keep working or go back for charging

D. Classify animals in their classes which are not known

Quiz

Which of the following is TRUE?

I. Bagging combines multiple weak classifiers to create a strong classifier
II. Boosting focuses on sequentially improving the performance of a single weak classifier
III. Bagging sequentially improves the performance of a single weak classifier
IV. Boosting combines multiple weak classifiers to create a strong classifier

A) I only
B) IV only
C) I and II
D) III and IV

End of Chapter 2