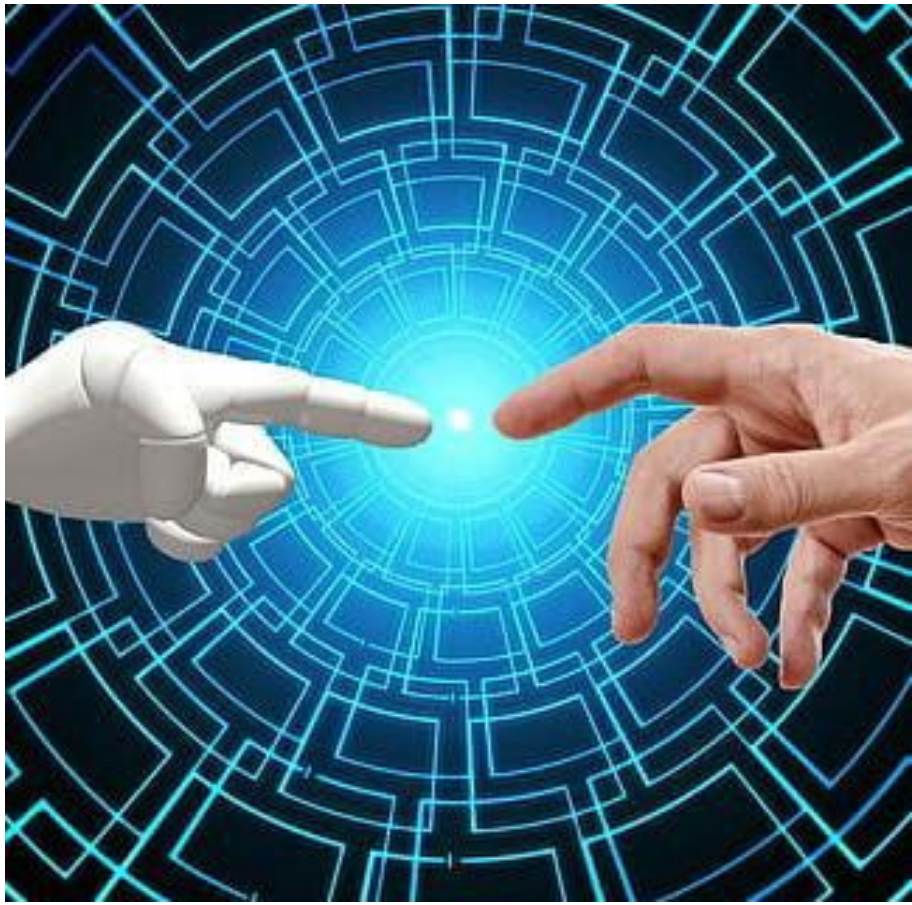
Artificial Intelligence (AI) for Engineering

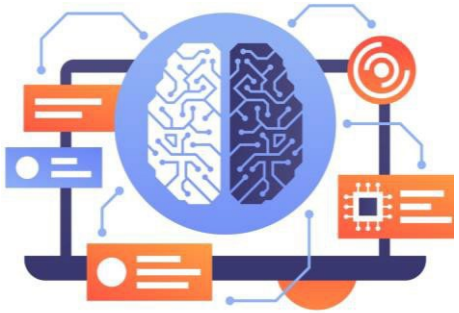


COS40007

Dr. Afzal Azeem Chowdhary

Lecturer, SoCET, Swinburne University of Technology

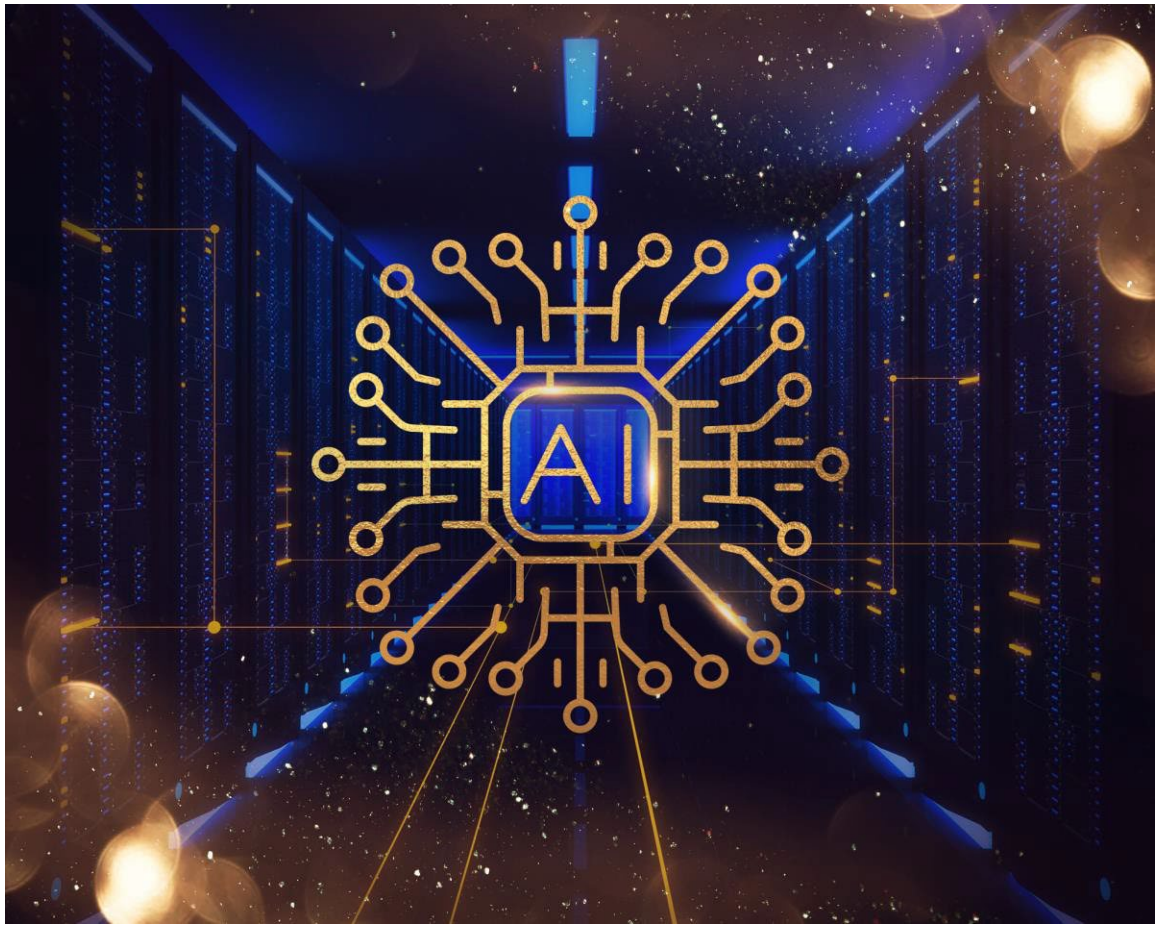
Seminar 3: 18th March 2025



SWINBURNE UNIVERSITY OF TECHNOLOGY

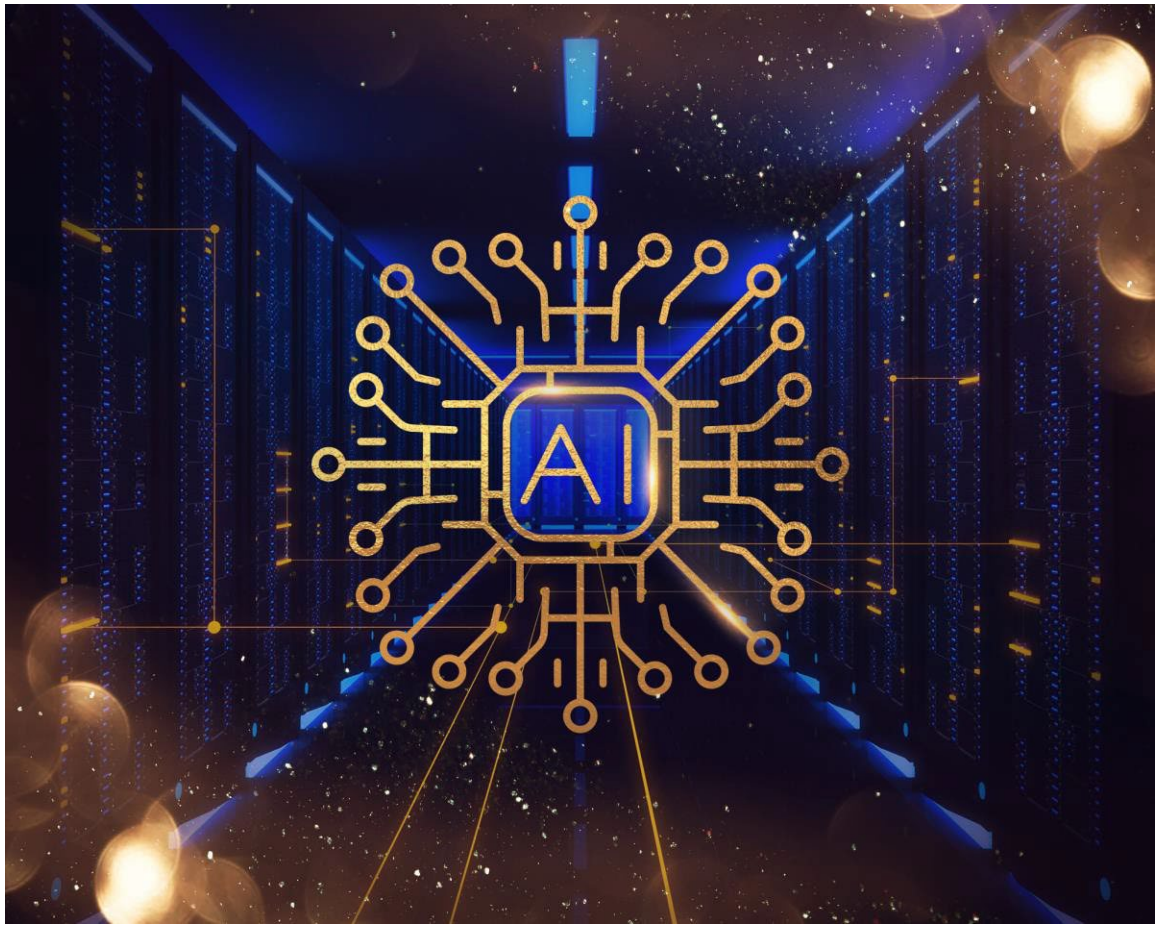
Overview

- Types of Machine Learning
- Supervised vs Unsupervised
- Classification vs Regression
- Machine Learning Algorithms
- Feature Deletion
- Dimensionality Reduction
- Machine Learning problem
- Key Data Science concepts



Required Reading

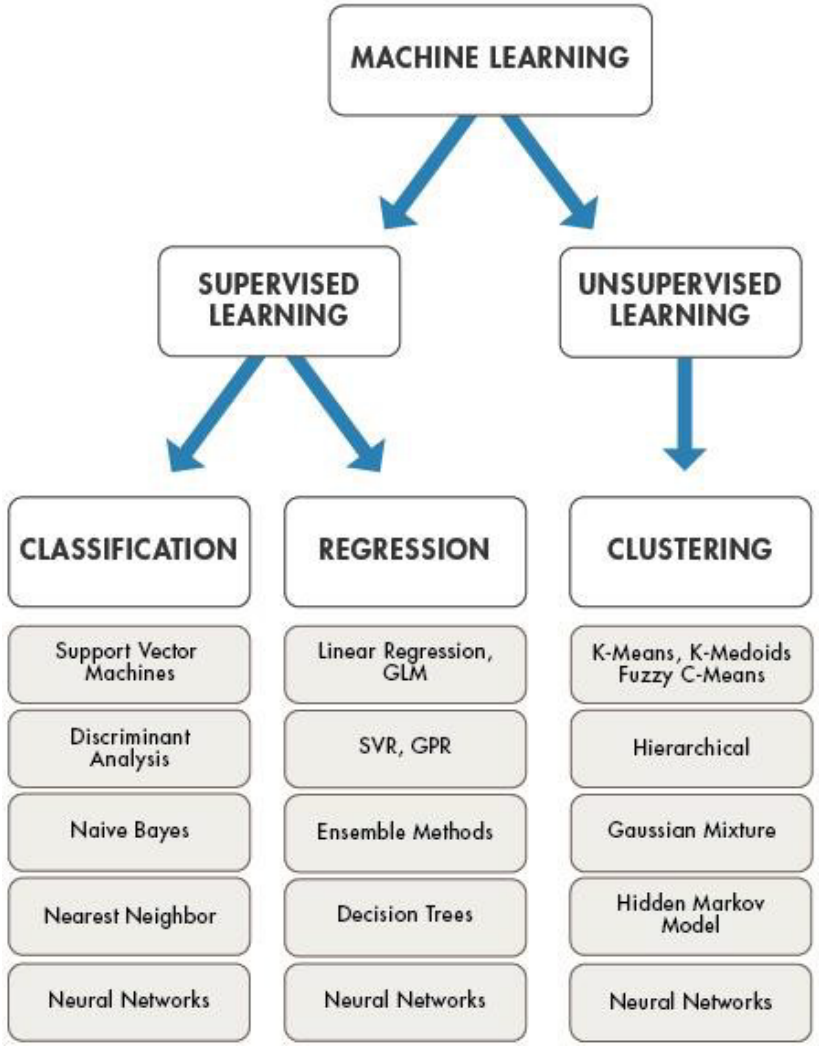
- Chapter 2,3,5 and 6 of “Applied Machine Learning and AI for Engineers”
- Chapter 5 and 6 of “Machine Learning with Pytorch and Scikit-Learn”



At the end of this you should be able to

- Understand the steps you need to complete to develop machine learning models.
- Understand how to perform data pre-processing.
- Understand model training and development process.

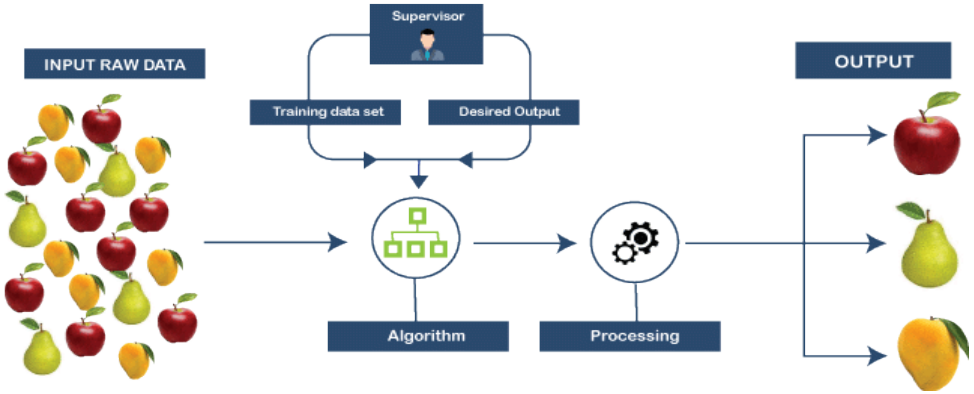
Type of machine learning models



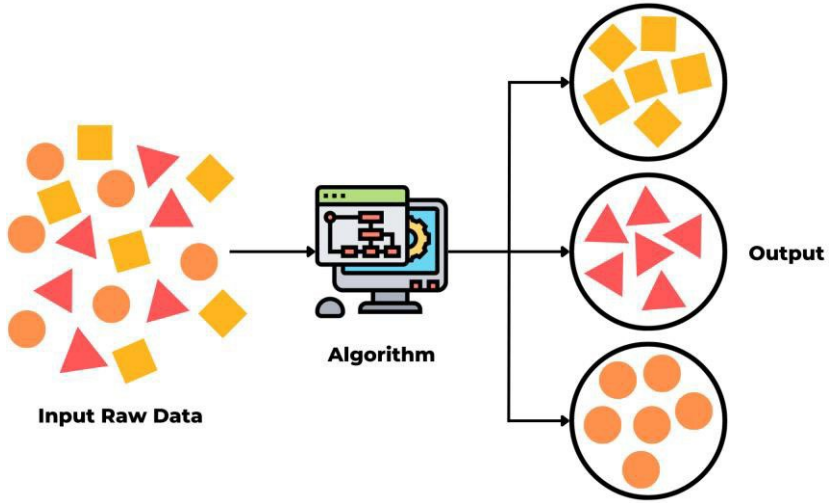
Supervised vs Unsupervised Learning

- Most ML algorithms falls into 2 categories
  - Supervised Learning: You train them with labelled data so that they can take future inputs and predict what the labels will be. E.g., human activity recognition using body worn sensors
  - Unsupervised learning: provide insights into existing data, or to group data into categories and categorize future inputs accordingly. Data that it consumes doesn't have to be labelled. E.g. group customer based on purchase behaviour.
  - Most of the ML models in use today are supervised learning models

Supervised Learning



Unsupervised Learning



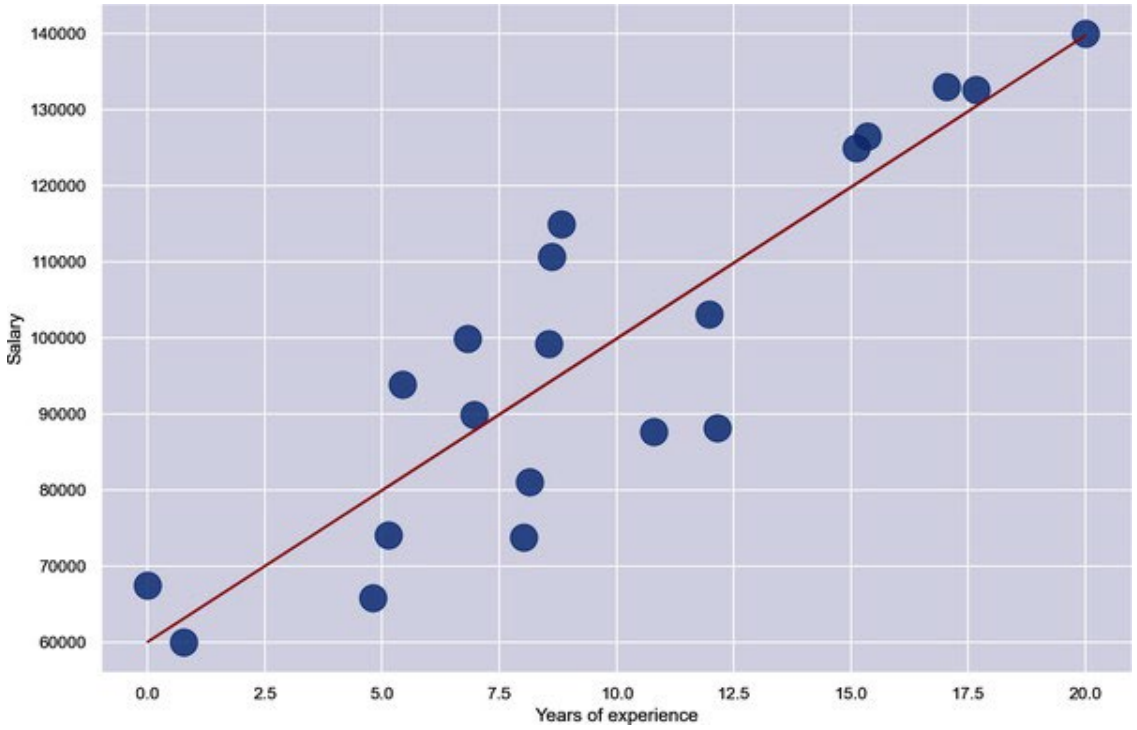
Classification and Regression algorithms

- Regression algorithms are used to determine continuous values mostly for predicting and forecasting values. E.g., Analysing raw material influence on product quality. Example of regression algorithms are - Liner regression.

- Classification: Classification algorithms are used to predict or classify the distinct values. E.g., True/False, good/average/bad, Human activity type (walking, standing, sitting). Example of classification algorithm are – Support Vector Machine, k-nearest neighbour.

- Some algorithms can be used both for regression and classification such as Decision Trees, Neural Network

Regression Analysis: Example – Linear Regression



Predicting salary of a plant worker with 10 years of experience

y = mx + b (y – salary value, x – years of experience).

The goal is to find ***m*** and ***b***. where ***m*** is the slope of the line and ***b*** is where the line intersects the y-axis. Use of mathematical formula (e.g. least square regression) to find ***m*** and ***b*** and using that compute ***y***.

Classification: Example

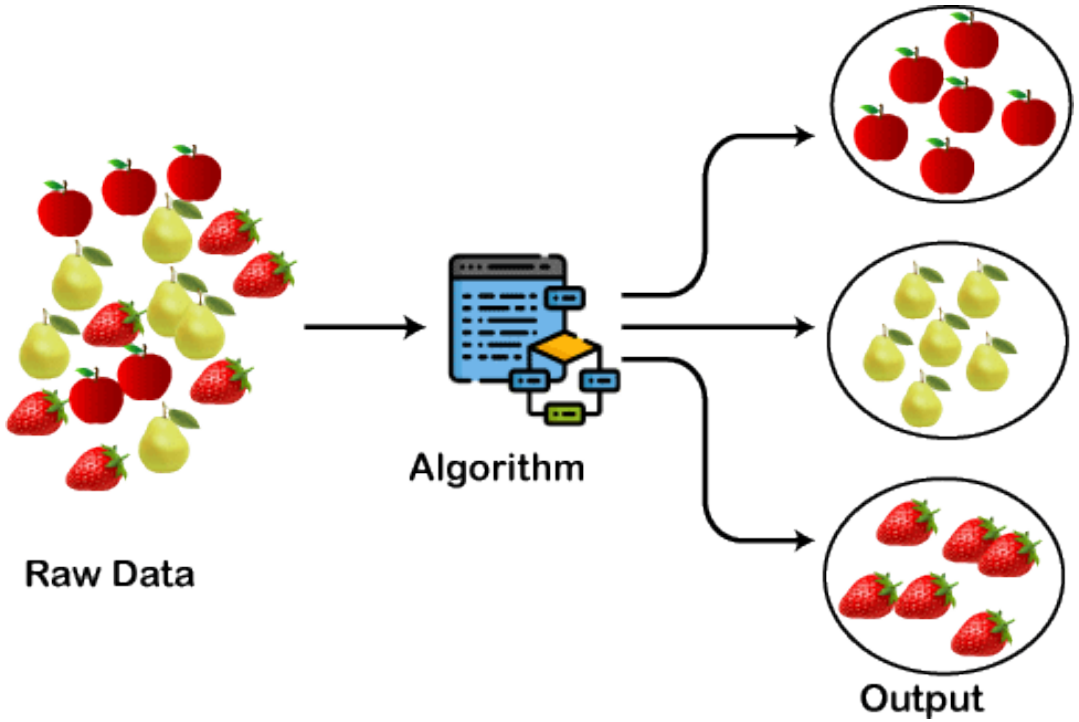
Prediction of whether or not an object is a mine or a rock, given the strength of sonar returns at different angles.

```
1 0.0200,0.0371,0.0428,0.0207,0.0954,0.0986,0.1539,0.1601,0.3109,0.2111,0.1609,0.1582,0.2238,0.0645,0.0660,0.227
2 3,0.3100,0.2999,0.5078,0.4797,0.5783,0.5071,0.4328,0.5550,0.6711,0.6415,0.7104,0.8080,0.6791,0.3857,0.1307,0.2
3 604,0.5121,0.7547,0.8537,0.8507,0.6692,0.6097,0.4943,0.2744,0.0510,0.2834,0.2825,0.4256,0.2641,0.1386,0.1051,0
4 .1343,0.0383,0.0324,0.0232,0.0027,0.0065,0.0159,0.0072,0.0167,0.0180,0.0084,0.0090,0.0032,R
5 0.0453,0.0523,0.0843,0.0689,0.1183,0.2583,0.2156,0.3481,0.3337,0.2872,0.4918,0.6552,0.6919,0.7797,0.7464,0.944
  4,1.0000,0.8874,0.8024,0.7818,0.5212,0.4052,0.3957,0.3914,0.3250,0.3200,0.3271,0.2767,0.4423,0.2028,0.3788,0.2
  947,0.1984,0.2341,0.1306,0.4182,0.3835,0.1057,0.1840,0.1970,0.1674,0.0583,0.1401,0.1628,0.0621,0.0203,0.0530,0
  .0742,0.0409,0.0061,0.0125,0.0084,0.0089,0.0048,0.0094,0.0191,0.0140,0.0049,0.0052,0.0044,R
  0.0262,0.0582,0.1099,0.1083,0.0974,0.2280,0.2431,0.3771,0.5598,0.6194,0.6333,0.7060,0.5544,0.5320,0.6479,0.693
  1,0.6759,0.7551,0.8929,0.8619,0.7974,0.6737,0.4293,0.3648,0.5331,0.2413,0.5070,0.8533,0.6036,0.8514,0.8512,0.5
  045,0.1862,0.2709,0.4232,0.3043,0.6116,0.6756,0.5375,0.4719,0.4647,0.2587,0.2129,0.2222,0.2111,0.0176,0.1348,0
  .0744,0.0130,0.0106,0.0033,0.0232,0.0166,0.0095,0.0180,0.0244,0.0316,0.0164,0.0095,0.0078,M
  0.0100,0.0171,0.0623,0.0205,0.0205,0.0368,0.1098,0.1276,0.0598,0.1264,0.0881,0.1992,0.0184,0.2261,0.1729,0.213
  1,0.0693,0.2281,0.4060,0.3973,0.2741,0.3690,0.5556,0.4846,0.3140,0.5334,0.5256,0.2520,0.2090,0.3559,0.6260,0.7
  340,0.6120,0.3497,0.3953,0.3012,0.5408,0.8814,0.9857,0.9167,0.6121,0.5006,0.3210,0.3202,0.4295,0.3654,0.2655,0
  .1576,0.0681,0.0294,0.0241,0.0121,0.0036,0.0150,0.0085,0.0073,0.0050,0.0044,0.0040,0.0117,R
  0.0762,0.0666,0.0481,0.0394,0.0590,0.0649,0.1209,0.2467,0.3564,0.4459,0.4152,0.3952,0.4256,0.4135,0.4528,0.532
  6,0.7306,0.6193,0.2032,0.4636,0.4148,0.4292,0.5730,0.5399,0.3161,0.2285,0.6995,1.0000,0.7262,0.4724,0.5103,0.5
  459,0.2881,0.0981,0.1951,0.4181,0.4604,0.3217,0.2828,0.2430,0.1979,0.2444,0.1847,0.0841,0.0692,0.0528,0.0357,0
  .0085,0.0230,0.0046,0.0156,0.0031,0.0054,0.0105,0.0110,0.0015,0.0072,0.0048,0.0107,0.0094,R
```

SWINBURNE UNIVERSITY OF TECHNOLOGY

Clustering

- Unsupervised learning frequently employs a technique called Clustering. The purpose of clustering is to group data by similarity.

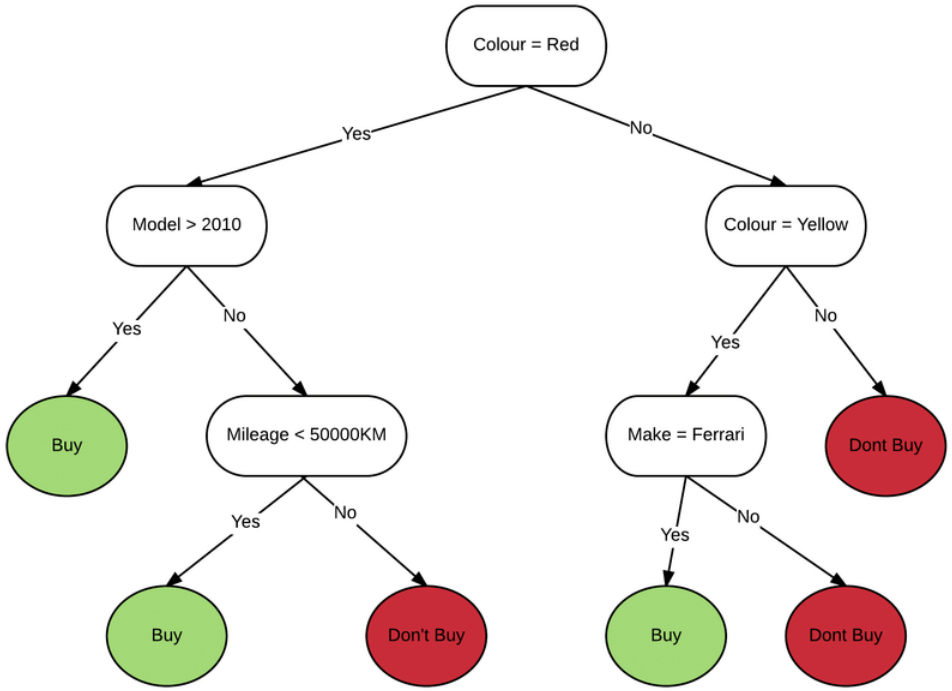


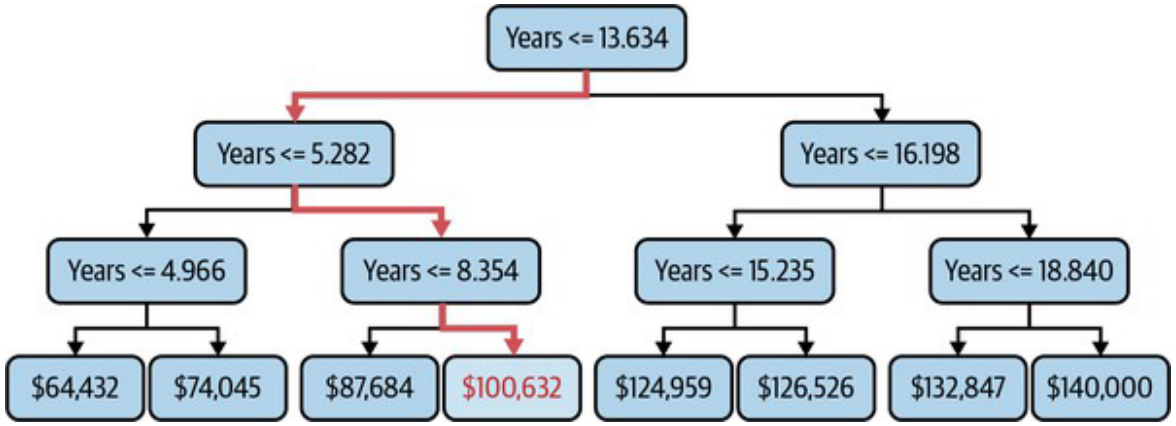
Popular Machine Learning Algorithms

Decision Tree (DT)

A decision tree is a tree structure that predicts an outcome by answering a series of questions. Most decision trees are binary trees, in which case the questions require simple yes-or-no answers.

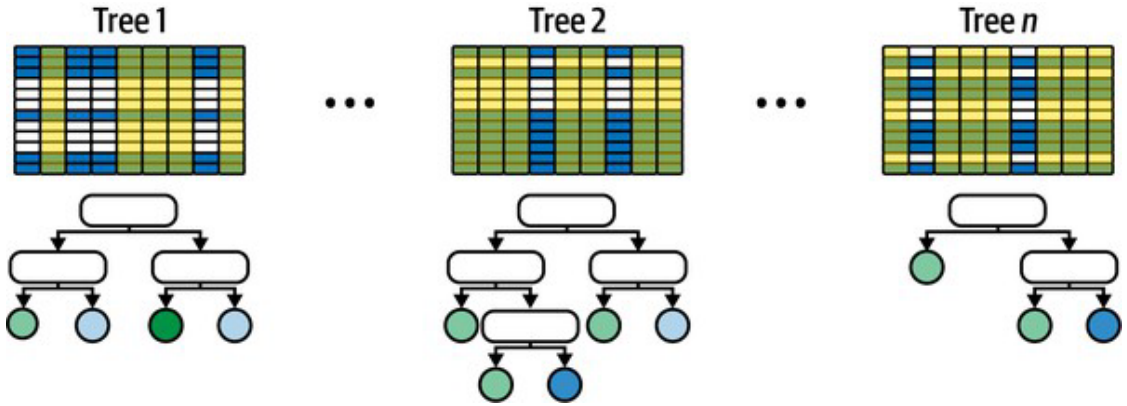
- Can be used for both regression and classification
- Work well with non-liner data
- Downside: overfitting issue (too tightly fit to training data and might not generalised well)
- That means it won't be as accurate when it's asked to make predictions with data it hasn't seen before





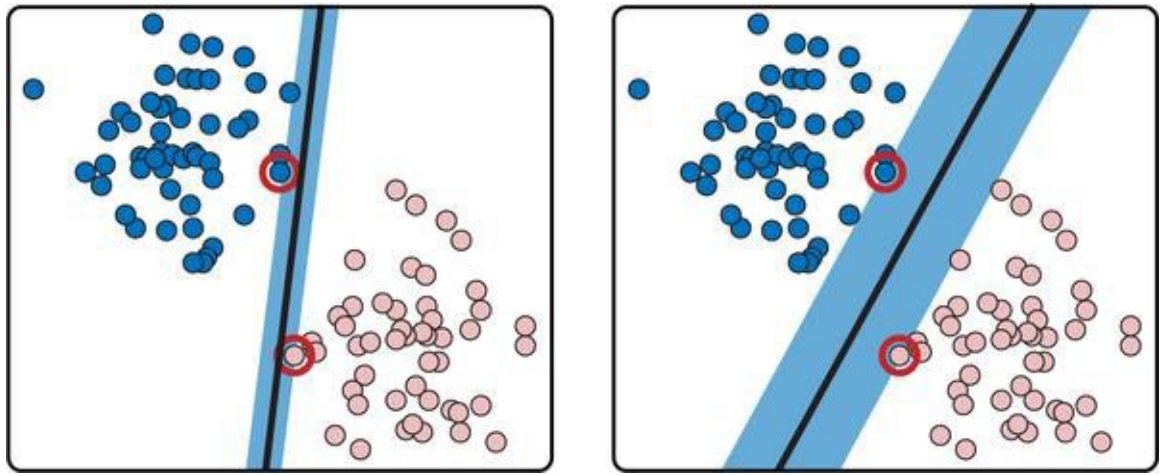
Random forests

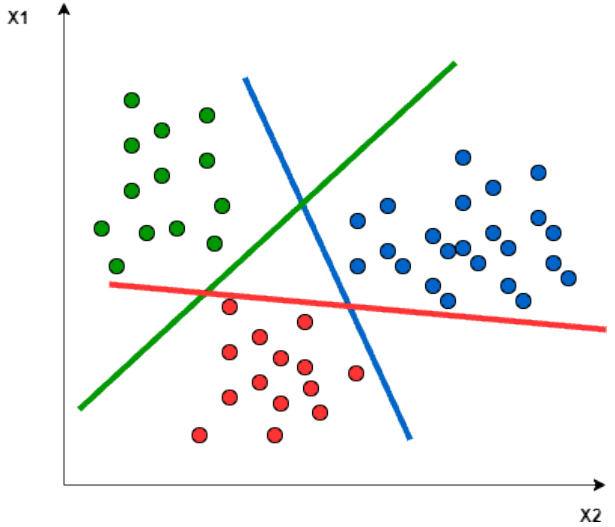
- A random forest is a collection of decision trees (often hundreds of them), each trained differently on the same data
- Each tree is trained on randomly selected rows in the dataset, and branching is based on columns that are randomly selected at every split.
- The model can't fit too tightly to the training data because every tree trains on a different subset of the data
- When the model makes a prediction, it runs the input through all the decision trees and averages the result.
- No individual tree in a random forest can predict an outcome with a great deal of accuracy. But put all the trees together and average the results and they often outperform other models



Support Vector Machine (SVM)

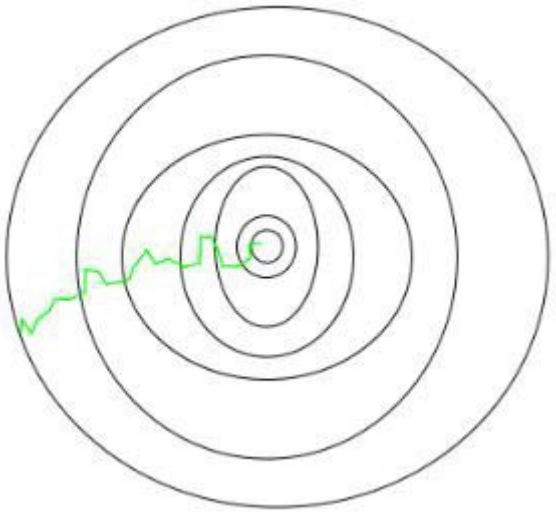
- Most often used to solve the classification problem.
- Based on a mathematical model, find a decision boundary that cleanly separates the classes.
- SVMs do this by finding a hyperplane in higher-dimensional space that allows them to distinguish between different classes with the greatest certainty possible.

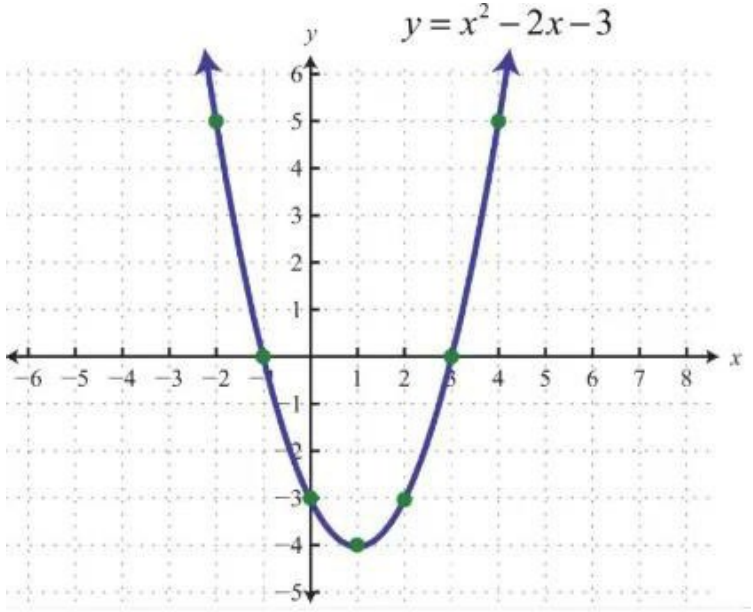




Stochastic Gradient Descent (SGD)

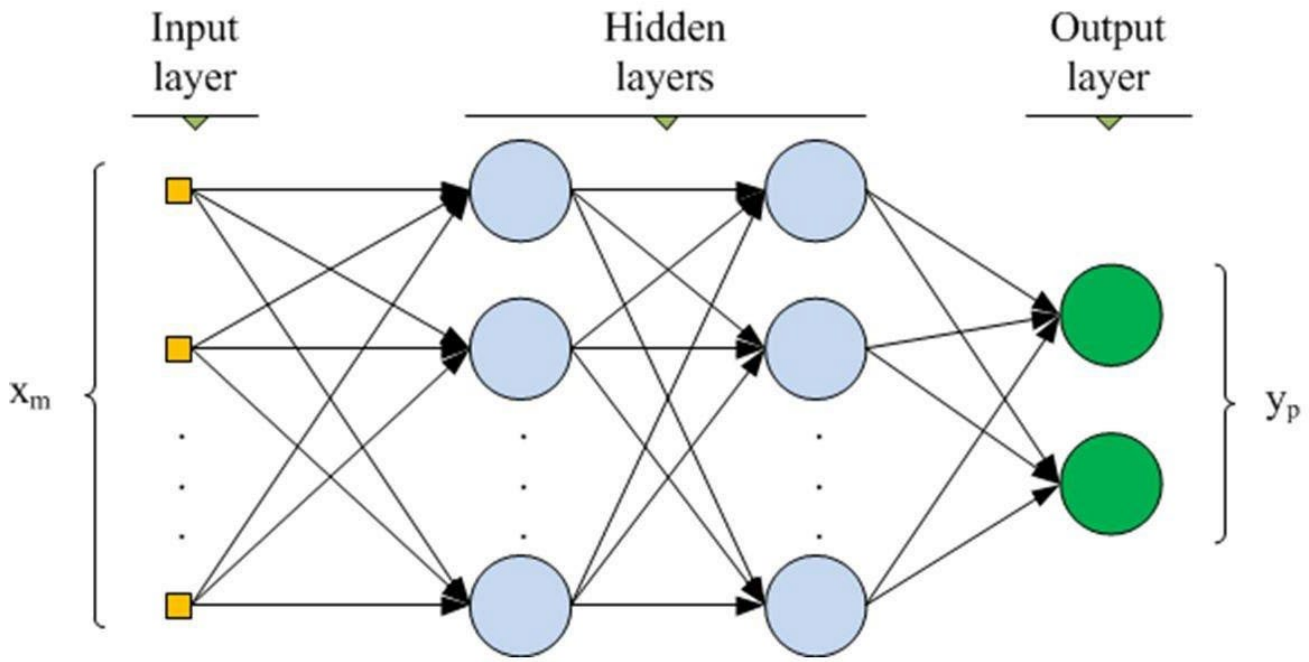
- Gradient descent means descending a slope to reach the lowest point on that surface
- Instead of using the entire dataset for each iteration, only a single random training example (or a small batch) is selected to calculate the gradient and update the model parameters.
- Computationally efficient
- SGD has been successfully applied to large-scale and sparse machine learning problems often encountered in text classification and natural language processing.





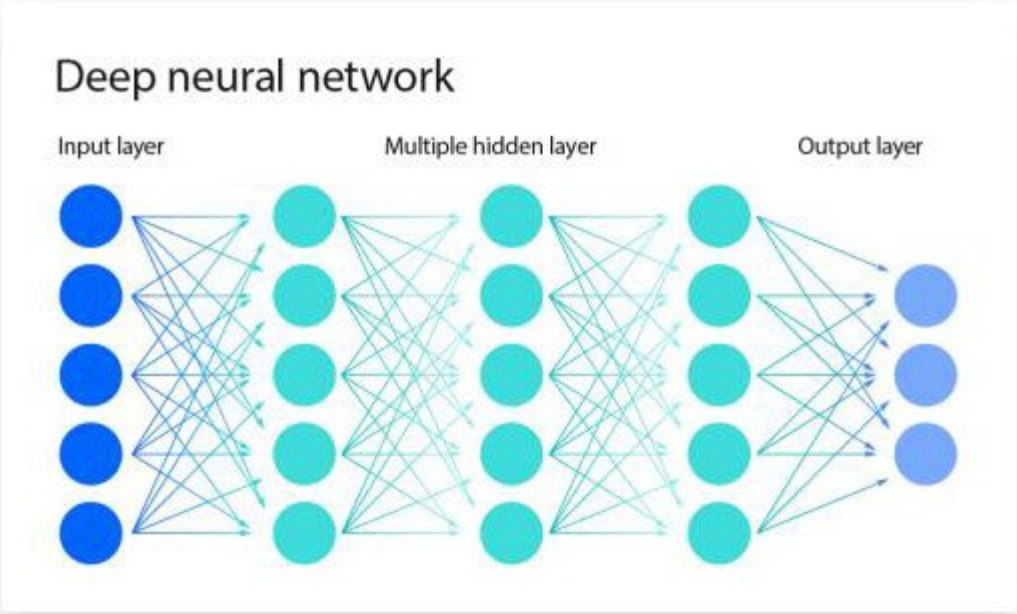
Neural Network (NN)

- Take inspiration from the learning process occurring in human brains
- A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output.
- The depth of the network is the number of layers; the width is the number of neurons in each layer, which can be different for every layer
- Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer.



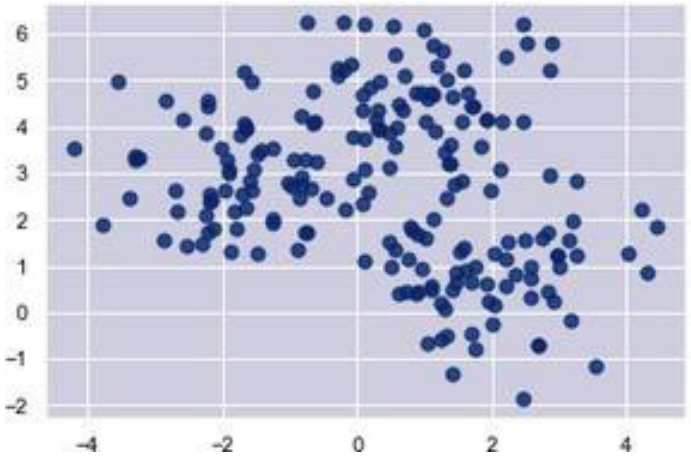
Deep Neural Network

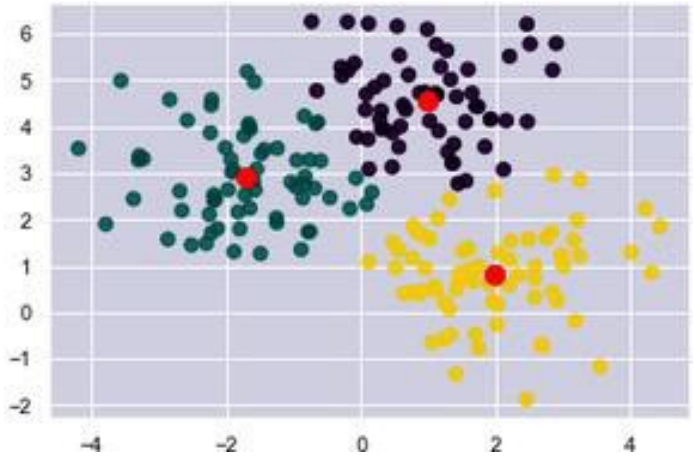
- Extension of Neural Network (high number of hidden layers)
- Computers are now more powerful (have GPUs, TPUS)
- Many variants of Deep Neural Network are as follows:
  - Convolutional neural networks (CNNs) – used in image classification
  - Recurrent neural networks (RNNs) used in handwriting recognition and natural language processing (NLP)
  - Generative adversarial networks, or GANs, enable computers to create images, music, and other content



k-means clustering

- The purpose of clustering is to group data by similarity
- The most popular clustering algorithm is *k*-means clustering, which takes n data samples and groups them into *k* clusters, where *k* is a number you specify.
- Grouping is performed using an iterative process that computes a centroid for each cluster and assigns samples to clusters based on their proximity to the cluster centroids.
- If the distance from a particular sample to the centroid of cluster 1 is 2.0 and the distance from the same sample to the centre of cluster 2 is 3.0, then the sample is assigned to cluster 1





Problem in Machine Learning: Overfitting

- Overfitting occurs when the model cannot generalize and fits too closely to the training dataset instead.
- The training data size is too small and does not contain enough data samples to accurately represent all possible input data values.
- The training data contains large amounts of irrelevant information, called noisy data.
- The model trains for too long on a single sample set of data.
- The model complexity is high, so it learns the noise within the training data.
- A high error rate in the testing data indicates overfitting

Preventing Overfitting

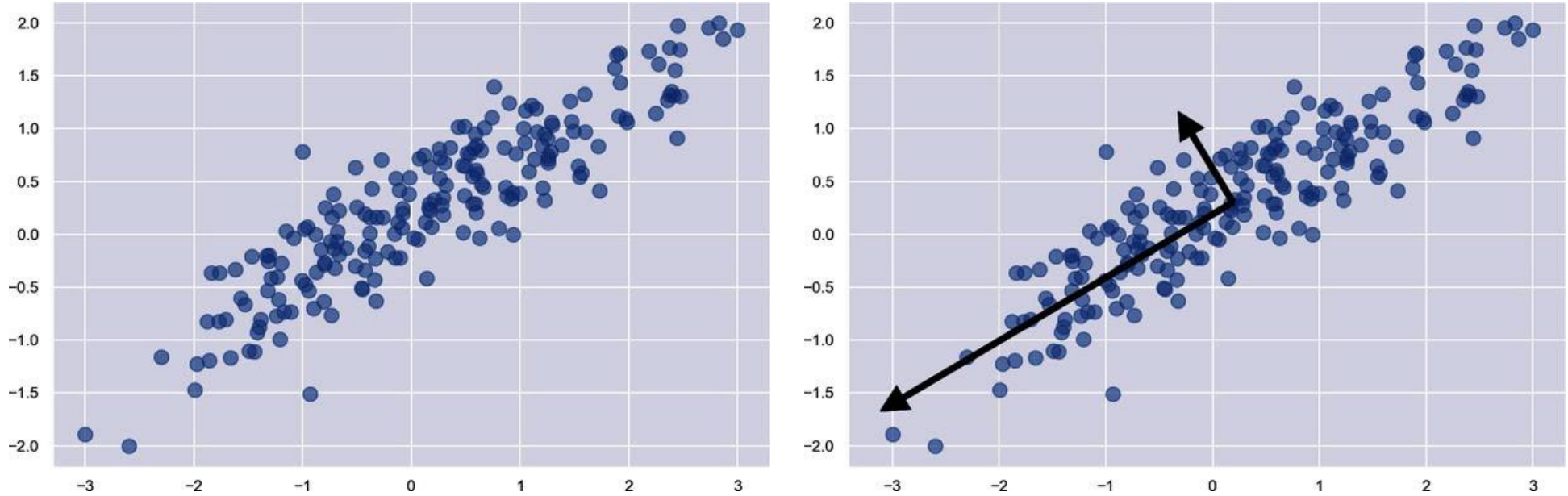
- Early stopping pauses the training phase before the machine learning model learns the noise in the data; however, it is difficult to understand the right time.
- Feature selection/ pruning—identifies the most important features within the training set and eliminates irrelevant ones.
- Grading features based on importance.
- Dimensionality Reduction.

Feature selection

- Find the best set of features to build optimised model
- Type of feature selection method
  - Correlation
  - Filter method
  - Statistical method
  - Feature Ranking
  - Trees

Dimensionality Reduction: Principal Component Analysis (PCA)

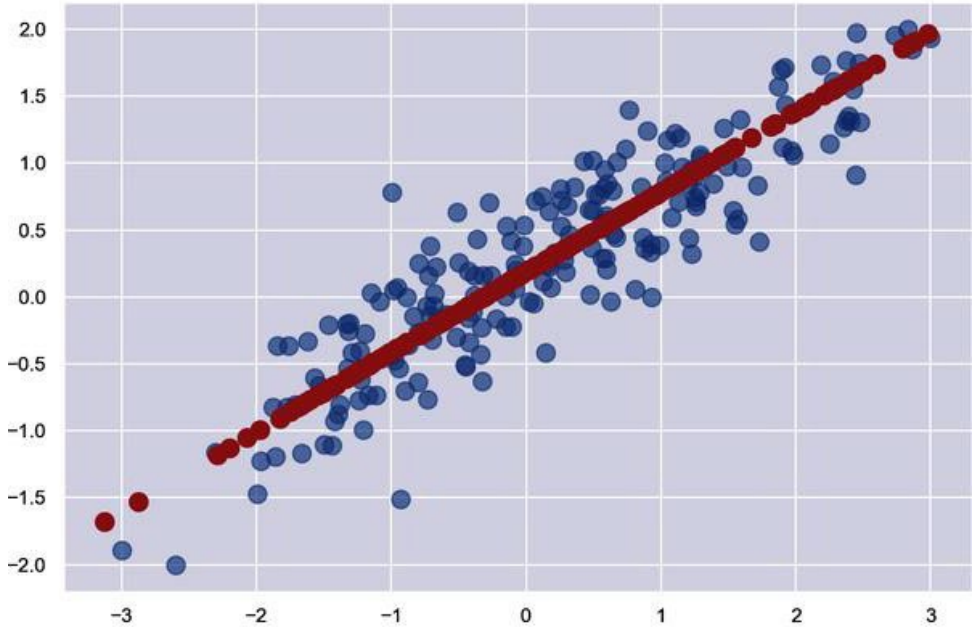
- A dimensionality reduction technique that reduces the number of dimensions in a dataset without sacrificing a commensurate amount of information
- You could take a dataset with 1,000 columns, use PCA to reduce it to 100 columns, and retain 90% or more of the information in the original dataset



SWINBURNE UNIVERSITY OF TECHNOLOGY

Principal Component Analysis (PCA)

- Reducing high-dimensional data to two or three dimensions so that it can be plotted and explored
- Reducing the number of dimensions in a dataset and then restoring the original number of dimensions, which finds application in anomaly detection and noise filtering
- Anonymizing datasets so that they can be shared with others without revealing the nature or meaning of the data



SWINBURNE UNIVERSITY OF TECHNOLOGY

Type of machine learning problem

- Prediction / Predictive analysis (Predict an activity using a classifier)
- Prescriptive analysis (Recommend a course of action)
- Forecasting (Especially in time-series data)
- Anomaly detection
- Optimisation

Key concepts

- Data Science: [mathematics, statistics, machine learning, AI, analytics]
- Data Engineering: Convert raw data to usable format
- Data mining: Use machine learning and statistical analysis to uncover patterns and other valuable information in large data sets
- Data analytics: Process of analysing raw data to find trends and answer questions

WEKA

A robust data mining tool (https://ml.cms.waikato.ac.nz/weka)

Learn, Practice and Enjoy the AI journey