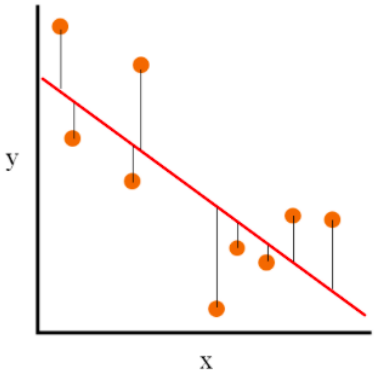
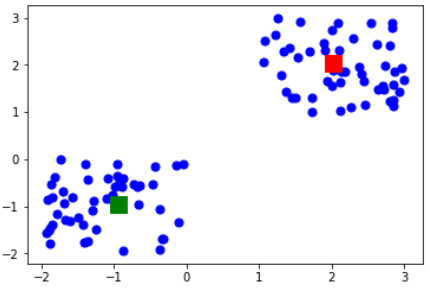
Linear Regression

Linear regression is supervised learning because it learns from previous data that has been labelled(independent and dependent variables). Simple linear regression has one independent variable and multiple linear regression has more than one independent variable. The algorithm finds a linear line/correlation between the data. It is best used for making predictions. It is used by companies such as Amazon to predict what customers will buy in the future based on past purchases and Walmart to predict what products will be popular in different regions. It can be used to predict how much cash a business will have in the future, how hot weather affects ice cream sales, stock predictions based on previous stock prices and sales.



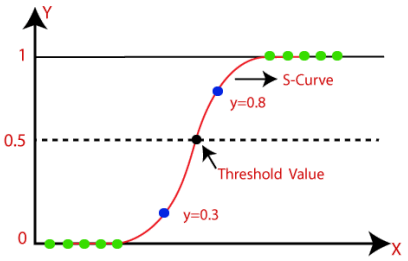
K-Means

K-means is unsupervised learning because there is only input data that hasn't been labelled so the algorithm speculates the outcome. K-means groups data together in clusters based on similarity to try and find patterns. It is most useful when data needs grouping together, for example crime data can be used to find out which specific crimes are happening in specific areas of cities.



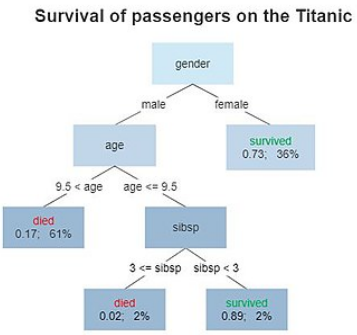
Logistic Regression

Logistic regression is supervised learning. It shows the probability for a classification problem with only two potential outcomes (e.g. Yes/No, True/False). It is shown as an S curve on a graph between the values of 0 and 1 with 0.5 being the threshold value. Examples of where it can be used include if a political candidate will win or not and whether a student will pass or fail an exam.



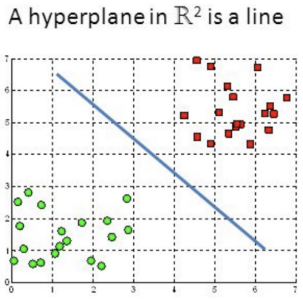
Decision Tree

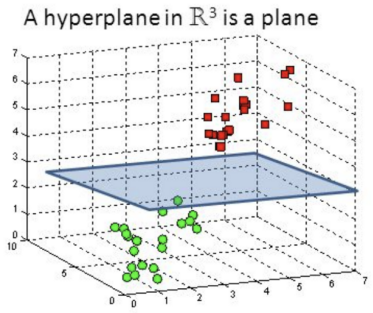
Decision trees are a supervised type of machine learning. They look like flowcharts with if/else conditions. Input variables are the branches and a prediction of the value of target variables are the leaves. It is possible to have a regression and classification tree. They are used for predictive modelling. They are used in retail to decide upon changing prices or product models.



SVM (Support Vector Machine)

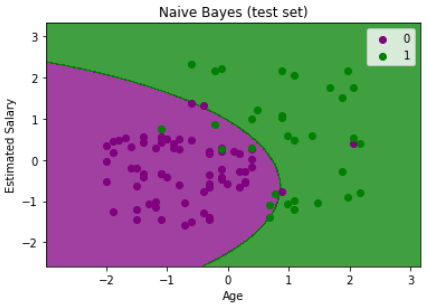
This is a supervised learning model. Where data has been labelled as one of two categories, SVM can then decide which category new data should go into. A hyperplane is used on a graph as a decision boundary to classify the data points. It is used in face detection to classify whether parts of the image are a face or not.





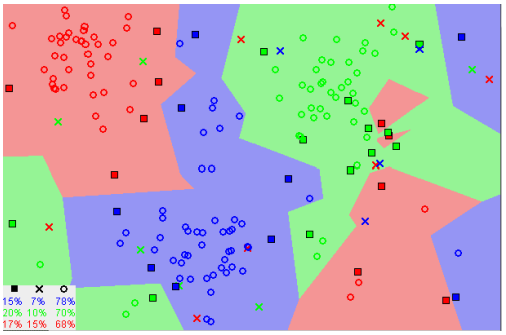
Naive Bayes

Naive Bayes is a supervised learning algorithm. It uses the Bayes theorem for probability to predict classification of data. Each feature within a class is unrelated to any other feature. It is best used for very large datasets. It is a good option for real time predictions and recommendation systems. Some email systems use this to detect spam.



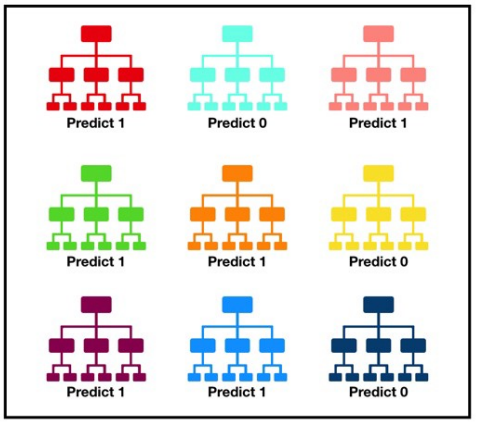
KNN (K- Nearest Neighbours)

KNN is a supervised algorithm. It assumes that similar things are near each other and calculates the distance between the points. It is most useful where it is necessary to have a solution based on identifying similar objects so it is good for recommendation systems but it can be a slow algorithm the larger the datasets. It can be used for credit card fraud detection to detect outliers.



Random Forest

This is supervised. A random forest is made up of multiple decision trees. Each one picks a class prediction and the class that comes out with the most votes is the overall prediction. It is used in banking to predict who will pay back debt on time. It can also analyse patient medical history to detect diseases.



Tally: Six 1s and Three 0s
Prediction: 1

Sources used:

https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

https://vitalflux.com/linear-regression-real-life-example/

https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

https://dzone.com/articles/10-interesting-use-cases-for-the-k-means-algorithm

https://www.javatpoint.com/logistic-regression-in-machine-learning

https://en.wikipedia.org/wiki/Decision_tree_learning

https://en.wikipedia.org/wiki/Support-vector_machine

https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

https://data-flair.training/blogs/applications-of-svm/

https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761#:~:text=Summary-,The%20k%2Dnearest%20neighbors%20(KNN)%20algorithm%20is%20a%20simple,that%20data%20in%20use%20grows.

https://towardsdatascience.com/k-nearest-neighbors-knn-algorithm-23832490e3f4

https://towardsdatascience.com/understanding-random-forest-58381e0602d2#:~:text=The%20random%20forest%20is%20a,that%20of%20any%20individual%20tree.