**CLASSIFICATION MODELS**

1. **Support Vector Machines (SVM):**

Support vector machine finds the best way to classify the data based on the position in relation to a border between positive class and negative class. This border is known as the hyperplane which maximize the distance between data points from different classes. Similar to decision tree and random forest, support vector machine can be used in both classification and regression, SVC (support vector classifier) is for classification problem.

A diagram of a graph

Description automatically generated

1. **Random Forest:**

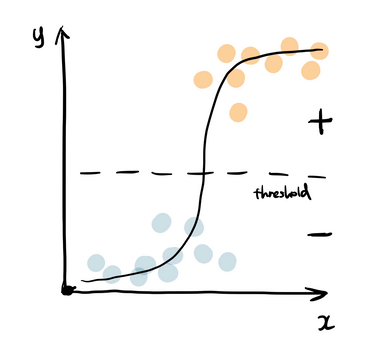
A random forest is a collection of decision trees. It is a common type of ensemble method which aggregate results from multiple predictors. Random forest additionally utilizes bagging technique that allows each tree trained on a random sampling of original dataset and takes the majority vote from trees. Compared to decision tree, it has better generalization but less interpretable, because of more layers added to the model.

A diagram of a company

Description automatically generated

1. **Logistic Regression:**

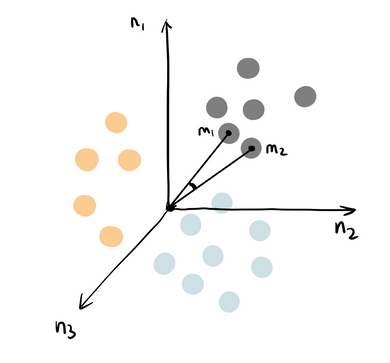
Logistics regression uses sigmoid function to return the probability of a label. It is widely used when the classification problem is binary — true or false, win or lose, positive or negative.



The sigmoid function generates a probability output. By comparing the probability with a pre-defined threshold, the object is assigned to a label accordingly.

1. **K-Nearest Neighbors (KNN):**

K-Nearest Neighbors algorithm as representing each data point in a n dimensional space — which is defined by n features. And it calculates the distance between one point to another, then assigns the label of unobserved data based on the labels of nearest observed data points. KNN can also be used for building recommendation systems, check out my article on “Collaborative Filtering for Movie Recommendation” if you are interested in this topic.



1. **Naive Bayes:**

Naive Bayes is based on Bayes’ Theorem — an approach to calculate conditional probability based on prior knowledge, and the naive assumption that each feature is independent to each other. The biggest advantage of Naive Bayes is that, while most machine learning algorithms rely on large amount of training data, it performs relatively well even when the training data size is small. Gaussian Naive Bayes is a type of Naive Bayes classifier that follows the normal distribution.



1. **Decision Tree:**

Decision tree builds tree branches in a hierarchy approach and each branch can be considered as an if-else statement. The branches develop by partitioning the dataset into subsets based on most important features. Final classification happens at the leaves of the decision tree.

A drawing of a diagram

Description automatically generated

1. **Artificial Neural Networks (Deep Learning):**

A substantial majority of heavily integrated processing components known as neurons comprise artificial neural networks, which are computational models that were modeled after biological neural networks.

Pattern recognition or data segmentation are two instances of applications for which an ANN (Artificial Neural network) is configured.

It is equipped to make sense of enormous or perplexing material.

It extracts patterns and detects trends that are too subtle for individuals or other computer technologies to grab on.