SyedBelal1

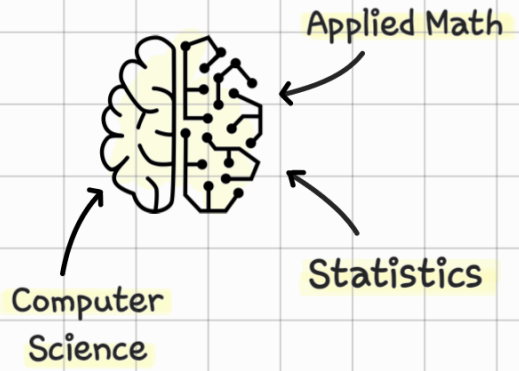
GoSleepBelal1

Machine Learning

Definitions

It provides ability to machine to learn automatically and improve without being explicity programmed.

A computer program is said to learn from experience E with respect to some class of task T and performance P, if it's performance at task T measured by P improves with experience E.



Types of learning

Supervised learning

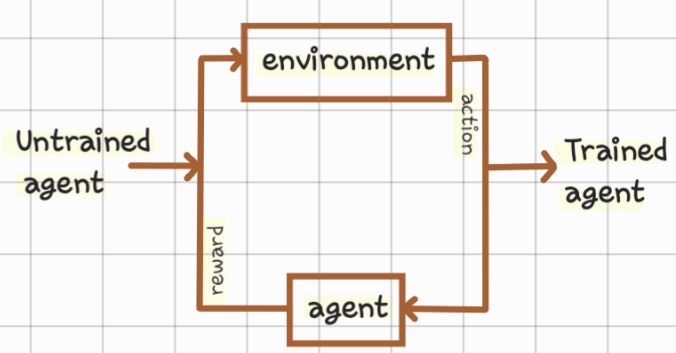
We train or teach the machine using data which is well labelled.

Unsupervised learning

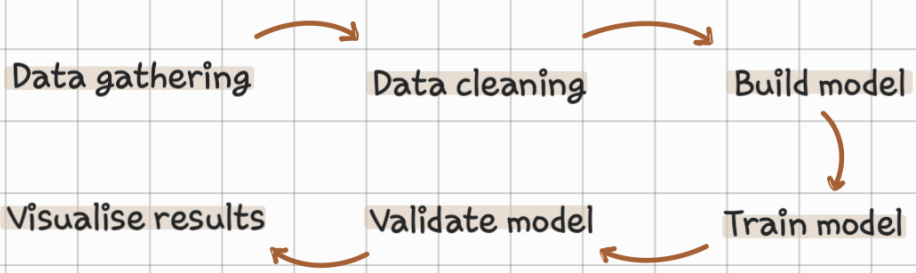
Training the machine using information that is unlabelled and allowing algorithms to act on that information without any guidance

Reinforcement learning

Learning where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the reward which it gets from those actions



Machine learning process

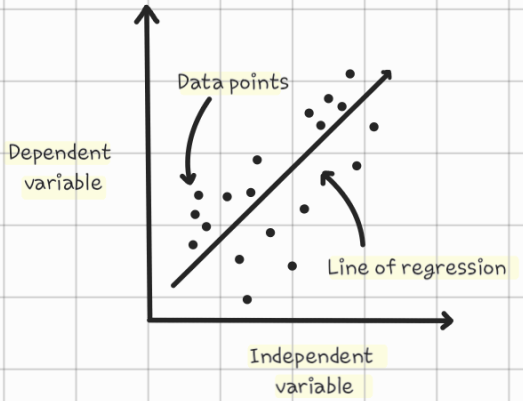


Types of problems

Regression

Predicting one based on one or more independent variables.

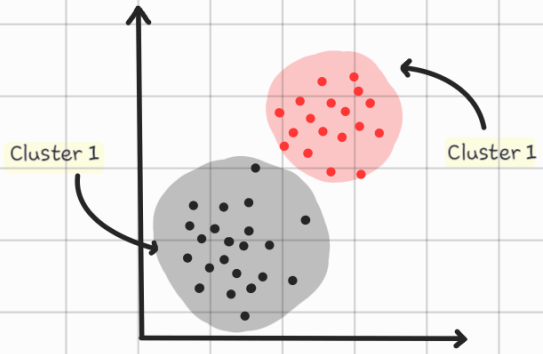
- gives a continuous output
- e.g., stocks prediction, house price estimation.



Clustering

Grouping similiar data based on their similarity.

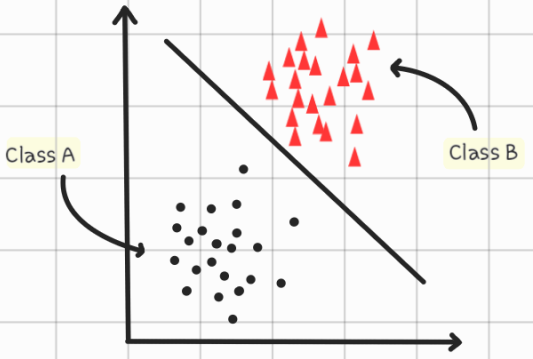
- e.g., image segmentation, customer segmentation.



Classification

Classifying data into multiple classes.

- e.g., Handwritten digits recognition, object detection, etc.
- gives a categorical value as output



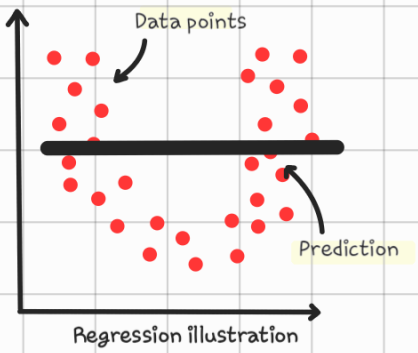
Types of classifiers

- Instance based (single sample based)
- Stream based (batch based)

Properties of model

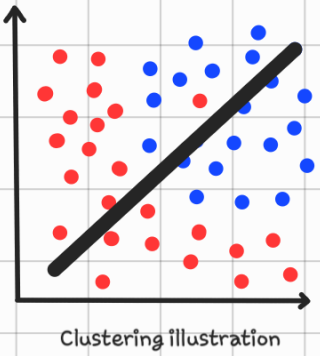
Underfitting

It occurs when the model is too simple and fails to capture the pattern of underlying data, in other words, poor performance models on both training and testing data.



Causes

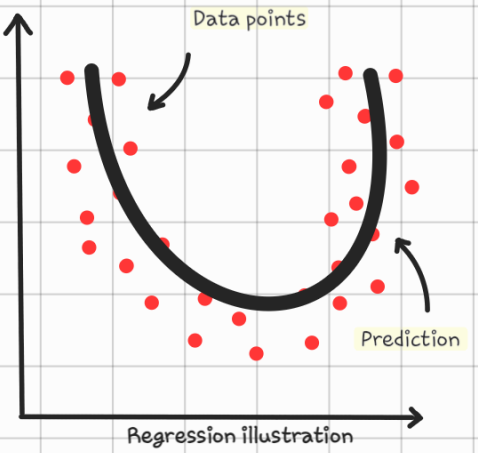
- The model is too simple for the complexity of problem.
- The training data is too small.
- The model is not trained long enough.

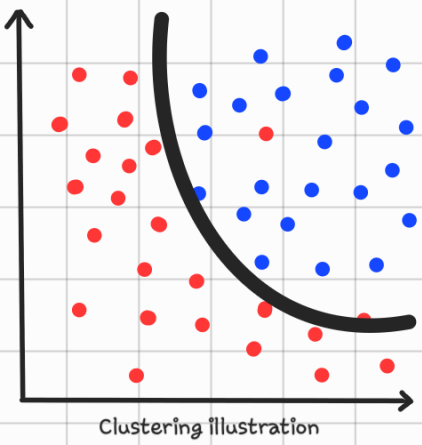


Generalisation

It is the ability of the model to perform well on new unseen data, it captures the underlying data patterns and make accurate predictions.

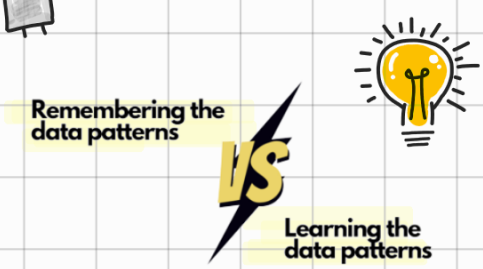
- to achieve good generalisation, the model must be trained on diverse and representative dataset that includes wide range of dataset.





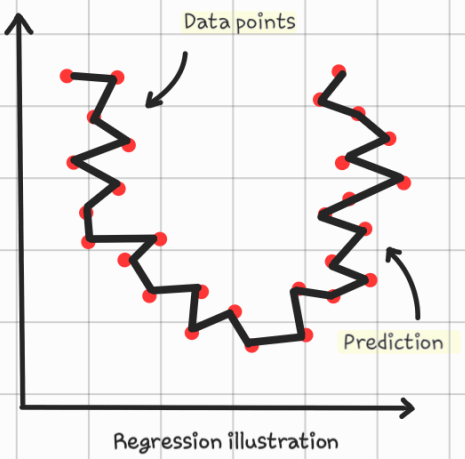
Overfitting

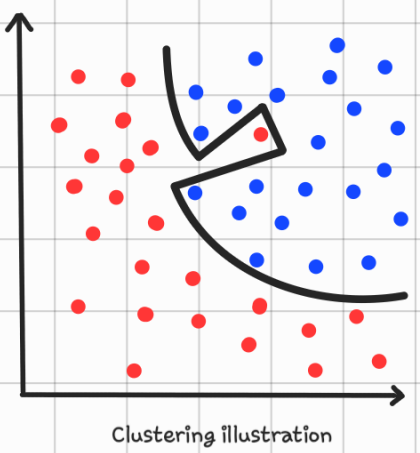
It occurs when the model is too complex and starts to fit noise in the training data, these models perform poor on unseen data.



Causes

- The model is too complex, has too many parameters.
- The model is trained for too long.
- The model uses a high-capacity algorithm.

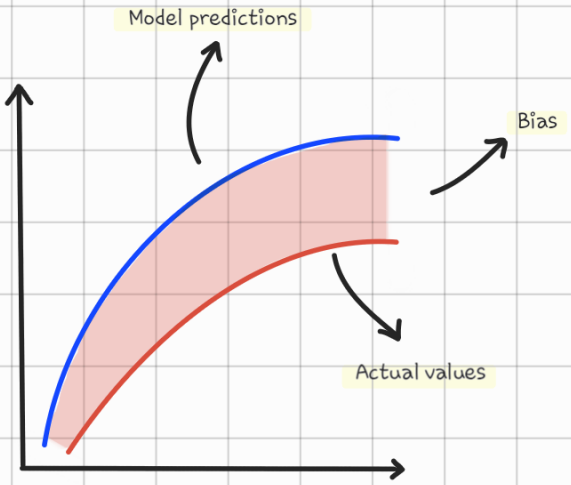




Bias

It is the difference between expected predictions of a model and the true values.

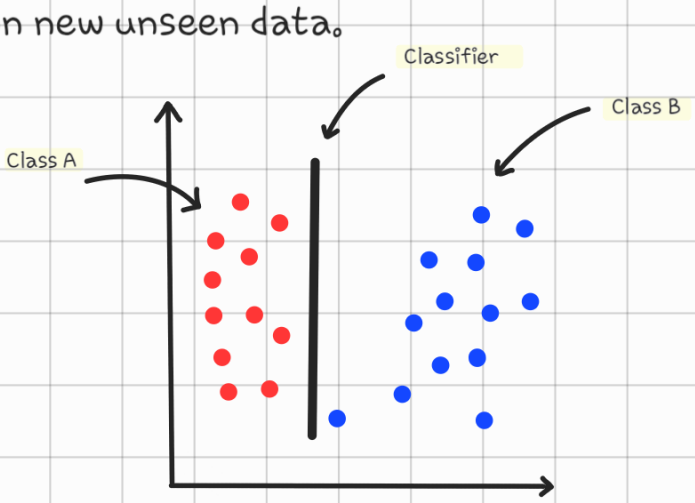
A model with high bias tends to underfit the training data. It is unable to capture the relevant features and relations the input and output variables.
In short, high bias is high error means poor performance.

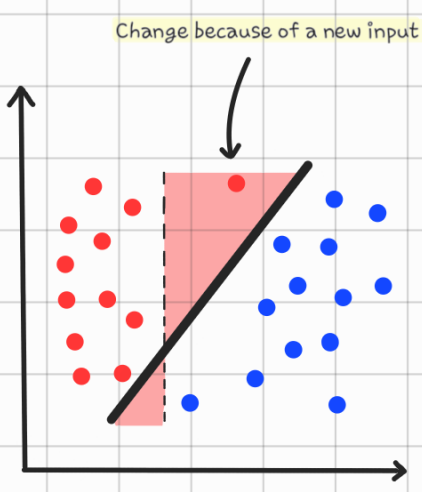


Variance

The sensitivity of the model's prediction to small fluctuations in the training data.

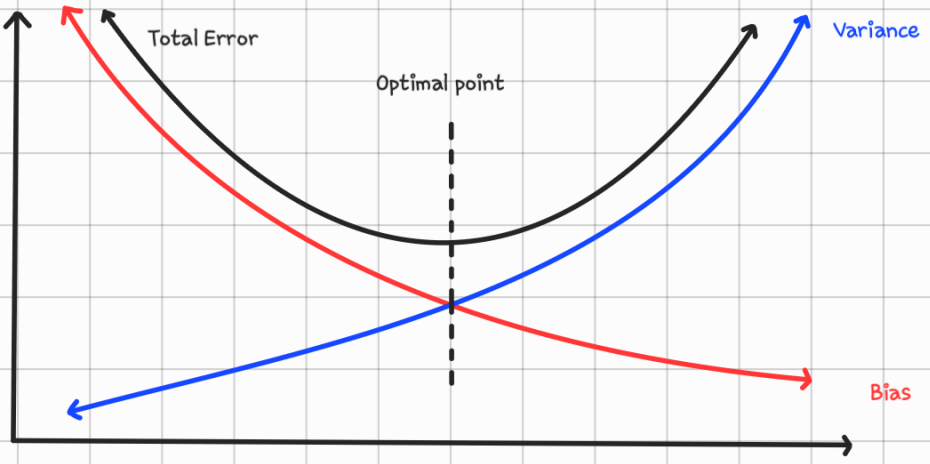
A model with high variance tends to overfit the training data and perform poorly on new unseen data.





This high change in the classifier shows that a model have high variance and is likely to cater all noise.

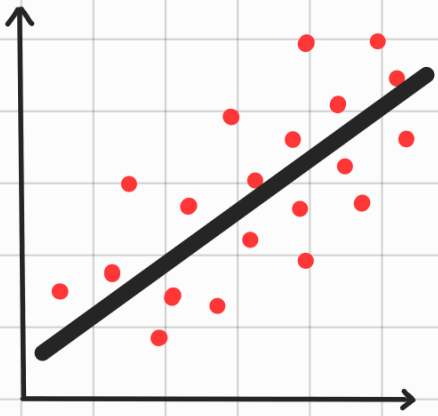
Bias (errors) and variance (senstivity) are opposite of each others. Balancing the trade off between bias and variance is essential for achieving good generalisation and avoiding overfitting or underfitting.



Supervised learning

Linear regression

It is the method to predict a dependent variable based on value of independent variable. The dependent variable is also caled target variable while the independent variable is also called as features or predictors. It is used when there exist a linear relationship b/w variables.



Always continuous

Y-Intercept

Slope of line

Error

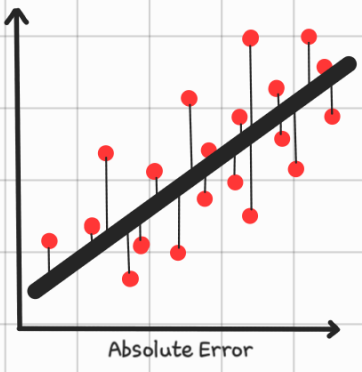
$$Y = \beta_0 + \beta_1 x + E$$

Can be both discrete or continuous

Errors in linear regression

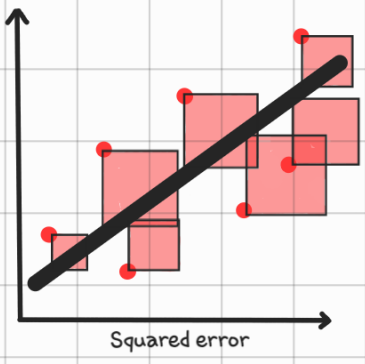
Sum of Absolute Error

Take absolute distance from each value to regression line and sum them.



Sum of Squared Error

Take square of distance from each value to regression line and sum the area of squares.



Issue

If we are comparing two models where A have 100 data points and the B have 1000 data points, B will obviously yield more sum of errors even with better performance, for this we have two other types

- Mean Absolute Error
- Mean Squared Error

and in cases like house prices predictions, if use $ as currency, squared Error will change units, which makes no sense. thus, we use

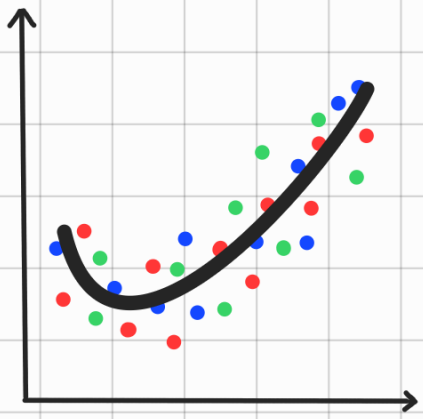
- Root Mean Squared Error

  RMSE is taking root of mean square error so we have matching units if the RMSE is 100$ that means our model makes around 100$ issue per prediction.

Polynomial regression

It is the method to predict a dependent variable based on value of multiple independent variable. It is used when there exist a non-lineaer relationship.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + E$$



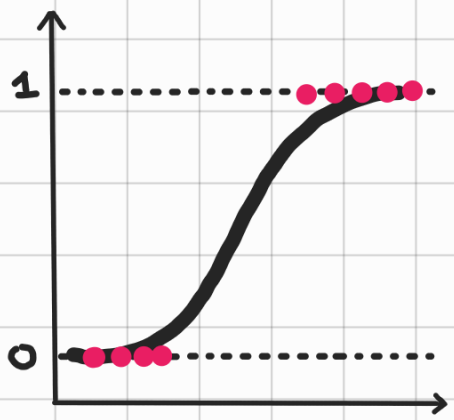
Logistic regression

It is the method to predict a dependent variable (Y) based on value of multiple independent variable where Y is categorical, predicting the probability of an outcome belonging to one of two classes.

$$P(Y) = \frac{e^{(\beta_0 + \beta_1 x)}}{e^{(\beta_0 + \beta_1 x)} + 1}$$

To make sure its > 0

To make sure its < 1



Decision Trees

A tree in which each node represents a predictor variable and the edge represents a decision and leaf represents an outcome. For example here is decision tree for understanding the risk to prevent heart attack.

Entropy

It measures the impurity or uncertainity present in data.

$$E_{(Parent)} = P_{(child\,1)} \log_2(P_{(child\,1)}) + P_{(child\,2)} \log_2(P_{(child\,2)}) + \ldots$$

Probability of that child in entire dataset

Information gain

It indicates how much information a variable gives us about the data, it is a measure to identify importance of the variable in the dataset and its effect on result.

$$gain = Ent_{(parent)} - P_{(child\,1)}\left(Ent_{(child\,1)}\right) - P_{(child\,2)}\left(Ent_{(child\,2)}\right) - \ldots$$

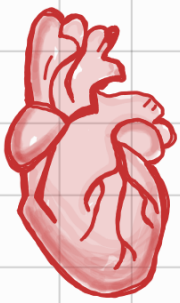
SyedBelall

GoSleepBelall

SyedBelall

GoSleepBelall

understanding the risk to prevent heart attack

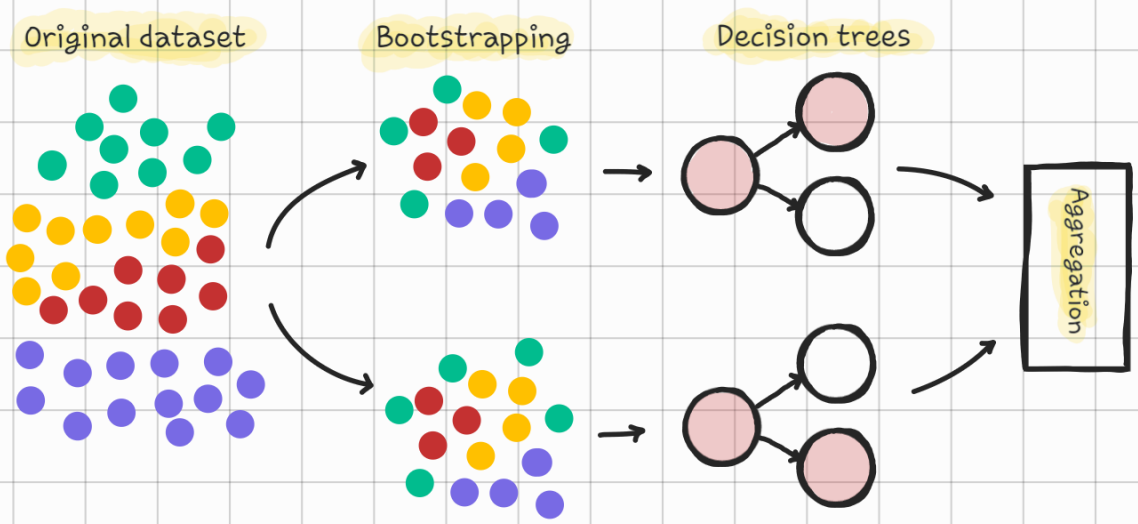


Decision trees usually face the problem of **Overfitting**, so Random forest comes into play to solve this issue.

Random Forest

It builds multiple decision trees and glue them togeather to get a more accurate and stable predictor. It classifies based on majority of decisions by all trees.

Random forest uses **Bagging** to subset the dataset using boot strapping.



Support Vector Machines

Hyperplane

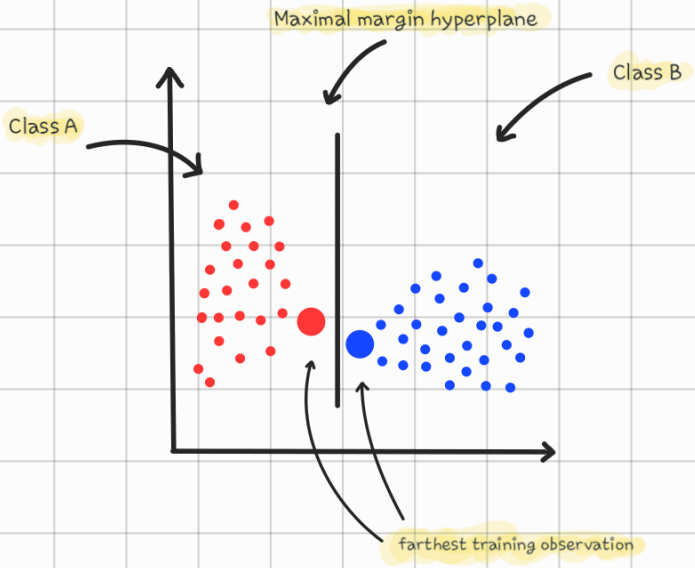
In a P dimensional space hyperplane is a flat affine subspace of P-1 dimension. For eg, a 1D line in a 2D coordinate system.

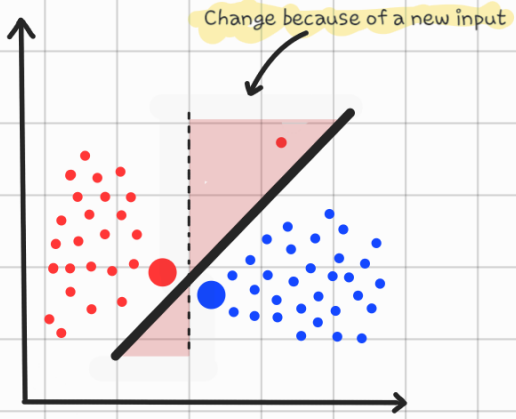
Maximal Margin Hyperplane

It is the seperating hyperplane for which the margin is largest -- that is the hyperplane that has the farthest minimum distance to training observations

It have several problems, for example:

- **Overfitting** in higher dimensions
- **Sensitive** to individual observation, it can show dramatic changes with addition of even a single observation.
- **Less confident** in case of small margins
- Inseperable cases cant be dealt



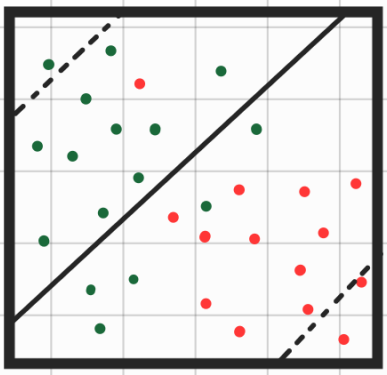


Support vectors classifier (soft margin)

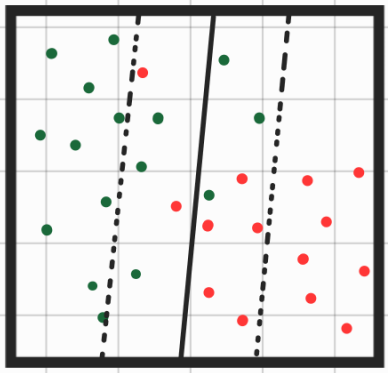
It allows some observation to be incorrect side of margin or even the incorrect side of hyperplane. The margin is soft because it can be violated to decrease senstivity.

Tolerance determines the number of severity of violations to the margin that it will tolerate. In general, tolerance is treated as tuning parameter that is generally choosen by cross-validation.

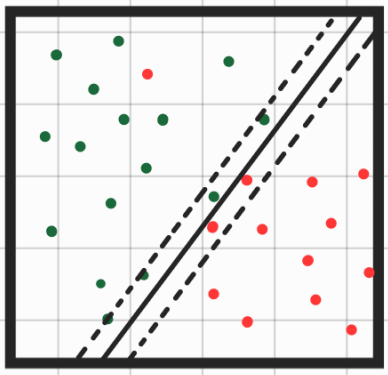
Support vectors are the observations that lie directly on the margin or on the wrong side of margins for their class.



tolerated: 1+5



tolerated: 2+3



tolerated: 1+0

The margin tightens as we decrease tolerance

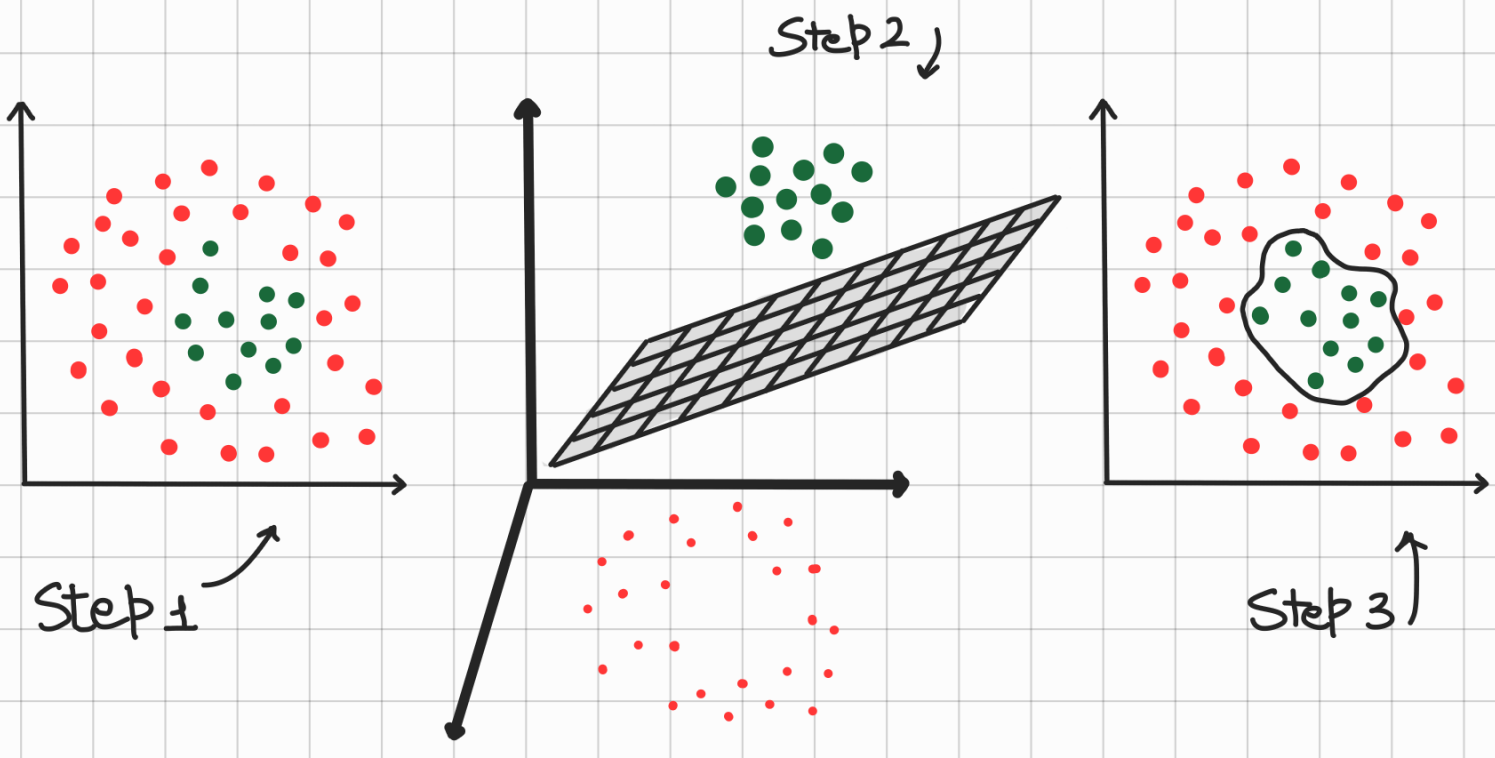
Support vectors machine

It is an extention of support vector classifier that results from enlarging the feature space in a specific way, using kernel tricks. by this, It can lead to much more flexible decision boundary

- sigmoid kernel
- Polynomial kernel
- Guassian kernel ...

Procedure

1. Takes the non linearly seperable data and project it to a higher dimension where it becomes linearly seperable.
2. Invoke the SVM to find the best linear decision boundary.
3. Project it back to the original space to get non-linear decision boundary.



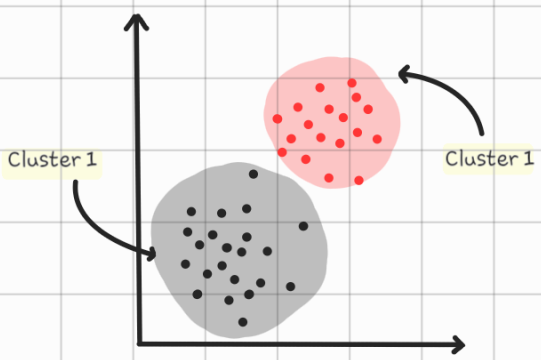
SyedBelall

GoSleepBelall

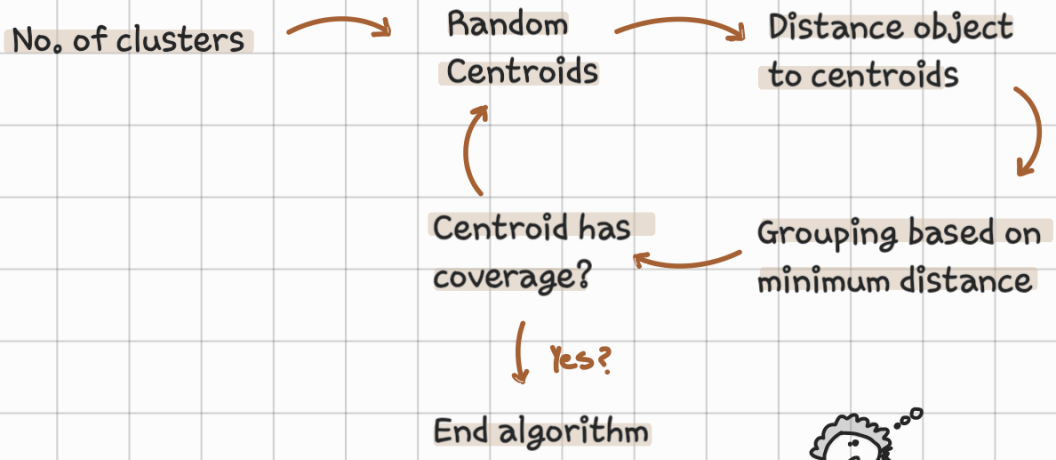
UnSupervised learning

K means clustering

The process by which objects are classified into a predefined number of groups so that, they have maximum similarity within the group and maximum disimilarity outside thegroup.

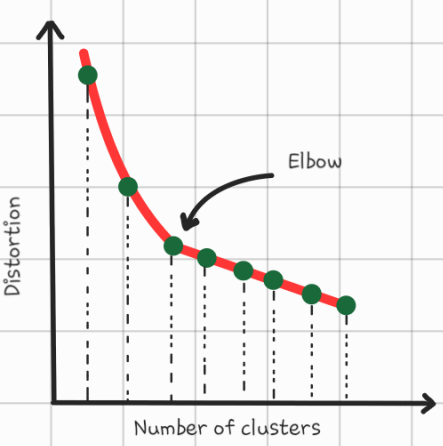


Procedure



But how to find optimal number of clusters ?

The elbow method helps us in this by identifying the point where the within-cluster sum of squares (WCSS) starts to decrease at a slower rate, resembling an "elbow" in the plot.



Problems

- Random centroid initialization
- Slow convergance

K means++ clustering

To resolve this K-means ++ was introduced for smart centroid initialization. It picks the centroid giving higher priority to farthest points from existing centroid.

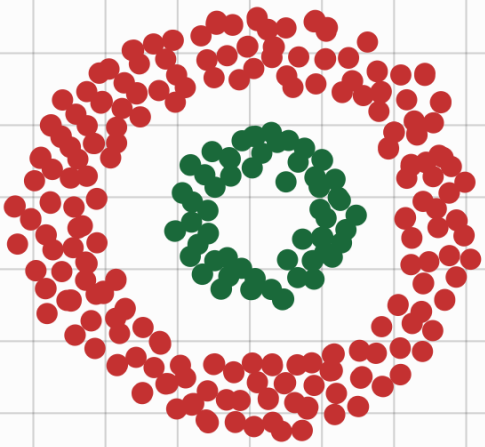
Still it have some problems like it's senstivity to outliers, its inability to identify clusters of irregular shapes and varying density, because of its nature of creating spherical clusters.

DBSCAN*

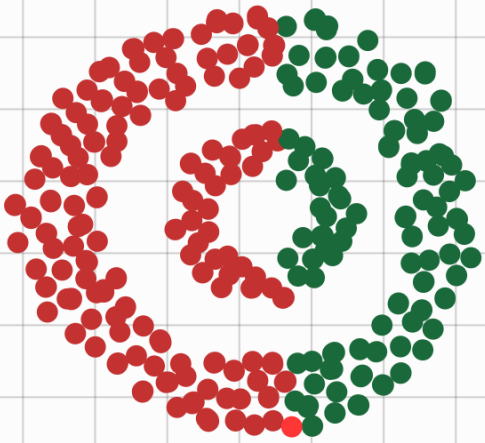
Density Based Spatial Clustering of Applications with Noise is a clustering algorithm that groups together points that are closely packed, defining clusters based on density and identifying outliers as noise.

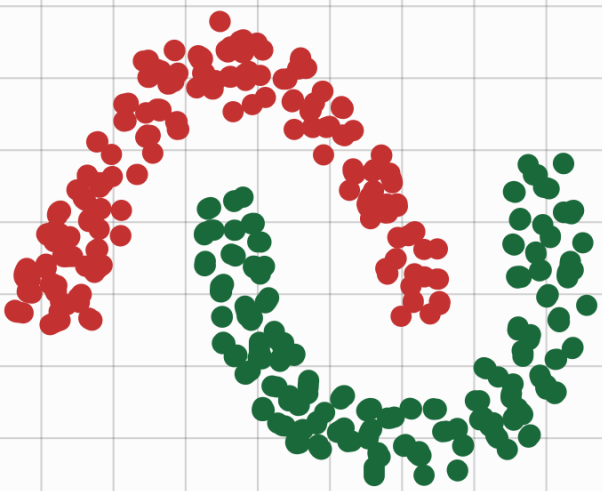
It automatically detects the number of clusters based on density.
It has the ability to handle clusters of arbitrary shapes and sizes.

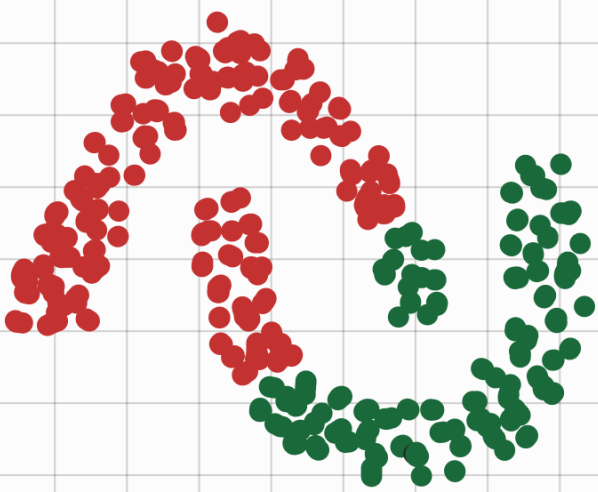
DBSCAN



K Means







Reinforcement learning

Reward

An instant return from the environment to appraise the last action of agent.

Policy

The approach that the agent uses to determine the next action based on current state.

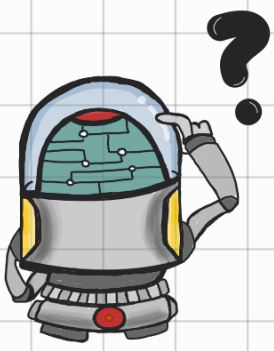
Value

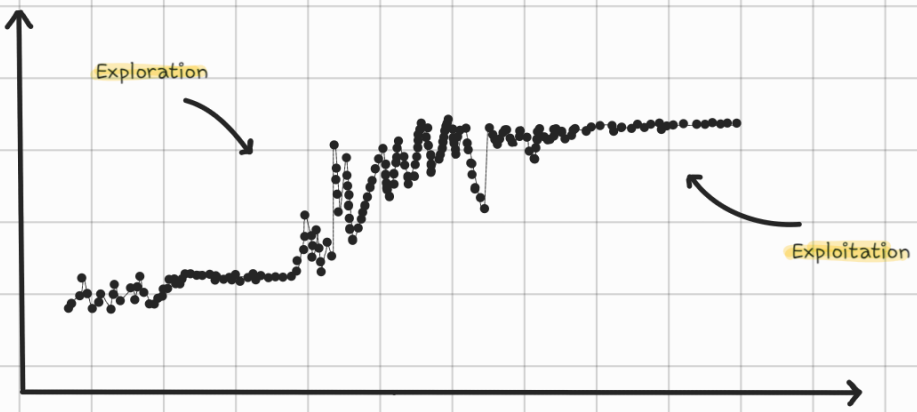
The expected long-term return with discounts, as opposed to the short-term reward.

Exploration vs. Exploitation

Exploration is about exploring and capturing more information about an environment, without prioritising reward.

Exploitation is about using the already known exploited information to maximise the reward.





Follow the Author

SyedBelall

GoSleepBelall

SyedBelalhyder

GoSleepBelall

SyedBelall