Advantages and Disadvantages of Linear Regression

Advantages:

- Simple and easy to understand
- Cheap computational cost
- Ground for more complex machine learning algorithms

Disadvantage:

- Oversimplify or fail in non-linear problems (only do well in linear modelling)
- Sensitive to outliers and noises and multicollinearity.
- Lack of practicality, most problems in our real world aren't "**linear**"

Advantages and Disadvantages of Logistic Regression

Advantages:

- Logistic Regression performs well when the **dataset is linearly separable**.
- Logistic regression is less prone to over-fitting but it can overfit in high dimensional datasets. You should consider Regularization techniques to avoid over-fitting in these scenarios.
- Logistic Regression not only gives a measure of how relevant a predictor (coefficient size) is, but also its direction of association (positive or negative).
- Logistic regression is easier to implement, interpret and very efficient to train.

Disadvantages:

- Main limitation of Logistic Regression is the **assumption of linearity** between the dependent variable and the independent variables.
- If the number of observations are lesser than the number of features, Logistic Regression should not be used, otherwise it may lead to overfit.
- Logistic Regression can only be **used to predict discrete functions**.

Advantages and Disadvantages of Cluster Analysis

Advantages:

- **Requires fewer resources**

Since cluster sampling selects only certain groups from the entire population, the method requires fewer resources for the sampling process.

- **More feasible**

The division of the entire population into homogenous groups increases the feasibility of the sampling. Additionally, since each cluster represents the entire population, more subjects can be included in the study.

Disadvantages:

- **Biased samples**

Cluster sampling is prone to biases. If the clusters that represent the entire population were formed under a biased opinion, the inferences about the entire population would be biased as well.

- **High sampling error**

Generally, the samples drawn using the cluster sampling method are prone to higher sampling error than the samples formed using other sampling methods.

Advantages and Disadvantages of Decision Tree

Advantages:

- Compared to other algorithms decision trees requires less effort for data preparation during pre-processing.
- A decision tree does not require normalization of data.
- A decision tree does not require scaling of data as well.
- Missing values in the data also does NOT affect the process of building decision tree to any considerable extent.
- A Decision trees model is very intuitive and easy to explain to technical teams as well as stakeholders.

Disadvantages:

- A small change in the data can cause a large change in the structure of the decision tree causing instability.
- For a Decision tree sometimes, calculation can go far more complex compared to other algorithms.
- Decision tree often involves higher time to train the model.
- Decision tree training is relatively expensive as complexity and time taken is more.
- Decision Tree algorithm is inadequate for applying regression and predicting continuous values.

Advantages and Disadvantages of Decision Tree

Advantages:

- Random Forest is based on the **bagging** algorithm and uses **Ensemble Learning** technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way it **reduces overfitting** problem in decision trees and also **reduces the variance** and therefore **improves the accuracy**.
- Random Forest can be used to **solve both classification as well as regression problems**.
- Random Forest works well with both **categorical and continuous variables**.
- Random Forest can automatically **handle missing values**.
- **No feature scaling required:** No feature scaling (standardization and normalization) required in case of Random Forest as it uses rule-based approach instead of distance calculation.
- **Handles non-linear parameters efficiently:** Non-linear parameters don't affect the performance of a Random Forest unlike curve-based algorithms. So, if there is high non-linearity between the independent variables, Random Forest may outperform as compared to other curve-based algorithms.
- Random Forest is usually **robust to outliers** and can handle them automatically.
- Random Forest algorithm is very **stable**. Even if a new data point is introduced in the dataset, the overall algorithm is not affected much since the new data may impact one tree, but it is very hard for it to impact all the trees.
- Random Forest is comparatively **less impacted by noise**.

Disadvantages:

- **Complexity:**

Random Forest creates a lot of trees (unlike only one tree in case of decision tree) and combines their outputs. By default, it creates 100 trees in Python sklearn library. To do so, this algorithm requires much more computational power and resources.

- **Longer Training Period:**

Random Forest require much more time to train as compared to decision trees as it generates a lot of trees (instead of one tree in case of decision tree) and makes decision on the majority of votes.

Advantages and Disadvantages of KNN

Advantages:

- Very simple implementation.
- Classes don't have to be linearly separable.
- Classifier can be updated at very little cost as new instances with known classes are presented.
- Few parameters to tune: distance metric and k.

Disadvantages:

- Expensive testing of each instance, as we need to compute its distance to all known instances.
- Sensitiveness to noisy or irrelevant attributes, which can result in less meaningful distance numbers.
- Scaling and/or feature selection are typically used in combination with kNN to mitigate this issue.
- Sensitiveness to very unbalanced datasets, where most entities belong to one or a few classes, and infrequent classes are therefore often dominated in most neighbourhoods. This can be alleviated through balanced sampling of the more popular classes in the training stage, possibly coupled with ensembles.

Advantages and Disadvantages of Naïve Bayes

Advantages:

- When assumption of independent predictors holds true, a Naive Bayes classifier performs better as compared to other models.
- Naive Bayes requires a small amount of training data to estimate the test data. So, the training period is less.
- Naive Bayes is also easy to implement.

Disadvantages:

- Main imitation of Naive Bayes is the **assumption of independent predictors**. Naive Bayes implicitly assumes that all the attributes are mutually independent. In real life, it is almost impossible that we get a set of predictors which are completely independent.
- If categorical variable has a category in test data set, which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as **Zero Frequency**. To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called **Laplace estimation**.