Linear Regression

1)Simple linear regression has a single independent variable whereas multiple linear regression has more than 1 independent variable.
2)Cost Function shows the difference of the original output from the expected output
3)A positive coefficient indicates that as the value of the independent variable increases, the mean of the dependent variable also tends to increase. A negative coefficient suggests that as the independent variable increases, the dependent variable tends to decrease.
4)There are four assumptions associated with a linear regression model:

1. Linearity: The relationship between X and the mean of Y is linear.
2. Homoscedasticity: The variance of residual is the same for any value of X.
3. Independence: Observations are independent of each other.
4. Normality: For any fixed value of X, Y is normally distributed

Logistic Regression

1.Linear Regression is used to handle regression problems whereas Logistic regression is used to handle the classification problems. Linear regression provides a continuous output but Logistic regression provides binary output.

2.Logistic regression is used for binary classification, where the goal is to predict one of two possible outcomes, typically represented as 0 and 1. The sigmoid function is at the core of logistic regression, serving as the link function that maps the linear combination of input features to a probability.

3.The most popular metric used to evaluate the performance of classification models is accuracy. However, accuracy isn't always the most reliable indicator of a good model, which is why often use measures like precision, recall, and the F1-score instead.

4.The potential solutions include the following:

- Remove some of the highly correlated independent variables.
- Linearly combine the independent variables, such as adding them together

Naive Bayes

1.It is based on the Bayes' Theorem for calculating probabilities and conditional probabilities.

2.In the context of Naive Bayes, conditional probability refers to the likelihood of an event occurring given that another event has already occurred.

3.Advantages

- This algorithm works quickly and can save a lot of time.
- Naive Bayes is suitable for solving multi-class prediction problems.
- If its assumption of the independence of features holds true, it can perform better than other models and requires much less training data.
- Naive Bayes is better suited for categorical input variables than numerical variables.

Disadvantages

- Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.
- This algorithm faces the 'zero-frequency problem' where it assigns zero probability to a categorical variable whose category in the test data set wasn't available in the training dataset. It would be best if you used a smoothing technique to overcome this issue.
- Its estimations can be wrong in some cases, so you shouldn't take its probability outputs very seriously.

4.Naive Bayes classifiers generally handle missing values by ignoring the instances with missing values during both training and classification.Naive Bayes classifiers handle categorical features by calculating the probabilities associated with each category within the feature and then using these probabilities to make predictions.

Decision Trees

1.A decision tree typically starts with a single node, which branches into possible outcomes. Each of those outcomes leads to additional nodes, which branch off into other possibilities

2.Main Criteria is calculating gini impurity and the entropy of each node.

3.Decision trees handle categorical variables by splitting the data based on the categories of the categorical variable. When building a decision tree, the algorithm selects the feature and the split point that maximizes the information gain or minimizes impurity at each node. For categorical variables, each category acts as a potential split point.

4.a)Pruning

b)Limiting maximum depth to prevent the tree from becoming too complex.

c)Feature selection

d)Cross-validation

e)ensemble methods

SVM

1)The basic idea behind Support Vector Machines (SVM) is to find the optimal hyperplane that best separates data points belonging to different classes in a high-dimensional space. The key objectives of SVM are maximizing the margin between classes and minimizing classification error.

2)The margin in SVM refers to the distance between the decision boundary (hyperplane) and the closest data points (support vectors) from each class.Support vectors are the data points that lie closest to the decision boundary (hyperplane) and have the most influence on determining the position and orientation of the hyperplane.

3)Kernel types:

a)The **linear kernel** is the simplest kernel function and is used when the data is linearly separable or when the number of features is very high.

b)**The polynomial kerne**l is useful for capturing non-linear relationships between features.It is suitable for scenarios where the decision boundary is expected to be polynomial.

c)Radial Basis Function (RBF) Kernel:It is suitable for scenarios where the decision boundary is expected to be non-linear and complex. However, it can be sensitive to the choice of parameters.

d)Sigmoid Kernel:It may be suitable for scenarios where the decision boundary is expected to be non-linear and sigmoid-shaped.

4)Here's how Svm handles Ouliers:

a)Robustness to outliers: SVM is generally robust to outliers because the optimization objective primarily focuses on maximizing the margin between classes.

b)Soft margin SVM: In scenarios where the data is not perfectly separable, SVM allows for some misclassification of data points through the use of a soft margin. This means that outliers can be tolerated to some extent without significantly affecting the decision boundary, as long as they do not violate the margin boundary excessively.

c)Choice of kernels affects how outliers are handled.The linear kernel is generally more robust to outliers compared to non-linear kernels like the RBF kernel. Linear SVMs tend to be less influenced by outliers because they aim to find a linear decision boundary that separates classes without requiring complex transformations.

d)Data preprocessing:Data preprocessing techniques such as outlier detection and removal, data scaling, and normalization can help mitigate the impact of outliers on SVM. Outliers can be identified using statistical methods or domain knowledge and either removed or downweighted during training. Additionally, scaling and normalization techniques can help ensure that features are on similar scales, reducing the influence of outliers.