1) What is the difference between simple linear regression and multiple linear regression?

   Simple linear regression involves predicting a target variables using only one predictor variable, while multiple linear regression involves predicting the target variable using 2 or more predictor variables. In simple linear regression, the relationship between the predictor and target variable is assumed to be linear, whereas in multiple linear regression, the relationship can be more complex as it considers multiple predictors simultaneously.

2) Explain the concept of the cost function in linear regression.

   The cost function in linear regression measures the difference between the actual target values and the predicted values generated by the model. It quantifies how well the model is performing. The cost function used in linear regression is the mean squared error, which calculates the average squared difference between the actual and predicted values. The goal of training a linear regression model is to minimize this cost function, to find the set of model parameters that results in the lowest MSE.

3) How do you interpret the coefficients in a linear regression model?

   In a linear regression model, coefficients represent the change in the target variable for a one-unit change in the corresponding predictor variable, holding other variables constant In simple linear regression model with one predictor variable, the coefficient represents the change in the target variable for a one unit change in the target variable when the corresponding predictor Variable changes by one unit, while holding all other predictor variables constant.

4) What are the assumptions of linear regression?

   Linearity, Independence, Homoscedasticity, Normality

5) How does logistic regression differ from linear regression?

   Logistic regression is used for binary classification tasks, whereas linear regression is used for predicting continuous numerical values. In logistic regression, the output is transformed using the sigmoid function to constrain it between 0 and 1, representing probabilities. Linear regression does not have such a transformation.Logistic regression models the probability of a binary outcome based on one or more predictor variables, while linear regression models the relationship between a continuous target variable and one or more predictor variables.

6) Explain the sigmoid function and its role in logistic regression

The sigmoid function, also known as the logistic function, is an S-shaped curve that maps any real-valued number to a value between 0 and 1. In logistic regression, the sigmoid function transforms the output of the linear combination of predictor variables into probabilities. It ensures that the output of the logistic regression model is interpretable as the probability of belonging to a particular class. The sigmoid function's output represents the likelihood of the event occurring, making it suitable for binary classification problems.
Sig(x) = 1/(1+ e^-x)

7) What are the key performance metrics used to evaluate a logistic regression model?

   Accuracy, Precision, Recall, F1-Score
8) How do you handle multicollinearity in logistic regression?

   Remove one of the correlated variables.
   Combine the correlated variables into a single composite variable.
   Regularized logistic regression model using techniques like Lasso or Ridge regression.
   Use PCA.

9) What is the Naive Bayes algorithm based on?

   The Naive Bayes algorithm is based on Bayes' theorem, which describes the probability of an event, based on prior knowledge of conditions that might be related to the event. It assumes that the presence of a particular feature in a class is independent of the presence of any other feature.

10) Explain the concept of conditional probability in the context of \Naive Bayes.

    Conditional probability in the context of Naive Bayes refers to the probability of an event occurring given that another event has already occurred. In Naive Bayes, it calculates the probability of a class given a set of features by multiplying the probabilities of each feature given the class.

11) What are the advantages and disadvantages of Naive Bayes?

    Advantages: Simple and easy to implement, Efficient for large datasets, Performs well in multi-class prediction
    Disadvantages: Assumation, Sensitivity to feature distribution, Can’t handle numerical features without discretization.

12) How does Naive Bayes handle missing values and categorical features?

    Missing Values: Naive Bayes can handle missing values by ignoring the missing values during probability estimation.

Categorical Features: Naive Bayes can handle categorical features by treating each category as a separate feature and calculating the probabilities accordingly.

13) How does a decision tree make decisions?

Decision trees make decisions by recursively partitioning the feature space into smaller regions based on the values of the predictor variables. At each node of the tree, a decision is made based on the value of a specific feature, and the data is split into two or more branches according to the decision rule. This process continues until a stopping criterion is met, such as reaching a maximum depth or having a minimum number of samples in a node.

14) What are the main criteria for splitting modes in a decision tree?

Gini Impurity, Information Gain, Gain Ratio

15) How do decision trees handle categorical variables?

Decision trees handle categorical variables by creating binary splits based on the categories. Each category becomes a separate branch in the tree, and the algorithm chooses the best split based on criteria such as Gini impurity or information gain.

16) What are some common techniques to prevent overfitting in decision trees?

Pruning, Setting minimum and maximum samples per leaf, Ensembling

17) What is the basic idea behind SVM?

The basic idea behind SVM is to find the hyperplane that best separates the data points into different classes. SVM aims to maximize the margin between the hyperplane and the nearest data points (support vectors) of each class while minimizing the classification error.

18) Explain the concepts of margin and support vectors in SVM.

Margin: The distance between the hyperplane and the closest data points (support vectors) from each class. SVM aims to maximize this margin.
Support vectors: The data points that lie closest to the hyperplane and have the most influence on determining the position of the hyperplane.

19) What are the different kernel functions used in SVM, and when would you use each?

Linear kernel: Suitable for linearly separable data.
Polynomial kernel: Suitable for non-linear data with polynomial decision boundaries.

Radial Basis Function (RBF) kernel: Suitable for non-linear data with complex decision boundaries.
Sigmoid kernel: Suitable for data that is not linearly separable.

20) How does SVM handle outliers?
Use outlier detection techniques to identify and remove outliers before training the model.
Adjust the regularization parameter (C) to control the influence of outliers on the decision boundary.
Use robust kernel functions like the RBF kernel, which are less sensitive to outliers compared to linear kernels.