Linear Regression:

1.What is the difference between simple linear regression and multiple linear regression?
Simple linear regression predicting a dependent variable based on a single independent variable. While multiple linear regression predicting a dependent variable based on two or more independent variables.
Simple linear regression fits a straight line to the data points, capturing the relationship between the two variables. While multiple linear regression fits a linear equation to the observed data by minimizing the sum of the squared differences between the observed and predicted values.

2.Explain the concept of the cost function in linear regression.
In linear regression the cost function quantifies the difference between the predicted values of the dependent variable and the actual observed values. The goal of linear regression is to minimize this cost function.
Mean Squared Error (MSE) is used as a cost function in linear regression. It calculates the average squared difference between the predicted values and the actual values across all data points.

MSE = 1/n(Summation from 1 to n (yi - y^i)^2)

3.How do you interpret the coefficients in a linear regression model?
In a linear regression model the coefficients represent the relationship between the independent variables and the dependent variable. These coefficients represent the change in the dependent variable for a one-unit change in the corresponding independent variable.
If the coefficient is positive, it means that independent variable increases, the dependent variable tends to increase as well.
If the coefficient is negative, it means that the independent variable increases, the dependent variable tends to decrease as well.

4.What are the assumptions of linear regression?
The relationship between the independent variables and the dependent variable should be linear.
The independent variables should not be highly correlated with each other. Multicollinearity can make it difficult to separate the individual effects of the independent variables on the dependent variable.
The errors should be normally distributed.
The errors should not be correlated with each other.

Logistic Regression:

1.How does logistic regression differ from linear regression?
Logistic regression is used when the dependent variable is categorical, whereas linear regression is used when the dependent variable is continuous.

Linear regression is suitable for modeling continuous outcomes, whereas logistic regression is appropriate for modeling binary or categorical outcomes and it is particularly useful in classification problems.

2.Explain the sigmoid function and its role in logistic regression.
The sigmoid function is a mathematical function that maps any real-valued number to a value between 0 and 1. The sigmoid function is a key component of logistic regression because it transforms the linear combination of the independent variables into a probability score.

Sigma = 1 / (1 + (e^(-z)))

3.What are the key performance metrics used to evaluate a logistic regression model?
Accuracy measures the proportion of correctly classified instances out of all instances.
Precision measures the proportion of true positive predictions out of all positive predictions.
Recall measures the proportion of true positive predictions out of all actual positive instances.
The F1 score provides a balance between precision and recall.

4.How do you handle multicollinearity in logistic regression?
Multicollinearity occurs when independent variables in a logistic regression model are highly correlated with each other.
Identify and remove one of the correlated variables from the model.
Create new independent variables by combining or transforming existing variables.
Regularization techniques can penalize large coefficients and shrink them towards zero,

Naive Bayes:

1.What is the Naive Bayes algorithm based on?
The Naive Bayes algorithm is based on Bayes theorem. By applying Bayes theorem Naive Bayes calculates the posterior probability of each class label and predicts the class label with the highest probability as the final prediction.
P(A/B) = (P(A) * P(B/A)) / P(B)

2.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 observing a particular feature given a specific class label. In Naive Bayes classification, we aim to predict the probability of each class label given a set of observed features. For that we calculate the posterior probability of each class label given the observed features using Bayes theorem

3.What are the advantages and disadvantages of Naive Bayes?
Naive Bayes is simple and easy to understand. It is based on probabilistic principles and requires minimal training data to estimate parameters. Naive Bayes is computationally efficient, especially for high-dimensional datasets with many features.
Naive Bayes tends to produce biased probability estimates, particularly for rare combinations of feature values. Naive Bayes can be sensitive to imbalanced datasets, where one class is much

more relevant than others. It tends to favor the majority class and may produce biased predictions for minority classes.

4.How does Naive Bayes handle missing values and categorical features?
Removing instances with missing values.
Imputing missing values with a certain value .
For categorical features with a small number of categories, the probabilities can be estimated using frequency counts.
For categorical features with a large number of categories, smoothing techniques such as Laplace smoothing are used to handle categories that may not appear in the training data.

Decision Trees:

1.How does a decision tree make decisions?
Decision tree makes decisions by recursively partitioning the feature space into smaller subsets based on the values of the features. It aims to maximize the purity of the target variable within each subset. This process results in a tree-like structure where decisions are made based on the paths taken from the root node to the leaf nodes.

2.What are the main criteria for splitting nodes in a decision tree?
The main criteria for splitting nodes in a decision tree are measures of impurity or information gain.
Gini impurity measures the probability of incorrectly classifying a randomly chosen element if it were randomly labeled according to the distribution of labels in the node.
Entropy measures the impurity or randomness in a set of data.
Information gain measures the reduction in entropy or Gini impurity achieved by splitting the dataset based on a particular feature.

3.How do decision trees handle categorical variables?
Decision Trees handle categorical features by one-hot-encoding in a preprocessing step. It means using label encoding decision tree handle categorical variables.

4.What are some common techniques to prevent overfitting in decision trees?
Pruning is a technique that involves removing branches from the tree that do not significantly improve its performance on validation data.
Pre-pruning: Stop growing the tree early by constraints on the maximum depth of the tree, the minimum number of samples required to split a node.
Post-pruning: Grow the tree to its full size and then prune it by iteratively removing branches that do not improve the tree performance on a validation dataset.

<u>Support Vector Machines (SVM):</u>

1.What is the basic idea behind SVM?

The basic idea behind SVM is to find the hyperplane that best separates the data points of different classes in a high-dimensional space. SVM works by transforming a non-linearly separable problem into a linearly separable one by the use of a kernel function.

2.Explain the concepts of margin and support vectors in SVM.

The margin in SVM refers to the distance between the decision boundary and the closest data points of different classes. The goal of SVM is to find the hyperplane that separates the classes with the maximum possible margin.
Support vectors are the data points that lie closest to the decision boundary. During the training process, SVM identifies the support vectors and uses them to define the margin and determine the optimal hyperplane.

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

Linear Kernel: The linear kernel is the simplest kernel function and is used when the data is linearly separable. It computes the dot product between the input vectors directly in the original feature space.
Polynomial Kernel: The polynomial kernel is used to handle nonlinear relationships between data points. It maps the input vectors into a higher-dimensional space using polynomial functions.
Gaussian Radial Basis Function (RBF) Kernel: The Gaussian RBF kernel is one of the most commonly used kernel functions in SVM. It maps the input vectors into an infinite-dimensional space using a Gaussian radial basis function. It is highly flexible and capable of capturing complex nonlinear relationships between data points.

4.How does SVM handle outliers?

SVM aims to maximize the margin between the decision boundary and the closest data points. Outliers that are far from the decision boundary are less likely to be considered support vectors and have minimal impact on the positioning of the decision boundary.
SVM incorporates a regularization parameter (C) that controls the trade-off between maximizing the margin and minimizing the classification error. A smaller value of C results in a larger margin and a higher tolerance for misclassified points, making the model less sensitive to outliers.