Linear Regression

1.What is the difference between simple linear regression and multiple linear regression?
Simple linear regression contains only one independent variable.
Multiple linear regression contains two or more independent variables.

2.Explain the concept of the cost function in linear regression.
Cost function is the formula that calculates the average error between predicted and actual value over the entire training dataset.
Cost function = Avg((predicted-actual)2)

3.How do you interpret the coefficients in a linear regression model?
Coefficient is equal to the average change in the dependent variable(y) for a unit change in the independent variable(x).

4.What are the assumptions of linear regression?
There should be a linear relationship between dependent and independent variables.
The independent variables should not be highly correlated with each other.
The error should follow normal distribution.

Logistic Regression:

1.How does logistic regression differ from linear regression?
Logistic regression is used for classification models where the target variable is categorical whereas in linear it is continuous.

2.Explain the sigmoid function and its role in logistic regression.
Sigmoid function is a mathematical function that transforms the input value into a range between 0 and 1 making it useful for binary classification in logistic regression.

3.What are the key performance metrics used to evaluate a logistic regression model?
The key performance metrics used in the logistic regression are
Accuracy = (TP+TN)/(TP+FP+TN+FN)
Recall = TP/(TP+FN)
Precision = TP/(TP+FP)
F1 score = (2*precision*recall)/(precision+recall)

4.How do you handle multicollinearity in logistic regression?
To handle multicollinearity we can remove one or more highly correlated variables and combine them into one or can use PCA to reduce the number of variables while retaining most of the information.

Naive Bayes:

1.What is the Naive Bayes algorithm based on?
Naive Bayes algorithm is based on Bayes theorem which describes the probability of an event based on the prior knowledge of the event that has already occurred.

2.Explain the concept of conditional probability in the context of Naive Bayes.
Conditional probability in the context of Naive Bayes is the probability of an event occurring given that another event has already occurred.

3.What are the advantages and disadvantages of Naive Bayes?
The advantage of naive bayes is that it is simple and fast, can perform well on small data.
The disadvantage of naive bayes is that it is sensitive to input data quality, estimates probability when the data is imbalance.

4.How does Naive Bayes handle missing values and categorical features?
Naive bayes handles missing values by simply skipping them during the calculation of probabilities and handles categorical features by considering them as separate class within the features.

Decision Trees:

1.How does a decision tree make decisions?
Decision tree recursively split data into classes based on the features. This continues until the stopping criteria is met such as attaining the purity that is containing only one class.

2.What are the main criteria for splitting nodes in a decision tree?
The main criteria for splitting nodes in a decision tree gini impurity and entropy.

3.How do decision trees handle categorical variables?
Decision tree handles categorical variables by treating each category as a separate branch in the tree. This splitting is done until stopping criteria is met or max depth is reached.

4.What are some common techniques to prevent overfitting in decision trees?
Overfitting can be prevented by pruning in which extra branches are removed whose branch does not add value to the model performance and also can use ensemble methods such as random forest, xgboost where parallel and series of decision trees are used.

Support Vector Machines (SVM):

1.What is the basic idea behind SVM?
The basic idea behind SVM is to find the optimal hyperplane that best separates the different classes in the feature space. This hyperplane maximizes the margin which is the distance between the hyperplane and the nearest data points from each class, making SVM effective for both linear and non-linear classification tasks.

2.Explain the concepts of margin and support vectors in SVM.
In SVM, margin is the distance between the decision boundary(hyperplane) and the nearest datapoint from each class. Support vectors are the data points that lie closest to the decision boundary(hyperplane) and are crucial in determining the position and orientation of the hyperplane.

3.What are the different kernel functions used in SVM, and when would you use each?
The different kernels in SVM are linear kernel which is used for linearly separable data, polynomial kernel which is used when data has non linear relationship between them, Sigmoid kernel which can be used for classification task.

4.How does SVM handle outliers?
SVM handle outliers to some extent because of the use of margin, which focuses on correctly classifying the majority of of data points while minimizing the influence of outliers