Linear Regression:

1. The main difference is that simple linear regression involves one independent variable, while multiple linear regression involves two or more independent variables.
2. The cost function in linear regression measures the difference between the predicted values and the actual values, aiming to minimize this difference.
3. The coefficients in a linear regression model represent the change in the dependent variable for a one-unit change in the independent variable, holding all other variables constant.
4. Assumptions include linearity, independence of errors, homoscedasticity, and normality of errors.

Logistic Regression:

1. Logistic regression is used for classification tasks with a binary outcome, while linear regression is used for continuous outcomes.
2. The sigmoid function maps the output of logistic regression to a probability between 0 and 1, making it suitable for binary classification.
3. Key performance metrics include accuracy, precision, recall, F1-score, and ROC-AUC.
4. Multicollinearity in logistic regression can be handled by removing one of the correlated variables or using techniques like ridge regression.

Naive Bayes:

1. Naive Bayes algorithm is based on Bayes' theorem and the assumption of conditional independence between features.
2. Conditional probability in Naive Bayes refers to the probability of a particular class given the values of the features.
3. Advantages include simplicity, scalability, and effectiveness for text classification. Disadvantages include the strong assumption of feature independence.
4. Naive Bayes can handle missing values by ignoring them during training and by using techniques like Laplace smoothing for categorical features.

Decision Trees:

1. Decision trees make decisions by recursively splitting the data based on the values of features to minimize impurity or maximize information gain.
2. Main criteria for splitting nodes include Gini impurity, entropy, and classification error.
3. Decision trees can handle categorical variables by creating binary splits based on different categories.
4. Techniques to prevent overfitting include pruning, setting minimum samples per leaf, and using ensemble methods like random forests.

Support Vector Machines (SVM):

1. SVM aims to find the hyperplane that maximizes the margin between different classes in a high-dimensional space.
2. Margin is the distance between the hyperplane and the closest data points (support vectors) from each class.
3. Common kernel functions include linear, polynomial, and radial basis function (RBF). Linear kernel is used for linearly separable data, polynomial for non-linear data, and RBF for complex data.
4. SVM is robust to outliers as it focuses on the support vectors near the decision boundary.