1. The logistic function, also known as the sigmoid function, is used in logistic regression to compute probabilities. It maps any input value to a value between 0 and 1, which can be interpreted as the probability of the input belonging to a certain class.
2. The criterion commonly used to split nodes in decision tree construction is the information gain. It is calculated by subtracting the weighted average of the entropy of the child nodes from the entropy of the parent node.
3. Entropy is a measure of impurity in a set of examples. Information gain is the reduction in entropy achieved by partitioning the examples based on a certain attribute. In decision tree construction, the attribute with the highest information gain is chosen as the splitting criterion.
4. The random forest algorithm utilizes bagging and feature randomization to improve classification accuracy. Bagging involves training multiple decision trees on different subsets of the training data, while feature randomization involves randomly selecting a subset of features for each tree.
5. The distance metric typically used in k-nearest neighbors (KNN) classification is the Euclidean distance. It measures the straight-line distance between two points in a multidimensional space. The choice of distance metric can impact the algorithm's performance.
6. The Naïve-Bayes assumption of feature independence assumes that the features used for classification are conditionally independent given the class label. This assumption simplifies the computation of the posterior probability of a class given a set of features.
7. The kernel function in SVMs is used to transform the input data into a higher-dimensional space where it is easier to separate the classes. Some commonly used kernel functions include the linear kernel, polynomial kernel, and radial basis function (RBF) kernel.
8. The bias-variance tradeoff refers to the tradeoff between model complexity and overfitting. A model with high bias (e.g., a linear model) may underfit the data, while a model with high variance (e.g., a complex model) may overfit the data. The goal is to find a model with an appropriate balance between bias and variance.
9. TensorFlow facilitates the creation and training of neural networks by providing a high-level API for building and training models. It also includes a variety of pre-built neural network layers and activation functions.
10. Cross-validation is a technique used to evaluate model performance by partitioning the data into training and validation sets. It involves training the model on multiple subsets of the data and evaluating its performance on the remaining subset. Cross-validation is important for detecting overfitting and selecting the best model.
11. Techniques that can be employed to handle overfitting in machine learning models include regularization, early stopping, and dropout. Regularization

involves adding a penalty term to the loss function to discourage overfitting, while early stopping involves stopping the training process when the validation error stops improving. Dropout involves randomly dropping out some neurons during training to prevent over-reliance on certain features.

12. The purpose of regularization in machine learning is to prevent overfitting by adding a penalty term to the loss function. The penalty term encourages the model to have smaller weights, which can help prevent over-reliance on certain features.

13. Hyper-parameters are parameters that are set before training the model and cannot be learned from the data. They include parameters such as the learning rate, regularization strength, and number of hidden layers in a neural network. Hyper-parameters are tuned for optimal performance using techniques such as grid search or random search.

14. Precision and recall are metrics used to evaluate the performance of a classification model. Precision measures the proportion of true positives among all positive predictions, while recall measures the proportion of true positives among all actual positives. Accuracy measures the proportion of correct predictions among all predictions.

15. The ROC curve is a graphical representation of the performance of a binary classifier. It plots the true positive rate (TPR) against the false positive rate (FPR) for different threshold values. The area under the ROC curve (AUC) is a commonly used metric for evaluating the performance of a binary classifier.