Q1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

Ans: Residual Sum of Squares is a better measure of goodness of fit model in regression as it measures the amount of error remaining between the regression function and the data set after the model has been run.

Q2. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.

Ans: The Equation is TSS = ESS + RSS

The TSS tells you how much variation there is in the dependent variable.

The ESS tells you how much of the variation in the dependent variable your model explained.

The RSS tells you how much of the dependent variable's variation your model did not explain.

Q3. What is the need of regularization in machine learning?

Ans: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Q4. What is Gini–impurity index?

Ans: Gini–impurity index calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly.

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans: Yes decision-trees prone to overfitting.

1. Overfitting Due to Presence of Noise.
2. Overfitting Due to Lack of Representative Instances.
3. Overfitting and the Multiple Comparison Procedure.

Q6. What is an ensemble technique in machine learning?

Ans: ensemble technique aims at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly.

Q7. What is the difference between Bagging and Boosting techniques?

Ans: Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification. It attempts to increase the weight of an observation if it was erroneously categorized. Boosting creates good predictive models in general.

Q8. What is out-of-bag error in random forests?

Ans: The out-of-bag (OOB) error is the average error for each Zi calculated using predictions from the trees that do not contain Zi in their respective bootstrap sample. This allows the Random Forest Classifier to be fit and validated whilst being trained.

Q9. What is K-fold cross-validation?

Ans: Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

Q10. What is hyper parameter tuning in machine learning and why it is done?

Ans: Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

Q11. What issues can occur if we have a large learning rate in Gradient Descent?

Ans: For the gradient descent algorithm to reach the local minimum we must set the learning rate to an appropriate value, which is neither too low nor too high. This is important because if the steps it takes are too big, it may not reach the local minimum because it bounces back and forth between the convex function of gradient descent.

Q12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Ans:

Q13. Differentiate between Adaboost and Gradient Boosting.
Ans: AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.

Q14. What is bias-variance trade off in machine learning?

Ans: If the algorithm is too simple (hypothesis with linear eq.) then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex (hypothesis with high degree eq.) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as Trade-off or Bias Variance Trade-off. This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can't be more complex and less complex at the same time.

Q15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Ans: Linear Kernel: It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for text-classification problems as most of these kinds of classification problems can be linearly separated. Linear kernel functions are faster than other functions.

RBF Kernel: It is one of the most preferred and used kernel functions in svm. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data.

Polynomial Kernel: It is a more generalized representation of the linear kernel. It is not as preferred as other kernel functions as it is less efficient and accurate.