1. **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?**

Answer: RSS is a better measure of goodness of fit model in the regression.

R-squared shows how well the data fit a regression curve i.e. it depends on the quality of the fit of the curve on the data and does not depend upon the number of data points in the data, while, RSS is the sum of the square of the error made on each data points of a model i.e. it depends upon the number of data points in the data as well as the quality of fit on the data.

2. **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.**

Answer: TSS(Total Sum of Square) is a measure of the total variability in your data.

ESS(Explained Sum of Squares) is the variability of the data which has been explained by the model itself, hence meaning a better model.

RSS is the sum of the square of the error made on each data points of a model i.e. it depends upon the number of data points in the data as well as the quality of fit on the data.

In Regression, the TSS equals the RSS + ESS i.e. TSS=RSS+ESS

3. **What is the need of regularization in machine learning?**

Answer: The need of regularization in machine learning is that the model does not become excessively complex and overfit the training data as well as the model gets the to capture the pattern from the training data which can be inexplicit to unseen data.

4. **What is Gini–impurity index?**

Answer: Gini–impurity measures the impurity of a dataset. The more impure the dataset is the higher is the Gini-impurity index i.e. if a dataset has datapoints belonging to more than one label then Gini-impurity index>0 and if the dataset has datapoints belonging to one label then Gini-impurity index=0.

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

Answer: Yes, unregularized decision-trees are prone to overfitting, because, the decision-trees will most likely learn each and every data point in the training dataset without any restriction which will lead to learn the training patterns too closely and it will be tasted on unseen data which is mostly going to perform poorly. Hence, we have to unregularized decision-trees either by Controlling the depth of the tree or controlling the maximum number of datapoints the tree have.

6. What is an ensemble technique in machine learning?

Answer: In this technique, the models act as complementary to each other to get a better predictive performance compared to a single model. The individual models in the technique act as complementary to each other so if any one model acts poorly then there is another model in the ensemble which takes up to the poor performance of the previous model.

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

Answer: Bagging is the ensemble technique in which the trees learn from each other independently in parallel and the final result to determine the model average is by combining.

Boosting is the ensemble technique in which trees are trained in series so that it learns sequentially where each tree works on the error of the previous trees until the errors are minimized to a level.

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

Answer: out-of-bag errors is used to evaluate the performance on the random forest unseen data which is computed using the samples that were not included in the training of the individual trees.

9. What is K-fold cross-validation?

Answer: K-fold cross-validation is an evaluation method to find out how well your trained model can predict the result of the unseen data where the data sample is split into K number of smaller samples or folds on the training data, one fold is used as the validation set and the remaining folds as the training set.

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

Answer: hyper parameter tuning is a technique to find the best possible values of a set of hyper parameters used in model on which the model gives the best performance. This process is repeated with another set of vales for the same hyperparameters until optimal error is attained.

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

Answer: The gradient may keep on oscillating and not settle at the optimal solution if there is large learning rate.

The gradient descent can also diverge from the optimal solution if there is large learning rate.

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

Answer: No, we cannot Logistic Regression for classification of Non-Linear Data because, the decision produced by logistic regression is linear i.e. the classifier needs the input data to be linearly separable.

13. Differentiate between Adaboost and Gradient Boosting.

Answer: In this both techniques, the trees are trained in series. The main difference is in Adaboost we assign the weights to each of the data points of the training data sets and the weight changes according to the errors made by the previous tree in series while in Gradient each tree is trained on the error made by the previous trees.

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

Answer: It refers to the balance between error made by the models(bias) and how much the model changes with the change in the training data(variance). For example, if there is a simple model that means the model bias increases but the variance decreases whereas if the model is complex then the bias decreases and variance increases, so there is an interchange or trade-off in bias and variance.

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

Answer: The SVM uses kernels to transform the data from one set to another set so that decision in the resultant space is simpler that the decision in the original space that means depending upon the nature of data we have we need to select the kernel.

*If original data is linearly separable then we will use Linear kernel.

*If original data is in a complex pattern, then we will use RBF.

* If original data is in the form of Polynomial of some degree n, then we will use Polynomial kernel.