MACHINE LEARNING

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?

Ans. The residual sum of squares is the absolute amount of explained variation whereas R-squared is the absolute amount of variation as a proportion of total variation. R-squared measures the goodness of fit of model. Hence, a higher R-squared indicates the model is a good fit while a lower R-squared indicates the model is not a good fit. On the other hand, RSS measures the residuals of regression model, the smaller amount of residuals - the better model fits the data, the larger amount of residuals – the poorer model fits 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.

Ans. TSS measures that how much variation there is in the observed data.

RSS is the absolute amount of explained variation, measures the residuals of model.

ESS is the sum of squares of deviations of predicted values from the mean value of response variable.

TSS = ESS + RSS

TSS = Total sum of squares

ESS = Explained sum of squares

RSS = Residual sum of squares

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

Ans. Regularization is used in machine learning to prevent the overfitting or underfitting the model. It tries to achieve the generalized model. Using the regularization, we can fit our model on a given dataset and reduce errors in it.

4. What is Gini–impurity index?

Ans. Gini-Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.

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

Ans. By learners, can be created over-complex trees that do not generalize the data well. So, model can be overfit. By small variations in data, result may be completely different.

6. What is an ensemble technique in machine learning?

Ans. It is a technique in Machine Learning which uses multiple base models and combines all and produces one final model for improving the accuracy of result of final model.

7. 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.

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

Ans. The out of bag error is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample.

9. What is K-fold cross-validation?

Ans. In k-fold cross-validation, we split dataset into k-sets, and then we use one set for testing and remaining all sets used for training. Similarly, each time we use one set for testing and remaining sets for training. We will take each time different set.

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

Ans. In machine learning, hyperparameter tuning is a process under which some values are used to control the learning process of model. It can give optimized values, which maximizes the model's predictive accuracy.

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

Ans. Gradient Descent is too sensitive to the learning rate. If it is too big, this may bypass the local minima and overshoot.

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

Ans. Non-linear data problems cannot be solved with Logistic Regression because it has a linear decision surface.

13. Differentiate between AdaBoost and Gradient Boosting.

Ans. AdaBoost is the first designed algorithm with a particular loss function. Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problems which makes Gradient Boosting more flexible than AdaBoost.

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

Ans. Bias-variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

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

Ans. The linear, polynomial and RBF or Gaussian kernel are simply different in case of making the hyperplane decision boundary between the classes. The kernel functions are used to map the original dataset (linear / nonlinear) into a higher dimensional space with view to making it linear dataset.