**MACHINE LEARNING 5**

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: - In regression models, R-squared is a goodness-fit-measure. It takes into account the strength of the relationship between the model and the dependent variable. Its convenience is measured on a scale of 0 – 100%.

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: -  The TSS tells you how much variation there is in the dependent variable. (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable. A residual sum of squares (RSS) is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself. Instead, it estimates the variance in the residuals, or error term.

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

Answer: - Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

4. What is Gini–impurity index?

Answer: - Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.

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

Answer: - Yes, **Decision trees** are **prone to overfitting**, especially when a **tree** is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

6. What is an ensemble technique in machine learning?

Answer: - Ensemble methods are **techniques that create multiple models and then combine them to produce improved results**. Ensemble methods usually produces more accurate solutions than a single model would.

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

Answer: - Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.

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

Answer: - The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample.

9. What is K-fold cross-validation?

Answer: - K-Folds cross validation is one method that attempts to maximize the use of the available data for training and then testing a model. It is particularly useful for assessing model performance, as it provides a range of accuracy scores across (somewhat) different data sets.

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

Answer: - In machine learning, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process.

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

Answer: - When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.

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

Answer: - No, Logistic Regression has traditionally been used as a linear classifier.

13. Differentiate between Adaboost and Gradient Boosting.

Answer: - 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**.

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

Answer: - the bias–variance tradeoff is **the property of a model that the variance of the parameter estimates 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.

Answer: - The linear, polynomial kernel and RBF 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.