MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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 (RSS) measures the level of variance in the error term, or residuals, of a regression model.
- The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.
- A value of zero means your model is a perfect fit.
- Statistical models are used by investors and portfolio managers to track an investment's price and use that data to predict future movements.
- The RSS is used by financial analysts in order to estimate the validity of their econometric models.

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- **1. Total sum of squares**

The total sum of squares is a variation of the values of a dependent variable from the sample mean of the dependent variable. Essentially, the total sum of squares quantifies the total variation in a sample. It can be determined using the following formula:

Ans- **2. Regression sum of squares (also known as the sum of squares due to regression or explained sum of squares)**

The regression sum of squares describes how well a regression model represents the modeled data. A higher regression sum of squares indicates that the model does not fit the data well.

Ans- **3. Residual sum of squares (also known as the sum of squared errors of prediction)**

The residual sum of squares essentially measures the variation of modeling errors. In other words, it depicts how the variation in the dependent variable in a regression model cannot be explained by the model. Generally, a lower residual sum of squares indicates that the regression model can better explain the data, while a higher residual sum of squares indicates that the model poorly explains the data.

Formula- total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

3. 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.

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- 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?

   Ans- An ensemble method is **a technique which uses multiple independent similar or different models/weak learners to derive an output or make some predictions**. For e.g. A random forest is an ensemble of multiple decision trees.

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- Out-of-bag (OOB) error, also called out-of-bag estimate, is **a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging)**. Bagging uses subsampling with replacement to create training samples for the model to learn from.

9. 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.

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

    Ans- Hyperparameter tuning is **choosing a set of optimal hyperparameters for a learning algorithm**. A hyperparameter is a model argument whose value is set before the learning process begins. The key to machine learning algorithms is hyperparameter tuning.

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

    Ans- The learning rate can seen as step size, η. As such, gradient descent is taking successive steps in the direction of the minimum. If the step size η is too large, **it can (plausibly) "jump over" the minima we are trying to reach, ie. we overshoot**.

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

    Ans- **Logistic Regression has traditionally been used as a linear classifier**, i.e. when the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary…

13. 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.

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

    Ans- In statistics and machine learning, the 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- RBF Kernel is popular because of its similarity to K-Nearest Neighborhood Algorithm. It has the advantages of K-NN and **overcomes the space complexity problem** as RBF Kernel Support Vector Machines just needs to store the support vectors during training and not the entire dataset