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) is the absolute amount of explained variation, whereas R-squared is the absolute amount of variation as a proportion of total variation. R squared, the proportion of variation in the outcome Y, explained by the covariates X, is commonly described as a measure of goodness of fit.
A higher R-squared value indicates a higher amount of variability being explained by our model and vice-versa. If we had a really low RSS value, it would mean that the regression line was very close to the actual points. This means the independent variables explain the majority of variation in the target variable.

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**.)** **RSS** : Residual Sum of Squares (RSS) is a statistical method that helps identify the level of discrepancy in a dataset not predicted by a regression model. Thus, it measures the variance in the value of the observed data when compared to its predicted value as per the regression model. Hence, RSS indicates whether the regression model fits the actual dataset well or not.

TSS: The Sum of Squared regression is **the sum of the differences between the predicted value and the mean of the dependent variable**.

ESS : The sum of squares is **a statistical measure of deviation from the mean**. It is also known as variation. It is calculated by adding together the squared differences of each data point. To determine the sum of squares, square the distance between each data point and the line of best fit, then add them together.

In statistics, the explained sum of squares (ESS), alternatively known as the model sum of squares or sum of squares due to regression (SSR – not to be confused with the residual sum of squares (RSS) or sum of squares of errors), is a quantity used in describing how well a model, often a regression model, represents the data being modelled. In particular, the explained sum of squares measures how much variation there is in the modelled values and this is compared to the total sum of squares (TSS), which measures how much variation there is in the observed data, and to the residual sum of squares, which measures the variation in the error between the observed data and modelled values.

TSS = ESS + RSS

where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Suqares. The aim of Regression Analysis is explain the variation of dependent variable Y.

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. Gini index or 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?

Ans.)Yes, Regularization in terms of decision tree means to control the growth of the tree. When decision tree becomes too large they tend to over-fit. To avoid over-fitting, we regularize the tree. By doing so, the tree does not grow to its full potential, i.e, it gets restricted.

6. What is an ensemble technique in machine learning?

Ans.) In ensemble technique we combine together a number of models to get a better performing model on the dataset. The individual models in the ensemble technique act as complementary to each other, so if one model work poorly on some area, then there is some other model in the ensemble which takes up on this weakness of the previous model. So, in this way the models act as complementary to each other and overall after ensemble them we get a better model.

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

Ans.)Bagging is the ensemble technique in which N decision trees are trained in parallel on N Randomly Generated datasets from the training data. The final result of the ensemble is produced by taking average of results of all the member decision trees for regression and for classification we take mode of the classes predicted by the member

trees. Example-Random Forest. Boosting is the ensemble technique in which trees are trained in series rather than parallelly. In boosting, each tree works on the errors of the previous class until errors are minimized to the level we want. examples are Gradient boosting, Ada boost.

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

Ans.) Out of bag error in the random forests is used to evaluate the random forest model. As we know in random forest a number of decision trees are trained in parallel on bootstrapped samples. while training a particular tree the data points which are not used for training of that tree act as unseen data. For each data point predictions are made by those trees in whose training these points were not included and error on these predictions are called out of bag error.

9. What is K-fold cross-validation?

Ans.)K-fold cross validation is the model evaluation technique which is generally used when we have limited data. In this technique we create k groups on the training data, train the model on k-1 groups and test the resultant model on the left-out group. In this we do it on every possible combinations of groups and then take average of the evaluation metric.

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

Hyper parameter tuning is the technique of finding best possible values of a set of hyper parameters used in model on which the model gives best performance. So, hyper parameter tuning is like fine tuning your model on the basis of hyper parameter values used in the model so that we get the best possible version of that model.

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

Ans.)The two main issues which can occur if we perform gradient descent with large learning rate are: • The gradient descent algorithm can diverge from the optimal solution if we try out a very large learning rate. The algorithm can go away from the optimal solution if we have a very large learning rate. • The gradient may simply keep oscillating around the optimal solution if the learning rate is high, and it will not settle at the optimal solution.

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

Ans.)We cannot use Logistic Regression for classification of Non-Linear Data because the decision boundary produced by logistic regression is linear and if we have nonlinear data where we have nonlinear decision boundaries then if we try to use the logistic regression it will perform poor on the data, as the decision boundary in the data is nonlinear.

13**. Differentiate between Adaboost and Gradient Boosting.**

Ans.)Adaboost and Gradient Boosting are ensemble techniques, in which the trees are trained in series. The major difference between Adaboost and Gradient Boosting is that in Adaboost we assign weights to each of the data points of the training data and the weights changes according to the errors made by the previous tree in the series. So basically a tree in Adaboost puts more emphasis on the datapoints on which the previous tree did not perform well. While in Gradient Boosting, each tree is trained on the errors made by the previous tree, so in this way the error made on the training data keeps on decreasing as we keep increasing the trees in the series.

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

Ans.)The bias variance tradeoff is the tradeoff which happens between bias (error made by the model) and variance (how much the model changes with change in training data) when the model complexity changes. If the model is too simple the model makes too much errors and its predictions becomes inaccurate, so when we decrease the model complexity the bias (or errors) increases but the variance decreases (that is the model does not change much with change in training data). On the other hand, if model is too complex the bias although becomes low but the variance increases. So, there is tradeoff between bias and variance. So, we always have to find that sweet spot where the model does not have much bias neither much high variance.

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

Ans.)The SVM uses kernel functions to transform the data from one set of dimensions to another set of dimensions so that the decision boundary in the resultant space is simpler than the decision boundary in original space. Now, the kernel to be used depend upon the nature of the data we have. • Linear kernel will be used if the original data is linearly separable. • Polynomial kernel is used when the data is the form of polynomial of some degree n. • RBF is used when the data follows some complex pattern which is neither linear nor polynomial