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?**

ANSWER:- R2 (R Squared) is statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model. In other hand 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. R-Squared is the better measure of goodness of fit compared to RSS. R-squared explains to what extent the variance of one variable explains the variance of the second variable. So, if the R2 of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.

2.What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sumof Squares) in regression. Also mention the equation relating these three metrics with each other.

ANSWER:- TSS: TSS or Total sum of squared is given by the summation of deviation of ground truth from the mean of the variable.

ESS: ESS of Explained sum of squares is given by the summation of squares of the deviation of the predicted value from the mean of the variable.

RSS: RSS or Residual sum of squares is given by the summation of squares of error values i.e., ground value – predicted value.

The relation between the above 3 could be linearly expressed as: TSS = RSS + ESS.

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

ANSWER:- Regressions come with a penalty known as regularization. Regularization limits or reduces the coefficient in the direction of zero. The term "regularization" describes methods for calibrating machine learning models to reduce the adjusted loss function and avoid overfitting or under fitting. We can properly fit our machine learning model on a particular test set using regularization, which lowers the mistakes in the test set.

4.What is Gini–impurity index?

ANSWER:- The Gini impurity index determines the probability that a particular characteristic will be incorrectly classified when selected at random. It can be said to as pure if every element is connected to a
single class. By deducting the sum of the squared probabilities of each class from one, the Gini impurity is identified. When tree algorithms like Decision Tree or Random Forest are used in the modelling, this metric is calculated.

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

ANSWER:- Yes, decision tress are prone to overfitting. But unlike other algorithms decision tree does not use regularization to fight against overfitting. Instead it uses pruning. There are mainly to types of pruning performed:
Pre-pruning that stop growing the tree earlier, before it perfectly classifies the training set. Post-pruning that allows the tree to perfectly classify the training set, and then post prune the tree.

6.What is an ensemble technique in machine learning?

ANSWER:- Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning. There are 2 types of Ensemble techniques, Bagging and Boosting.

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

ANSWER:- A machine learning ensemble meta-algorithm called bagging, also known as bootstrap aggregating, and aims to increase the stability and precision of machine learning algorithms. By combining several weak classifiers, the ensemble modelling technique known as "boosting" aims to create a powerful classifier. It is accomplished by using weak models in series to develop a model.
In Bagging each model receives equal weight. But in Boosting, models are weighted according to their performance.
In Bagging each model is built independently. But in Boosting, new models are influenced by the performance of previously built models.
The base classifiers are trained parallelly in Bagging. But in Boosting, classifiers are trained sequentially.

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. This allows the Random Forest Classifier to be fit and validated whilst being trained. This is an advantage if:

a- Your data set is too small to split in to training, validation and test.
b- Gives a second validation on the model allowing.

9.What is K-fold cross-validation?

ANSWER:- In k-fold cross-validation, the input data is splitted into k subsets of data (also known as folds). We train an ML model on all but one (k-1) of the subsets, and then evaluate the model on the subset that was not used for training. This process is repeated k times, with a different subset reserved for evaluation (and excluded from training) each time. The K in K fold is the integer defining how any times does the subset should be created and trained and tested. For example a 5 Fold cross validation will create 5 subsets in both training and testing dataset, train and predict are output 5 accuracy values.

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

ANSWER:- Hyper parameter tuning consists of finding a set of optimal hyper parameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyper parameters maximizes the model’s performance, minimizing a predefined loss function to produce better results with fewer errors. A model hyper parameter is a configuration that is external to the model and whose value cannot be estimated from data. They are often used in processes to help estimate model parameters.

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. It also can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.

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

ANSWER:- No, we cannot use Logistic Regression for classification of Non-Linear Data. Non-linear problems can't be solved with logistic regression because it has a linear decision surface. It has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries.

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. In the case of Gradient Boosting, the shortcomings of the existing weak learners can be identified by gradients and with AdaBoost, it can be identified by high-weight data points.
AdaBoost minimises the exponential loss function that can make the algorithm sensitive to the outliers. With Gradient Boosting, any differentiable loss function can be utilized. Gradient Boosting algorithm is more robust to outliers than AdaBoost.

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

ANSWER:- There is a tradeoff between a model's ability to minimize bias and variance. Understanding these two types of error can help us diagnose model results and avoid the mistake of over- or under-fitting. This is knwn as bias-variance tradeoff.

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

ANSWER:- Support Vector Machine, or SVM for short, is a supervised machine learning technique that can be applied to classification or regression problems. SVM employs many kernels, such as linear, RBF, and polynomial kernels. The extremely underdeveloped linear functions can be used with a linear kernel. The size of the function class grows as the order of the polynomial kernel rises. By raising the polynomial kernel's power, we can easily calculate the dot product. For more, Gaussian RBF (Radial Basis Function) is a well-liked Kernel technique used in SVM models. The value of an RBF kernel relies on how far it is from the origin or another location.