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 1 – R-squared is a better measure of goodness of fit model in regression as is shows the absolute amount of variation as a proportion of total variation.

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 2 – The equation is
   Total sum of squares = residual sum of squares + explained sum of squares
   Total sum of squares is defined as the sum of the squares of the difference between actual data points and mean of target variable. Formula for TSS is = from i to n ∑(yi – mean(y))^{2} , this shows the amount of variance in dependent variable.
   Explained Variance is defined as the sum of squares of the difference between predicted target variable and mean of the target variable. Formula for ESS is = from i to in ∑(y(predicted) – mean(y))$^{2.}$
   Residual sum of squares is the sum of the squares of the difference between the predicted target variables and actual target variable.

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

   ANSWER 3 – Regularization is needed to reduce the error and model complexity. Regularization helps in preventing overfitting in a model. The most common regularization techniques are L1 regularization and L2 regularization.

4. What is Gini–impurity index?

   ANSWER 4 – As we know that for building a decision tree, we use either Gini entropy or Gini impurity to calculate the information gain for the splits. Gini entropy shows us homogeneity of a sample. If the sample is completely homogeneous then the entropy value is 0 and if the sample is an equally divided it has the entropy of 1. This shows how much purity the split has. If the entropy is 0 that means that the split is pure, if the entropy is 0 that means the split is not pure (include both the classes equally).
   In case of Gini Impurity instead of 1 we get 0.5 if the sample is equally divided. And the formula for Gini Impurity is = 1 – from i to n ∑(p)^{2}. It means the sum of the squares of probability of positive class and negative class. Gini Impurity is computationally easy than Gini Entropy

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

   ANSWER 5 – Yes, if a decision tree is very dense and there is no regularization is done it is more likely that the model will be an overfitting model. Regularization helps to prevent a model from overfitting, in case of Decision tree pruning, Max depth, ensemble (bagging or boosting), max number of leaf nodes etc., can be used as a regularization.

6. What is an ensemble technique in machine learning?

   ANSWER 6 – Ensemble technique in machine learning is widely used. In this technique many models are created and then combine all the models to produce/get an improved result. There 4 different types of Ensemble techniques: Bagging, Boosting, Stacking, Cascading.
   Bagging and Boosting are most widely used technique. Bagging technique is used in Random Forest Decision models, while Boosting techniques are used in XGboost, Adaboost classification techniques.

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

   ANSWER 7 - In Bagging (boot strap aggregation) we train multiple models using dense decision tree, each decision tree shows high variance, number of Decision trees are known as number of estimators. Each Decision tree will train on random subset of data. Then the result from each decision tree is calculated and based on the majority final output is classified.

   However, in Boosting technique, we train shallow/weak decision trees (stump decision trees). In this method the accuracy gets improved with each iteration, that iteratively adjusts the weight of . observation as per the last classification done by the model. If an observation is incorrectly classified, it increases the weight of that observation

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

   ANSWER 8 – In Random Forest we use Bootstrap aggregation technique (bagging), in which we take number of samples from a bag of data with replacement. Suppose out of population of 100 observation we take 75% of data for bagging method, means 75% of the data will be taken randomly with replacement to create samples. Now 75% of the data will be training data for each sample, remaining 25% will be considered as out of bag sample. The model prediction is done on Out of Bag samples. Based on majority votes we finalise the results or classify the label for test data. When there is a wrong classification done for any observation in out of bag, that will be called as an out of bag error. Out of bag samples are used only for prediction not for training. When we aggregate the error from each of the out of bag samples that will be considered as out of bag error.

9. What is K-fold cross-validation?

   ANSWER 9 – One main problem in machine learning is less data and imbalanced data. A machine learning model needs data to learn and give more accurate results. In general, we split the data into train and test, however in that case if the data size is small or imbalanced our model will not be able to learn the data better and it will have more bias towards domination class variables in case of classification. To solve this issue, we use cross validation. Cross validation is a technique where we train our data with different training and testing sets of data. K fold cross states that we partition the data set into K bins of equal size for example if we have 200 data points, and we want 5 equal bins of data, it will be 5 set of 40 data points each. Now from these sets we will assign 1 set as test and other 4 as train. And we will train the model K times so that each bin will be tested. By performing this we will get rid of the problem of less data set and also the problem of Imbalanced data set will be reduced.

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

    ANSWER 10 – Hyper parameter tuning is one of the most important things for any machine learning or deep learning model building. With the use of hyper parameter, we control and optimise the machine learning or deep learning model. The main of aim for every machine learning model and deep learning model is to have a perfect trade off between bias and variance. Hyper Parameter tuning helps us to maintain that trade off so that our model will not be overfitted or under fitted.
    Different algorithms have different hyper parameters associated with them.

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

    ANSWER 11 – Learning rate controls the step size for a function to converge to its minimum, is the learning rate is too small then we will end up to the problem of vanishing gradient which means it will take a lot of time for the function to converge to its minimum or may be it will not converge as the weights will not get updated after a period of times due to small learning rate, And if the learning rate is high we will end up to the problem of Exploding gradient, which means the function will get out of control and will not be converge to its minimum. Exploding gradient causes underfitting of the model.

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

ANSWER– Logistic regression is indeed non linear in terms of Odds and Probability

13. Differentiate between Adaboost and Gradient Boosting.

ANSWER 13 – In Adaboost at each iteration, adaptive boosting changes the sample distribution by modifying the weights attached to each of the instances. It increases the weights of the wrongly predicted instances and decreases the ones of the correctly predicted instances. The weak learner thus focuses more on the difficult instances.

gradient boosting doesn’t modify the sample distribution. Instead of training on a newly sample distribution, the weak learner trains on the remaining errors (so-called pseudo-residuals) of the strong learner.

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

ANSWER 14 – Bias Variance trade-off is the most important aspects for any machine learning or deep learning model. Bias means simplifying assumption made by the model however variance means changing the function with respect to the training data. High bias means under fitting as we are simplifying our model where High variance means Overfitting as we are trying to make a more complex decision boundary and changing it with respect to our training data. High variance leads to drastic change in model if there is a slight change in the training data set. For machine learning or deep learning setting this trade-off between bias and variance is very crucial.

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

ANSWER 15 – LINEAR KERNEL - Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. Training a SVM with a Linear Kernel is Faster than with any other Kernel.

RBF Kernel stands for radial based kernel, It is used when the data is non linearly separable, It is a default kernel used in SVM. RBF kernel is a function whose value depends on the distance from the origin or from some point.

Polynomial Kernels – In polynomial kernel we simply calculate the dot product by increasing the power of the kernel.

Kernels in svm are formed only due to the dual formation of SVM using LaGrange multiplier.