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 ->

2. 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 -> Total Sum of Square -> TSS(Total Sum of Square) The Difference between acutal data and the mean is known as TSS .

   Lets Under see the .

   TSS -> $\mathrm{TSS} = \sum (y_i - \bar{y})^2$

   Yi = Actual Data
   Y Bar = Mean of all the actual data .

   Exlained Sum of Square -> The Explained SS tells you how much of the variation in the dependent variable your model explained.

   ESS = Σ(Y-Hat – mean of Y)^2.

   Residual Sum of Squares -> In Residual Sum of square we can find the difference between Actual data and Predicted Data . and then we are square all the data ..

   Formula of RSS =

   $$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

   Here ,

   Yi= Actual data
   Ycap = Predicted Data

The relationship between the three types of sum of squares can be summarized by the following equation:

Relationship Formula TSS=SSR+SSE

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

Ans -> your model is learning too fast we have to basically pull it we have to penalise , we have to make it learn as slow as possible so that it achieve the better result thats why we need regularization.

4. What is Gini–impurity index?
   Ans ->
   Ginni indexing what happen it will check each and every feature . and it will check how much impurity each features has ,
   How much non sense data . techincall we called as a Ginni Impurity.
   means less the impurity more then information.

   Gain(T.X)= Entropy(T)-Entropy(T.X)

   in the ginni indexing we can find ginni impurities. which feature have less impurities it will provide you each feature impurities in terms of an number. so which ever is having less impurities so that is going to root node.

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

   Ans -> 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?
   Ans -> We regularly come across many games show on television and you must have noticed an option of 'Audiance Poll' most of the times a contestent goes with the option which has the highest vote from the audiance and most of the times they win. we can generalize this in real life as well where taking opinions from a majority of people is much more preferred than the opinion of a single person. Ensembles technique has a similar underlying idea where we aggregate predictions from a group of a predictors . which may be classifier or regressor . and most of the time the prediction is better than the one obtained using a single predictor such algorithm are called ensemble methods and such predictors are called ensembles.

   When i say ensemble approaches there are two types .
   1- Bagging - In bagging we have couple of classifier algorithms.
   2- Boosting - Boosting also we are going to discuss couple of algorithm . we get into this boosting lates .
7. What is the difference between Bagging and Boosting techniques?

1- Training data subsets are drawn randomly with replacement from the entire training dataset.
2- Bagging attempts to tackle the over-fitting issue.
3- Every model receives an equal weight.10Every model receives an equal

4- Objective to decrease variance, not bias.
5- Every model is built independently.

1. Each new subset contains the components that were misclassified by previous models.
2. Boosting tries to reduce bias.
3. Models are weighted by their performance.10:42
4. Objective to decrease bias, not variance.
5. New models are affected by the performance of the previously developed model.

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

   Ans -> We build multiple models every model will give independent decision then we will go with a final decision based on the majority. So , whenvever they use the data they will get trained every thing will be so for the testing . so there is a small set of a data . Some portion of that data . i am keep a side . That i will use for a testing so that set of data called out of bag evaluation . All things happen internally . Some portion of data will never be sampled will never go to anything so that data used for the testing . that set of data is called out of bag evaluation . Because this models never trained on this and have not seen this so everything happen internally and directly come with a result other wise you have to do train test
9. What is K-fold cross-validation?

Ans -> k fold cross validation comes into a picture .. Lets say whenever i say cross validation = 5 , we have to tell how many iteration and folds K means --> How many fold there is 5 Fold.

How its Work ?

This dataset trained 5 times . and within iteration 1 in that this whole dataset .it is going to split in 5 pieces / 5 sets (5 iteration/5split) it will divide in equal parts suppose there is a total 100% means 80% of the data kept for model training and 20 % will be use for testing.

Similarly 2nd iteration will take second part of the iteration for testing and left 5 part is use for training . similary its happen to all iteration. it depend on how much you give cross validation .

--> It means every iteration our testing data is going to be different and our training data is also changing partially.

Then after testing we are going to find accuracy of each training and test data set and we claculate the mean of all set of iteration.

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

    Ans ->

Just look the picture in above picture . what tell this picture to us . Just understand this points suppose we are listning a radio the one which has a nob , not the electronic.

So you are trying to listen songs and one situation is playing good song and you are not able to hear it properly . so you are trying to adjust your knob turn, right , left, slowly some time fast . untill you get there is no noise in the music . the knob you are going to tune it . why are adjusting to listen quality sound.

Similarly we going to adjust our parameter.

Q- What are those parameter ? i need to change some parameter , some of the internel things that model is assuming and building the model . by default value it will take and build and it will run.

i wanted to change the default values like .

K=3

K=5

We dont know 3 is best or 5 is best we dont know which k values is best it is taking which ever it wants to take and execute it and train it .

Not i wanted to change that number so that may be increase the accuracy . so that is what tuning here.

What are you tuning -- The parameter

What are the parametere - Every algorithm has different parameter .

Maximum time it will increase the accuracy.

There are two popular Hyperparameter tuning

1- GridSearchCV

2-RandomSearchCV

This are the technique we used to tune the parameter.

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

    Ans -> When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error. When the learning rate is too small, training is not only slower, but may become permanently stuck with a high training error.

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

    Ans -> No because Logistic regression is considered a generalized linear model because the outcome always depends on the sum of words, the output cannot depend on the product (or quotient the inputs and parameters. Or in other, etc.) of its parameters.

13. Differentiate between Adaboost and Gradient Boosting .

    Ans -> Adaboost is more about ‘voting weights’ and Gradient boosting is more about ‘adding gradient optimization’. Adaboost increases the accuracy by giving more weightage to the target which is misclassified by the model. At each iteration, Adaptive boosting algorithm 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.

    Adaboost is more about ‘voting weights’ and Gradient boosting is more about ‘adding gradient optimization’. Adaboost increases the accuracy by giving more weightage to the target which is misclassified by the model. At each iteration, Adaptive boosting algorithm 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.

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

    Ans -> Bias - tendency to prefer one person or thing to another, and to favour that person or thing. To bias someone means to influence them in favour of a particular choice.

Variance: Variance is a statistical measure that tells us how measured data vary from the average value of the set of data. In other words, a variance is the mean of the squares of the deviations from the arithmetic mean of a data set.

Trade-off: Trade-off is tension between the error introduced by the bias and the variance.

Variance Trade off-> it is the case that if the variance is high, the bias is low, and vice versa. The model will overfitting and underfitting..

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

Ans- > 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. It is mostly used when there are a Large number of Features in a particular Data Set.

k(x,x1) = sum(x*x1)

RBF Kernel -> Radial Basis Kernel function that is used in machine learning to find a non-linear classifier or regression line.

Kernel Function is used to transform n-dimensional input to m-dimensional input, where m is much higher than n then find the dot product in higher dimensional efficiently

k(x,x1) = exp(-gamma*sum(x-x1^2))
Here gamma is a parameter which ranges from 0 to 1.

Polynomial kernels -> : A polynomial kernel is a more generalized form of the linear kernel. The polynomial kernel can distinguish curved or non linear input space.In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors (training samples) in a feature space over polynomials of the original variables, allowing learning of non-linear models
k(x,x1) = 1+sum(x*x1)^d where d = degree of polynomia