Machine Learning

Q1. 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: 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.

e1= (x1, y1), e2= (x2, y2), e3= (x3, y3)

RSS=(e1) **2+(e2) **2+(e3) **2+(en) **2

R-squared is a statistical measure that represents the goodness of fit of a regression model. The ideal value for r-square is 1. The closer the value of r-square to 1, the better is the model fitted.

R-square is a comparison of the residual sum of squares (SS res) with the total sum of squares (SS tot).

Formula **R2=1- (SS res/SS tot)**

Q2. 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:

Residual Sum of Square (RSS)

RSS is the sum of squares of error terms and this is an absolute number. This number can vary greatly based on the number of data points. RSS will increase if the data points are more and the value will decrease If the data points are less.

$$RSS = (y_1 - y_{pred})^2 + (y_2 - y_{pred})^2 + - + (y_n - y_{pred})^2$$

Total Sum of Squares

The total sum of squares (TSS) is a variation from the actual value from the mean value of data points.

TSS=sum of where (I=1) to (n= is total number) (yi data point – Y Mean) squared.

Where

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Where:

$\bar{y} = mean\ of\ y\ data\ points$

$y_i = the\ value\ of\ ith\ data\ ponts$

R-squared

R squared is a measure of variance for dependent variables. That is variance in the output that is explained by the small change in input.

$$R^2 = 1 - \frac{RSS}{TSS}$$

The value of *R-sq* is always between 0 (0%) and 1 (100%). The bigger the value better the fit to the model.

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

Ans: While training a machine learning model, the model can easily be overfitted or underfitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization technique help reduce the chance of overfitting and help us get an optimal model.

This Regularization technique is to prevent the model from overfitting by adding extra information to it. It mainly regularizes or reduces the coefficient of features towards zero, In regularization technique we reduce the magnitude of the features by keeping the same number of features.

Technique of Regularization.

- Ridge Regression (L2 Regularization)
- Lasso Regression (L1 Regularization)
- Elastic Net

The Elastic Net Regularization technique that uses both L1 and L2 regularizations to produce most optimized output.

Lasso regression and Ridge regression is an extension to linear regression where we want to minimize the following loss function.

Q4. What is Gini-impurity index?

Ans: Decision trees are often used while implementing machine learning algorithms. The hierarchical structure of a decision tree leads us to the final outcome by traversing through the nodes of the tree. Each node consists of an attribute or feature which is further split into more nodes as we move down the tree.

But how do we decide:

- Which attribute/feature should be placed at the root node?
- Which features will act as internal nodes or leaf nodes?

To decide this, and how to split the tree, we use splitting measures like Gini Index, Information Gain, etc. 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. More precisely, the Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans: Yes, unregularized decision trees are prone to overfitting. 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. 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 stops 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.

Q6. What is an ensemble technique in machine learning?

Ans: Ensemble Technique is one of the Machine Learning approaches for improving the model by combining several Models.

- The main advantage of the Ensemble Technique is to reduce the variance and bias factor.
- It also helps to increase the model accuracy and reduce variability in prediction.

The Ensemble Technique includes:

- Bagging
- Boosting

Q7. What is the difference between Bagging and Boosting techniques?

Answer: Both Bagging and Boosting Technique are two Main types of Ensemble Learning Methods. The Main Difference between them is the way they are trained. In Bagging, the weak learners being trained in parallel but in boosting the they learn sequentially. (This Means that the series of Models are constructed and with each new model iteration, the weight of the miss classified data in the previous model are increased.) This redistribution of weights helps the algorithm identify the parameters that it needs to focus on to improve the performance. The Ada Boost stands for adaptive boosting algorithm, which is one of the most popular Boosting Algorithm. Other Difference is Bagging methods are typically used on weak learners which exhibit high variance and low bias, whereas boosting methods are leveraged when low variance and high bias is observed.

Q8. What is out-of-bag error in random forests?

Ans: Out-of-bag error, also called known as 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). When bootstrap aggregating is performed, two independent sets are created. One set, the bootstrap sample, is the data chosen to be "in-the-bag" by sampling with replacement. The out-of-bag set is all data not chosen in the sampling process.

Q9. What is K-fold cross-validation?

Ans: In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k − 1 sub samples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results can then be averaged to produce a single estimation.

Q10. What is hyper parameter tuning in machine learning and why it is done?

Ans: Models are full of Mathematics and it is beyond the human abilities to check each and every parameter of a model, if someone tries, it would be lengthy, time consuming and tuff task to work on single-single parameters. To avoid this, we would use Hyper parameter tunning, the models we use for hyper parameter tuning is grid search cv and random search cv. Grid search will thoroughly check the parameters and would give us the best parameters and model score. Whereas Random Search will go searching randomly. Then the best parameters will be determined along with score. These make model get the best model.

Q11. What issues can occur if we have a large learning rate in Gradient Descent?

Ans: The learning rate is the step size; Gradient descent is taking successive steps in the direction of the minimum. If the step size is too large then it would overshoot the minima we are trying to reach.

Q12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Ans: Logistic Regression can be used only for two class classification but if its multiclass then this can't be used for multiclass problems. Logistic Regression is known as a linear classifier. Logistic regression is considered as a linear model because the decision boundary it generates is linear, which can be used for classification purposes.

Q13. Differentiate between Ada boost and Gradient Boosting?

AdaBoost

AdaBoost or Adaptive Boosting is the first Boosting ensemble model. The method automatically adjusts its parameters to the data based on the actual performance in the current iteration. Meaning, both the weights for re-weighting the data and the weights for the final aggregation are re-computed iteratively.

Gradient Boosting

Gradient Boost is a robust machine learning algorithm made up of Gradient descent and boosting. The word 'gradient' implies that you can have two or more derivatives of the same function. Gradient Boosting has three main components: additive model, loss function and a weak learner.

Q14. What is bias-variance trade off in machine learning?

Ans: - If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand, if our model has large number of parameters then it's going to have high variance and low bias.

We need to find the right/good balance without overfitting and under fitting the data. This trade off in complexity is why there is a tradeoff between bias and variance. An algorithm can't be more complex and less complex at the same time.

Q15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Ans:

- 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 many Features in a particular Data Set.

- Gaussian RBF (Radial Basis Function) is another popular Kernel method used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format

- In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel.