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 ofgoodness of fit model in regression and why?**

Answer:-

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(RSS) with the total sum of squares(TSS). The total sum of squares is calculated by summation of squares of perpendicular distance between data points and the average line Whereas

The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model.

The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data.

Hence R2 is a better measure of goodness of fit model in regresson .

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:-

The total sum of squares (TSS) measures how much variation there is in the observed data, while the **residual sum of squares** measures the variation in the error between the observed data and modeled values.

RSS is a statistical method used to detect the level of discrepancy in a dataset not revealed by regression. If the residual sum of squares results in a lower figure, it signifies that the regression model explains the data better than when the result is higher. In fact, if its value is zero, it's regarded as the best fit with no error at all.

The **explained sum of squares (ESS)** is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model

TSS = ESS + RSS,

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

Answer:-

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.
It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features."

4. **What is Gini–impurity index?**

Answer:-

The Gini impurity index measures the diversity in a set. It also calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. Let's say, for example, that we have a bag full of balls of several colors. A bag where all the balls have the same color, has a very low Gini impurity index (in fact, it is zero). A bag where all the balls have different colors has a very high Gini impurity index.

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

Answer:-

The answer is YES because Overfitting can be one problem that describes if your model no longer generalizes well.

Overfitting happens when any learning processing overly optimizes training set error at the cost test error. While it's possible for training and testing to perform equality well in cross validation, it could be as the result of the data being very close in characteristics, which may not be a huge problem. In the case of decision tree's they can learn a training set to a point of high granularity that makes them easily overfit. Allowing a decision tree to split to a granular degree, is the behavior of this model that makes it prone to learning every point extremely well — to the point of perfect classification — ie: overfitting.

6. **What is an ensemble technique in machine learning?**

Answer:-

Ensembling is nothing but the technique to combine several individual predictive models to come up with the final predictive model. This type of machine learning algorithm helps in improving the overall performance of the model.
ensemble methods in machine learning employ a set of models and take advantage of the blended output, which, compared to a solitary model, will most certainly be a superior option when it comes to prediction accuracy.

Types of ensemble techniques:-

Bagging and Boosting,

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

Answer:-

Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data.
Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification. It attempts to increase the weight of an observation if it was erroneously categorized. Boosting creates good predictive models in general.

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

Answer:-

Generally, in machine learning and data science, it is crucial to create a trustful system that will work well with the new, unseen data. Overall, there are a lot of different approaches and methods to achieve this generalization. Out-of-bag error is one of these methods for validating the machine learning model.

This approach utilizes the usage of bootstrapping in the random forest. Since the bootstrapping samples the data with the possibility of selecting one sample multiple times, it is very likely that we won't select all the samples from the original data set. Therefore, one smart decision would be to exploit somehow these unselected samples, called out-of-bag samples.

Correspondingly, the error achieved on these samples is called out-of-bag error. What we can do is to use out-of-bag samples for each decision tree to measure its performance. This strategy provides reliable results in comparison to other validation techniques such as train-test split or cross-validation.

9. **What is K-fold cross-validation?**

Answer:-

K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation.

Cross-validation is usually used in machine learning for improving model prediction when we don't have enough data to apply other more efficient methods like the 3-way split (train, validation and test) or using a holdout dataset.

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

Answer:-

Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors. Hyperparameter tuning takes advantage of the processing infrastructure of Google Cloud to test different hyperparameter configurations when training your model. It can give you optimized values for hyperparameters, which maximizes your model's predictive accuracy.

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

Answer:-

Gradient Descent is too sensitive to the learning rate. If it is too big, the algorithm may bypass the local minimum and overshoot.

In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.

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

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

Answer:-

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.

In practice, this boosting technique is used with simple classification trees or stumps as base-learners, which resulted in improved performance compared to the classification by one tree or other single base-learner.

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.

The technique yields a direct interpretation of boosting methods from the perspective of numerical optimisation in a function space and generalises them by allowing optimisation of an arbitrary loss function.

In the case of AdaBoost, the shifting is done by up-weighting observations that were misclassified before, while Gradient Boosting identifies the difficult observations by large residuals computed in the previous iterations.

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

Answer:-

In machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

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. So we need to find the right/good balance without overfitting and underfitting the data.
This tradeoff 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.

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

Answer:-

In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification.

Linear Kernel
It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for text-classification problems as most of these kinds of classification problems can be linearly separated.
Linear kernel functions are faster than other functions.

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.