MACHINE LEARNING

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: R-squared is better than RSS.

   Reason: Residual Sum of Squares (RSS) range can vary by a large amount depending on the scale we used on the target. So, it is difficult to know if that RSS value is good or not. So, we need a scale-invariant statistic which is nothing but R-squared. R-squared value always lies between 0 and 1. Closer the value to 1, better is the result which is very easy to interpret.

****
****

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:

   TSS: Total variation in target variable is the sum of squares of the difference between the actual values and their mean. TSS or Total sum of squares gives the total variation in Y. We can see that it is very similar to the variance of Y. While the variance is the average of the squared sums of difference between actual values and data points, TSS is the total of the squared sums.

   ESS: 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 dependent variable.

   RSS: The residual sum of squares (RSS) is the sum of the squares of residuals (difference between estimated values and actual values)

   Formula: total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).

****
****

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

   Answer: We need regularization techniques in machine learning to reduce overfitting in our model. Ridge (L2) regularization modifies over-fitted and under-fitted models by adding penalty equivalent to the sum of the squares of the magnitude of the coefficients. Lasso (L1) regularization modifies over-fitted and under-fitted models by adding penalty equivalent to the sum of absolute values of coefficients.

****
****

4. What is Gini–impurity index?

   Answer: 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.

****
****

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

   Answer: 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. It could perform well for training data but would perform very bad for unseen data.

****
****

6. What is an ensemble technique in machine learning?

   Answer: Ensemble Techniques means combining different models. We will build multiple models and combine and make decisions. An ensemble technique in machine learning helps to make a better decision. Rather than just relying on one decision tree and hoping we made the right decision at each split, ensemble methods enables us to take a sample of decision trees into account, calculate which features to use or questions to ask at each split, and make a final predictor based on the aggregated results of the sampled decision trees.

****
****

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

Answer:

Bagging

1) Training data subsets are drawn randomly with replacement from the entire training data set

2) Bagging attempts to tackle the over fitting issue

3) Every model receives an equal weight

4) Objective to decrease variance, not bias

5) Every Model is built independently

Boosting

1) Each new subset contains the components that were misclassified by previous n models

2) Boosting tries to reduce bias

3) Models are weighted by their performance

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?

Answer: Out-of-bag (OOB) error, also called 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). Bagging uses subsampling with replacement to create training samples for the model to learn from. OOB error is the mean prediction error on each training sample xi, using only the trees that did not have xi in their bootstrap sample

****
****

9. What is K-fold cross-validation?

Answer: Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model. It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

****
****

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

Answer: Hyperparameter tuning is an essential part of controlling the behaviour of a machine learning model. If we do not correctly tune our hyperparameters, our estimated model parameters produce suboptimal results, as they do not minimize the loss function. This means our model makes more errors. In practice, key indicators like the accuracy or the confusion matrix will be worse.

****
****

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

Answer: The learning rate controls how quickly the model is adapted to the problem. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution.

****
****

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

Answer: Logistic Regression is a statistical approach and a Machine Learning algorithm that is used for classification problems and is based on the concept of probability. It cannot be used for classification of non-linear data. Logistic regression is considered a generalized linear model because the outcome always depends on the sum of the inputs and parameters. Or in other words, the output cannot depend on the product (or quotient, etc.).

****
****

13. Differentiate between Adaboost and Gradient Boosting.

Answer:

Adaboost:

1) In case of Adaptive Boosting or AdaBoost, it minimises the exponential loss function that can make the algorithm sensitive to the outliers.
2) AdaBoost is the first designed boosting algorithm with a particular loss function.
3) AdaBoost minimises loss function related to any classification error and is best used with weak learners. The method was mainly designed for binary classification problems and can be utilised to boost the performance of decision trees
4) AdaBoost is the first Boosting ensemble model. The method automatically adjusts its parameters to the data based on the actual performance in the current iteration

Gradient Boosting

1) With Gradient Boosting, any differentiable loss function can be utilised. Gradient Boosting algorithm is more robust to outliers than AdaBoost.
2) 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.

3) Gradient Boosting is used to solve the differentiable loss function problem. The technique can be used for both classification and regression problems.
4) 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.

****
****

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

Answer: In statistics and machine learning, the bias–variance trade-off 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.

Bias and variance are complements of each other" The increase of one will result in the decrease of the other and vice versa. Hence, finding the right balance of values is known as the Bias-Variance Tradeoff. Target Function. An ideal algorithm should neither underfit nor overfit the data.

****
****

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

Answer:

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 many features in a particular Data Set.

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

Polynomial kernel: 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.

****
****