MACHINE LEARNING

Q 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.- R-Squared is a goodness of fit model in regression because it indicates the percentage of the variable that the independent variables explain collectively.

Q 2.- What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and the RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.

Ans.- TSS (Total Sum of Squares) :- TSS is a variation of the values of a dependent variable from the sample mean of the dependent variable.

ESS (Explained Sum of Squares) :- ESS is the sum of the difference between the predicted value and the mean of the dependent variable.

RSS (Residual Sum of Squares) :- RSS is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model itself.

$$TSS = ESS + RSS$$

$$\sum(Yi - \bar{Y})^2 = \sum(Yi - \hat{Y}i)^2 + \sum(\hat{Y}i - \bar{Y})^2 + 2\sum(Yi - \hat{Y}i)(\hat{Y}i - \bar{Y})$$

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

Ans.- Regularization is used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting.

Q 4.- What is Gini-impurity index?

Ans.- Gini-impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset.

Mathematically, we can write Gini impurity –

$$I_{Gini} = 1 - \sum_{i=1}^{j} p_i^2$$

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

Ans.- Yes. 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.

Q 6.- What is an ensemble technique in machine learning?

Ans.- Ensemble technique is a machine learning paradigm where multiple models (weak learners) are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtained more accurate and / or robust models.

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

Ans.- The Bagging Technique combines multiple models trained on different subsets of data, whereas The Boosting Technique trains the model sequentially, focusing on the error made by the previous model.

The Bagging Technique reduces variance by averaging out individual model error, whereas The Boosting Technique reduces both bias and variance correcting misclassification of the previous model.

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

Ans.- Out -of-bag error is a method measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating. Out-of-bag error is the mean prediction error on each training sample x_i, using only the trees that did not have x_i in their bootstrap sample.

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

Ans.- K-fold cross-validation is a technique for evaluating predictive models. The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance.

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

Ans.- Hyperparameter tuning is the process of selecting the optimal values for a machine learning model's hyperparameters. Hyperparameter tuning is used to find the values that lead to the best performance on a given task.

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

Ans.- The learning rate is an important hyperparameter that greatly affects the performance of gradient descent. If we have a large learning rate, gradient descent can suffer from divergence. This means that weights increase exponentially, resulting in exploding gradients which can cause problems such as instabilities and overly high loss values.

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

Ans.- No, we cannot use Logistic Regression for classification of Non-Linear-Data because it has a linear decision surface.

Q 13.- Differentiate between Adaboost and Gradient Boosting.

Ans.- Adaboost -

1. In Adaboost, shift is done by up-weighting observations that were misclassified before.
2. In Adaboost "shortcomings" are identified by high-weight data points.
3. Exponentials loss of Adaboost gives more weights for those samples fitted worse.

Gradient Boosting –

1. Gradient Boost identifies difficult observations by large residuals computed in the previous iterations.
2. In Gradient Boost "shortcomings" are identified by gradients.
3. Gradient boost further dissect error components to bring in more explanation.

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

Ans.- Bias-Variance trade off is a fundamental concept in machine learning that deals with the balance between model bias and variance. It refers to the tradeoff between a model's ability to accurately represent the underlying data patterns (low bias) and its susceptibility to fluctuations with changes in the training data (high variance).

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

Ans.- Linear Kernel :- A linear kernel can be used as normal dot product any two given observations. The product between two vectors is the sum of multiplication of each pair of input values.

$$K(x, xi) = sum(x*xi)$$

Radial Basis Function(RBF) Kernel :- The Radial Basis Function Kernel is a popular kernel function commonly used in support vector machine classification. RBF can map an input space in infinite dimensional space.

$$K(x, xi) = exp(-gamma*sum (x – xi^2))$$

Polynomial Kernel :- A polynomial kernel is a more generalized form of the linear kernel. The polynomial kernel can distinguish curved or nonlinear input space.

$$K(x, xi) = 1 + sum(x*xi)^d$$