MACHINE LEARNING ASSIGNMENT-5

Q1 to Q15 SUBJECTIVE ANSWERS

Q1. Both R-squared or Residual Sum of Squares (RSS) is a better measure of goodness of fit model in regression because 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.

Q2. TSS is the sum of the squared deviations of the predicted values from the observed mean of y, ESS is the explained sum of squares, and RSS is the residual sum of squares. EQUATION: $TSS = ESS + RSS + 2\sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$

Q3. Regularization is a technique to prevent the model from overfitting by adding extra information to it.

Q4. Gini-impurity or gini-index in machine learning is a metric to measure the randomness in a feature.

Q5. Unregularized decision-trees are prone to overfitting because they can learn a training set to a point of high granularity that makes them easily overfit.

Q6. Ensemble learning is a technique in machine learning which takes the help of several base models and combines their output to produce an optimized model.

Q7. Bagging gives equal weight to each model, whereas in Boosting technique, the new models are weighted based on their results.

Q8. The out-of-bag error is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample.

Q9. K-fold cross-validation is a method for estimating the performance of a model on unseen data.

Q10. Hyperparameters are the knobs or settings that can be tuned before running a training job to control the behavior of an ML algorithm because it needs to tune hyperparameters in machine learning to minimize the error.

Q11. If a large learning rate in Gradient Descent occur as seen as step size, η. As such, gradient descent is taking successive steps in the direction of the minimum. If the step size η is too large, it can (plausibly) "jump over" the minima trying to reach, i.e., it overshoot that. This can lead to osculations around the minimum or in some cases to outright divergence.

Q12. Yes, it use Logistic Regression for classification of Non-Linear Data. It is a statistical analysis method used to determine relationships between continuous variables and categorical variables.

Q13. AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem.

Q14. 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.

Q15. SHORT DESCRIPTION:

Linear Kernel : They are used when data is linearly separable.

Radial Basis Function[RBF] Kernel: It uses trick actually refers to using efficient and less expensive ways to transform data into higher dimensions.

Polynomial Kernel: It is a kernel function commonly used with support vector machines (SVMs) and other kernelized models.

THANK YOU