WORKSHEET 5

MACHINE LEARNING

Q1 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 (SS_{res}) with the total sum of squares (SS_{tot}). The total sum of squares is calculated by summation of squares of perpendicular distance between data points and the average line.

Q2 answer :

TSS : The sum of squares total, denoted SST, is the squared differences between the observed dependent variable and its mean. It is a measure of the total variability of the dataset.

ESS : sum of squares error, or SSE. The error is the difference between the observed value and the predicted value.

RSS : sum of squares due to regression, or **SSR**. It is the sum of the differences between the predicted value and the mean of the dependent variable.

Reletion is denoted as

TSS=RSS+ESS

Q3 answer : Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting**.** Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

WORKSHEET 5

Q4 answer : The Gini impurity measure is one of the methods used in decision tree algorithms to decide the optimal split from a root node, and subsequent splits.it is the most popular and the easiest way to split a decision tree and it works only with categorical targets as it only does binary splits.

Q5 answer : yes,

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 that is overfitting.

Q6 answer : Ensemble Techniques learning is a technique in machine learning which takes the help of several base models and combines their output to produce an optimized model. This type of machine learning algorithm helps in improving the overall performance of the model.

Q7 answer :

1)Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions.

2)In Bagging, each model receives an equal weight. In Boosting, models are weighed based on their performance.

3)Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.

4) Models are built independently in Bagging. New models are affected by a previously built model's performance in Boosting.

5) In Bagging, training data subsets are drawn randomly with a replacement for the training dataset. In Boosting, every new subset comprises the elements that were misclassified by previous models

WORKSHEET 5

Q8 answer : OOB score is a very powerful Validation technique used espiecilly for the Random Forest algorithm for least variance results.

Q9 answer : 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.

Q10 answer : Hyperparameters are the knobs or settings that can be tuned before running a training job to control the behavior of an ML algorithm. They can have a big impact on model training as it relates to training time, infrastructure resource requirements (and as a result cost), model convergence and model accuracy.

Q11 answer : This defeats the purpose of gradient descent, which was to use a computationally efficient method for finding the optimal solution. On the other hand, if you choose a learning rate that is too large, you might overshoot the minimum value of the error function, and may even never reach the optimal solution.

Q12 answer : Logistic Regression is not suitable for complex non-linear relationships between the dependent variable and independent variables. it can only handle binary classification.

Q13 answer : The technique of Boosting uses various loss functions. In case of Adaptive Boosting or AdaBoost, it minimises the exponential loss function that can make the algorithm sensitive to the outliers. With Gradient Boosting, any differentiable loss function can be utilised. Gradient Boosting algorithm is more robust to outliers than AdaBoost.

Q14 answer : 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.

WORKSHEET 5

Q15 answer : 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 a Large number of Features in a particular Data Set.

This example illustrates the effect of the parameters gamma and C of the Radial Basis Function (RBF) kernel SVM. Intuitively, the gamma parameter defines how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close'.

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.

WORKSHEET 5

STATICS WORKSHEET

1 c) Predicted

2 c) Frequencies

3 b) 6

4 b) Chisqared distribution

5 c) F Distribution

6 b) Hypothesis

7 a) Null Hypothesis

8 a) Two tailed

9 b) Research Hypothesis

10 a) np