

MACHINE LEARNING ASSIGNMENT -5

ANSWERS:

1. R-Squared is a better measure of goodness of fit model as RSS is just the absolute amount of explained variation, whereas the R Squared is the (RSS/TSS) absolute amount of variation as a proportion of total variation.
2. ESS or Explained Sum of Squares marks the variation in data explained by the regression model. RSS or Residual Sum of Squares marks the variations marked by discrepancies in the dataset not explained by the model. The summation of both ESS and RSS is TSS or Total Sum of Squares. Equation: **$TSS = RSS + ESS$**
3. 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.
4. Gini impurity calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. If all the elements are linked with a single class, then it can be called pure.
5. Yes, unregularized decision-trees prone to overfitting. In decision trees, over-fitting occurs when the tree is designed so as to perfectly fit all samples in the training data set. Thus, it ends up with branches with strict rules of sparse data. So this effects the accuracy when predicting samples that are not part of the training set.
6. Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. Popular Ensemble techniques: Bagging, Boosting.
7. Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.
8. The out-of-bag (OOB) error is the average error for each training observation (i) calculated using predictions from the trees that do not contain (i) in their respective bootstrap sample. This allows the Random Forest Classifier to be fit and validated whilst being trained.
9. 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.
10. 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.

11. 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. Logistic Regression has traditionally been used as a linear classifier, i.e., when the classes can be separated in the feature space by linear boundaries but it can also be used as a non-linear classifier.
13. 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. This makes Gradient Boosting more flexible than AdaBoost.
14. 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.
15. Kernels in SVM:
 - 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.
 - The RBF kernel works by mapping the data into a high-dimensional space by finding the dot products and squares of all the features in the dataset and then performing the classification using the basic idea of Linear SVM.
 - A polynomial kernel is a kind of SVM kernel that uses a polynomial function to map the data into a higher-dimensional space. It does this by taking the dot product of the data points in the original space and the polynomial function in the new space.