

MACHINE LEARNING ASSIGNMENT

ANSWERS:

1. The RSS is just the absolute amount of explained variation, the R squared is the (RSS/SST) , i.e. the absolute amount of variation as a proportion of total variation.

RSS alone doesn't explain anything so R-squared would be the appropriate measure to check the goodness of the fit of the model.

2. **TSS:** The total sum of squares, denoted TSS, is the squared differences between the observed dependent variable and its mean.

ESS: The second term is the explained sum of squares, or ESS. It is the sum of the differences between the predicted value and the mean of the dependent variable.

RSS: The last term is residual sum of squares, or RSS. It is the error is the difference between the observed value and the predicted value.

$$\text{TSS} = \text{ESS} + \text{RSS}$$

3. Regularization is an important concept in machine learning because it is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid over-fitting.
4. Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. If all the elements belong to a single class, then it can be called pure. The degree of Gini index varies between 0 and 1, where, 0 denotes that all elements belong to a certain class or if there exists only one class, and 1 denotes that the elements are randomly distributed across various classes. A Gini Index of 0.5 denotes equally distributed elements into some classes.
5. 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. Thus 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.
7. **Bagging:** It is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.

Boosting: It is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm.

8. In Breiman's original implementation of the random forest algorithm, each tree is trained on about 2/3 of the total training data. As the forest is built, each tree can thus be tested (similar to leave one out cross validation) on the samples not used in building that tree. This is the out of bag error estimate - an internal error estimate of a random forest as it is being constructed.
9. K-Fold CV is where a given data set is split into a K number of sections/folds where each fold is used as a testing set at some point. Lets take the scenario of 5-Fold cross validation(K=5). Here, the data set is split into 5 folds. In the first iteration, the first fold is used to test the model and the rest are used to train the model. In the second iteration, 2nd fold is used as the testing set while the rest serve as the training set. This process is repeated until each fold of the 5 folds have been used as the testing set.
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. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.
12. No. Non-linear problems can't be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.
13. Adaboost is an additive model where shortcomings of previous models are identified by high-weight data points whereas gradient boosting is an additive model where shortcomings of previous models are identified by the gradient. In Adaboost each classifier has different weights assigned to the final prediction based on its performance whereas in gradient boosting all classifiers are weighed equally and their predictive capacity is restricted with learning rate to increase accuracy.
14. In 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 and vice-versa.
15. **Linear Kernel:** It is used when the data is linearly separable, that is, it can be separated using a single Line.

RBF Kernel: It is used in SVM because it helps the SVM to become non-linear rather than linear. RBF kernel function is similar to Normal distribution.

Polynomial Kernel: It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.