Worksheet 5 - ML

Submitted by Arjun B

1) A smaller or lower value for the RSS is ideal in any model since it means there's less variation in the data set. In other words, the lower the sum of squared residuals, the better the regression model is at explaining the data.
2) TSS-The sum of squares total, denoted SST, is the squared differences between the observed dependent variable and its mean. You can think of this as the dispersion of the observed variables around the mean. The total sum of squares (TSS) measures how much variation there is in the observed data, while the residual sum of squares measures the variation in the error between the observed data and modeled values.
   total sum of squares (TSS) = explained sum of squares (ESS) + residual sum of squares (RSS).
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 Index, also known as the Gini impurity, calculates the amount 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) If a decision tree is fully grown, it may lose some generalization capability. This is a phenomenon known as overfitting.
6) Ensemble methods are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.
7) Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.
8) The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained
9) K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-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.

Worksheet 5 - ML

Submitted by Arjun B

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) Logistic regression is indeed nonlinear in terms of Odds and Probability, however, it is linear in terms of Log Odds.

$$\textbf{Probability of } (Y=1): p = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2}}$$

$$\textbf{Odds of } (Y=1): \left(\frac{p}{1-p}\right) = e^{\alpha+\beta_1 x_1+\beta_2 x_2}$$

$$\textbf{Log Odds of } (Y=1): \log\left(\frac{p}{1-p}\right) = \alpha+\beta_1 x_1+\beta_2 x_2$$

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) In statistics and machine learning, 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.

15) Linear kernel: The Linear kernel is the simplest kernel used in SVM. It simply calculates the dot product between two vectors and is used when the data is linearly separable.

Radial basis function (RBF) kernel: RBF is a non-linear kernel that maps the input data into a higher-dimensional space to make it linearly separable. RBF is defined by a Gaussian function that centers around a landmark data point and decreases as the distance from the landmark increases.

Polynomial kernel: The Polynomial kernel is used to model the relationship between variables when the data is not linearly separable. The kernel maps the input data into a higher-dimensional space and models the relationship as a polynomial equation of degree "d".