WORKSHEET – 5

STATISTICS

1.d] Expected

2.c] Frequencies

3.c] 6

4.b] Chi-squared distribution

5.c] F distribution

6.b] Hypothesis

7.a] Null Hypothesis

8.a] Two Tailed

9.b] Research Hypothesis

10.a] np

MACHINE LEARNING

3. One of the major aspect of training your machine learning model is avoiding overfitting. The model will have a low accuracy if it is overfitting. This happens because your model is trying too hard to capture the noise in your training dataset. This is a form of regression, that constrains / regularizes or shrinks the coefficient estimates toward zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

4. Gini Index, also known as Gini impurity, calculates the amount of probability of a specifics features that is classified incorrectly when selected randomly. If all the elements are linked with a single class then it can be called pure. The Gini index varies between values 0 and 1, where 0 express the purity of classification, i.e. all the elements belong to a specified class or only one class exists there.

And 1 indicates the random distribution of elements across various classes. The value of 0.5 of the Gini Index shows an equal distribution of elements over some classes.

5. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusion.

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. The most popular ensemble methods are **' Boosting , Bagging & Stacking '**. Ensemble methods are ideal for regression and classification, where they reduce bias and variance to boost the accuracy of models.

7. Bagging and **Boosting** are two types of **Ensemble Learning**. These two decrease the variance of a single estimate as they combine several estimates from different models. So the result may be a model with higher stability. Let's understand these two terms in glimpse.

- **Bagging** : Bootstrap Aggregating, also known as bagging, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It decreases the **variance** and helps to avoid **overfitting**. It is usually applied to **decision tree methods**. Bagging is a special case of the model averaging approach.
- **Boosting** : Boosting is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is built from the training data. Then

the second model is built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models is added.

8. Out-of-bag (OOB), also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging). Bagging uses subsampling with replacement to create training samples for the model to learn from.

9. Cross validation is an evaluation method used in machine learning to find out how well your machine learning model can predict the outcome of unseen data. It is a method that is easy to comprehend, works well for a limited data sample and also offers an evaluation that is less biased, making it a popular choice. The data sample is split into 'k' number of smaller samples, hence the name: k-fold Cross Validation. You may also hear terms like four fold cross validation, or ten fold cross validation, which essentially means that the sample data is being split into four or ten smaller samples respectively.

10. Hyperparameter tuning consist of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameter maximizes the model's performances, minimizing a predefined loss function to produces better results with fewer errors.

11. In order for Gradients 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. If it is too small we will need too many iterations to converge to the best values. So using a good learning rate is crucial.

12. Logistic Regression has traditionally been used as a linear classifier. Logistic Regression is known and used as a linear classifiers. It is used to come up with a hyperplane in feature space to separate observations that belong to a class from all the other observation that do not belong to that class.

13. Both AdaBoosting and Gradient Boosting build weak learner in a sequential fashion. Originally, AdaBoost was designed in such a way that at a every step the sample distribution was adapted to put more weight on misclassified samples. The final prediction is a weighted average of all the weak learners, where more weight is placed on stronger learners. The main difference , therefore are that Gradient Boosting is a generic algorithm to find approximate solution to the additive modelling problem, while AdaBoost can be seen as a special case with a particular loss function. Hence, Gradient Boosting is much more flexible. On the other hand AdaBoosting can be interpreted from much more intuitive perspective and can be implemented without the reference to gradient by reweighting the training samples based on classifications from previous learners.

14. In statistics and 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. The bias-variance problem is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised learning algorithms from generalizing beyond their training set.

15. A Kernel is a function used in SVM for helping to solve problems. With the help of the kernel, we can go to higher dimensions and perform smooth calculations. We can go up to an infinite number of dimensions using kernels. Kernels plays a vital role in classifying and analyzing some given dataset patterns. They are very helpful in solving a non-linear problem by using a linear classifier.

- **Linear Kernel:** It is the most basic kernel type, usually one-dimensional in nature. It proves to be the best function when there are lots of features. Linear kernels functions are faster than other functions.
- **Polynomial Kernel:** It is a more generalized representation of the linear kernel. It is not as preferred as other kernel functions as it is less efficient and accurate.
- **RBF:** It is one of the most preferred and used kernel functions in SVM. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data.