

# WORKSHEET-5

---

## STATISTICS

- 1 c) Predicted
- 2 c) Frequencies
- 3 b) 12
- 4 b) Chisquared 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

1 Typically, however, 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

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 More precisely, the Gini Impurity of a dataset is **a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset**.

**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 conclusions.

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**: 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 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.

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. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary...

13 Just like AdaBoost, Gradient Boost also **combines a no. of weak learners to form a strong learner**. Here, the residual of the current classifier becomes the input for the next consecutive classifier on which the trees are built, and hence it is an additive mode

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 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.