**Machine Learning Assignment - 5**

1. **R-squared or Residual Sum of Squares (RSS) Which one of these two is a better measure of the goodness of fit model in regression and why?**

**Ans** **=** R-squared is generally considered a better overall measure of goodness of fit for the entire model. It gives a sense of the proportion of variance explained, providing a global assessment.

RSS is important for understanding prediction errors on an individual level.

R-squared is preferred because it gives a percentage that tells you how well your model explains the variability in the data, making it easier to understand and compare different models.

1. **What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.**

**Ans = Total sum of squares** - TSS measures the total variability in the dependent variable (the variable we trying to predict) without considering the effect of the independent variables we using to make predictions.

**Explained Sum of Squares** – ESS measures the variability in the dependent variable that is explained by your regression model, it represents how well your model fits the data.

**Residual Sum of Squares** – RSS measures the Variability in the dependent variable that is not explained by your regression model

**TSS = ESS + RSS**

1. **What is the need of regularization in machine learning?**

**Ans** = Regularization consists of different techniques and methods used to address the issue of over-fitting by reducing the generalization error without affecting the training error much. Imagine training the model is like fitting a suit. You want is to fit well, not too loose or too tight.

1. **What is Gini–impurity index?**

**Ans =** The Gini impurity index is a tool used by decision tree algorithmsto guide the process of sorting data into groups in a way that minimizes the chance of misclassifying elements. It helps build effective decision tree by reducing the disorder within each group.

1. **Are unregularized decision-trees prone to overfitting? If yes, why?**

**Ans =** Yes, unregularized decision trees are prone to overfitting. Overfitting occurs when a model learns the training data too well, capturing noise and specific patterns that might not generalized well to new, unseen data. Decision trees, by nature, are capable of fitting the training data very closely, and this can lead to overfitting for several reasons.

1. **What is an ensemble technique in machine learning?**

**Ans =** Imagine you have a big decision to make, like where to go for dinner. Instead of relying on just one person’s opinion, you might ask a bunch of your friends for their recommendations and then choose the restaurant that gets the most votes.

Ensemble techniques in machine learning work kind of like that. Instead of relying on just one model to make predictions, you use multiple models and then combine their predictions to come up with a final answer.

1. **What is the difference between Bagging and Boosting techniques?**

**Ans =** Bagging and Boosting are both ensemble techniques used in machine learning.

**Bagging (Bootstrap Aggregation):**

* Bagging involves training multiple models independently, each on a random subset of the training data.
* The simplest way of combining predictions that belong to the same type.
* Aim to decrease variance, not bias.
* Each model receives equal weight.
* If the classifier is unstable (high variance), then apply bagging.

**Boosting:**

* Boosting, on the other hand, focuses on training multiple models sequentially, where each subsequent model tries to correct the errors of the previous one.
* A way of combining predictions that belong to the different types.
* Aim to decrease bias, not variance.
* Models are weighted according to their performance.
* New models are influenced by the performance of previously built models.

1. **What is out-of-bag error in random forests?**

**Ans =** In random forests, each tree in the forest is trained using a random subset of the original data. But here’s the cool part: not every data is used to train every tree.

When we’re randomly selecting data for each tree, some data points might not get picked. These untouched data points are called “out-of-bag” (OOB)samples.

So, the out-of-bag is like a bonus After training each tree we can use the untouched OOB samples to test how well that tree predicts data it hasn’t seen before. Then we do this for every tree in the forest.

1. **What is K-fold cross-validation?**

**Ans =** K-fold cross-validation is one of the most popular strategies widely used by data scientists. It is a data partitioning strategy so that you can effectively use your dataset to build a more generalized model. The main intention of doing any kind of machine learning is to develop a more generalized model which can perform well on unseen data. One can build a perfect model on the training data with 100% accuracy or 0 error, but it may fail t generalize for unseen data.

1. **What is hyper parameter tuning in machine learning and why it is done?**

**Ans =** In machine learning, algorithms have settings like this too, called hyperparameters. Hyperparameter tuning is like finding the best settings for these parameters to make your model perform its best. Different hyperparameters can affect how well your model learns from data. Tuning them can make your model more accurate and reliable.

So, just like adjusting the sugar in your cookies to make them perfect, hyperparameter tuning helps adjust your machine learning model to make it perform its best on your data.

1. **What issues can occur if we have a large learning rate in Gradient Descent?**

**Ans =** If your steps are too big, you might leap past the bottom of the valley you’re trying to reach. Similarly, in gradient descent, a large learning rate can cause the algorithm to overshoot the minimum point of the loss function. It might keep bouncing back and forth around the minimum without ever settling down. When you take huge steps, you might miss small rocks or obstacles on the way down. Similarly, a large learning rate might cause the algorithm to skip over important details in the data, leading to less accurate models.

1. **Can we use Logistic Regression for classification of Non-Linear Data? If not, why?**

**Ans =** Logistic Regression is a powerful tool for classification, but it works best when the relationship between the input feature and the output (or the target variable) is linear. This means that the decision that the decision boundary separating different classes is a straight line(in 2D) or a hyperplane (in higher dimensions).

So, while Logistic Regression is a fantastic choice for linearly separable data, for nonlinear data, it’s better to use other algorithms that can handle the complexity of the relationships between features and target variables.

1. **Differentiate between Adaboost and Gradient Boosting.**

**Ans = Adaboost (Adaptive Boosting):**

* Adaboost works by combining multiple weak learners (simple models that perform slightly better than random guessing) to create a strong learner (A powerful model).
* It does this by giving more weight to be data points that were misclassified by the previous weak learners.

**Gradient Boosting:**

* Gradient Boosting also combines multiple weak learners, but it does so by building them sequentially.
* It starts by creating a simple model (like a decision tree) that tries to predict the target variable.

1. **What is bias-variance trade off in machine learning?**

**Ans = Bias** is like using a recipe that’s too simple. No matter how many times you bake it, the cake never testes quite right. In machine learning, bias refers to the error introduced by overly simplistic assumptions in the model.

**Variance** is like using a recipe that’s too complicated. Sometimes the cake turns out perfectly, but other times it’s a disaster because you made too many adjustments. In machine learning, variance refers of the error introduced by the model being too sensitive to small fluctuations in the training data.

**Trade-off** when we are trying to find the perfect balance between using a simple recipe and a complicated one. If you stick too much to the simple recipe, your cake might always taste mediocre because you’re missing out on some key ingredients.

1. **Give short description each of Linear, RBF, Polynomial kernels used in SVM**

**Ans = Linear Kernel:**

* The linear kernel is the simplest one.
* It works like drawing a straight line to separate different classes in your data.
* Imagine you have two groups of points, and you can draw a straight line between them to separate them. That’s what the linear kernel does it finds the best straight line to divide the data

**RBF (Radial Basis Function) Kernel:**

* The RBF kernel is more flexible than the linear one.
* It works by creating a ‘bump’ around each data point, making it like hills and valleys.

**Polynomial Kernel:**

* The Polynomial kernel is also flexible but in a different way than the RBF.
* It works by creating curved boundaries instead of straight line.
* Creates curved boundaries to capture nonlinear relationships in the data.

**Statistics Worksheet - 5**

1. **Using a goodness of fit,we can assess whether a set of obtained frequencies differ from a set of frequencies.**

**Ans = c)** Predicted

1. **Chisquare is used to analyse**

**Ans = c)** Frequencies

1. **What is the mean of a Chi Square distribution with 6 degrees of freedom?**

**Ans = c)** 6

1. **Which of these distributions is used for a goodness of fit testing?**

**Ans = b)** Chisqared distribution

1. **Which of the following distributions is Continuous**

Ans = F distribution

1. **A statement made about a population for testing purpose is called?**

**Ans =** **b)** Hypothesis

1. **If the assumed hypothesis is tested for rejection considering it to be true is called?**

**Ans = a)** Null Hypthesis

1. **If the Critical region is evenly distributed then the test is referred as?**

**Ans = b)** One tailed

1. **Alternative Hypothesis is also called as?**

**Ans =** Simple Hypothesis

1. In a Binomial Distribution, if ‘n’ is the number of trials and ‘p’ is the probability of success, then the mean value is given by \_\_\_\_\_\_\_\_\_\_\_

**Ans = a)** np