MACHINE LEARNING – WORKSHEET 5

1. **R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?**
   Ans. R-squared is the absolute amount of variation as a proportion of total variation. The residual sum of squares (RSS) is the absolute amount of explained variation

   The residual sum of squares (RSS) measures the level of variance in the error term, or residuals, of a regression model. The smaller the residual sum of squares, the better your model fits your data; the greater the residual sum of squares, the poorer your model fits your data. RSS=Σi=1 to n [(yi – y'i)^2] Where yi is a given data point and y'i is the fitted value for yi. The actual number we get depends largely on the scale of our response variable. Taken alone, the RSS isn't so informative. Therefore, R2 is a better measure.

2. **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. TSS: The total sum of squares (TSS) measures how much variation there is in the observed data.
   TSS = Σi=1 to n [(y'i – ymean)^2] + Σi=1 to n [(yi – y'i)^2]

   ESS- The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model.
   ESS=Σi=1 to n [(y'i – ymean)^2]

   RSS-The residual sum of squares (RSS) is the sum of the squared distances between actual versus predicted values.
   RSS=Σi=1 to n [(yi – y'i)^2]

The relation between the above three can be linearly expressed as: TSS = RSS + ESS .

3. **What is the need of regularization in machine learning?**
   Ans. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting .
   Need of Regularization-
   Regularization constraints or shrinks the coefficient towards zero. This means that this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting. Regularization significantly reduces the variance of the model, without substantial increase in its bias.

4. **What is Gini–impurity index?**
   Ans. Gini index or Gini impurity measures the probability of a particular variable to be wrongly classified when chosen randomly. This measure is calculated where the modelling contains Tree Algorithms like Decision Tress or random forest.

Gini impurity is calculated by subtracting the sum of the squared probabilities of each class
from one.

$$Gini = 1 - \sum_{i=1}^{n} (p_i)^2$$

5. **Are unregularized decision-trees prone to overfitting? If yes, why?**
   Ans. Yes, unregularized decision trees are prone to overfitting. 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. But unlike other algorithms decision tree does not use regularization to fight against overfitting. Instead, it uses pruning. There are mainly to types of pruning performed: -
   Pre-pruning that stops growing the tree earlier, before it perfectly classifies the training set.
   Post-pruning that allows the tree to perfectly classify the training set, and then post prune the tree.

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

Ans. - An ensemble method is a technique which uses multiple independent similar or different models/weak learners to derive an output or make some predictions.
Two of the most used techniques in machine learning.
Bagging and Boosting

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

Ans. Bagging is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.
Boosting 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.

Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types.

Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.

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

Ans. Out-of-bag error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests. 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. **What is K-fold cross-validation?**
   Ans. 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. Cross-validation is usually used in machine learning for improving model prediction when we don't have enough data to apply other more efficient methods like the 3-way split (train, validation and test) or using a holdout dataset**.**

10. **What is hyper parameter tuning in machine learning and why it is done?**
    Ans. In machine learning, hyper parameter optimization or tuning is the problem of choosing a set of optimal hyper parameters for a learning algorithm. A hyper parameter is a parameter whose value is used to control the learning process.

11. **What issues can occur if we have a large learning rate in Gradient Descent?**
    Ans. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution.

12. **Can we use Logistic Regression for classification of Non-Linear Data? If not, why?**
    Ans. No.
    Non-linear problems can't be solved with logistic regression because it has a linear decision surface.
    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.

13. **Differentiate between Ada boost and Gradient Boosting.**
    Ans.Gradient Boosting is a generic algorithm to find approximate solutions to the additive modelling problem, while AdaBoost is a special case with a particular loss function. Hence, Gradient Boosting is much more flexible. On the other hand, AdaBoost can be interpreted from a much more intuitive perspective and can be implemented without the reference to gradients by reweighting the training samples based on classifications from previous learners.
    Gradient Boosting algorithm is more robust to outliers than AdaBoost.
    Gradient Boosting is more flexible than AdaBoost.
    In the case of Gradient Boosting, the shortcomings of the existing weak learners can be identified by gradients and with AdaBoost, it can be identified by high-weight data points.

14. **What is bias-variance trade off in machine learning?**
    Ans. In 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.
    If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand, if our model has large number of parameters then it's going to have high variance and low bias. So, we need to find the right/good balance without overfitting and under fitting the data. This trade off in complexity is why there is a trade-off between bias and variance. An algorithm can't be more complex and less complex at the same time.

15. **Give short description each of Linear, RBF, Polynomial kernels used in SVM.**
    Ans.
    - 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 many Features in a particular Data Set.

- RBF (Radial Basis Function) is another popular Kernel method used in SVM models. RBF kernel is a function whose value depends on the distance from the origin or from some point.
- In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel.