**Worksheet\_5**

MACHINE LEARNING

**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:** R2: it represents the proportion of the variance in our data which is explained by our model; the closer to one, the better the fit. The residual sum of squares (RSS) is the sum of the squared distances between actual versus predicted values:

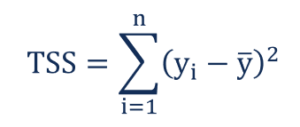
RSS=∑i=1 to n [(yi – y’i)^2]

Where yi is a given datapoint 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:** The total sum of squares is a variation of the values of a [dependent variable](https://corporatefinanceinstitute.com/resources/knowledge/terms/dependent-variable/) from the sample mean of the dependent variable. Essentially, the total sum of squares quantifies the total variation in a [sample](http://www.webmath.com/sampledata.html). It can be determined using the following formula:

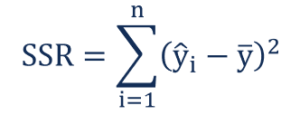


Where:

* yi– the value in a sample
* ȳ – the mean value of a sample

ESS (Explained Sum of Squares) describes how well a regression model represents the modeled data. A higher regression sum of squares indicates that the model does not fit the data well.

The formula for calculating the ESS is:

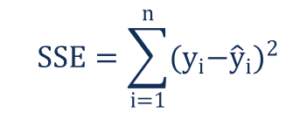


Where:

* ŷi– the value estimated by the regression line
* ȳ – the mean value of a sample

The residual sum of squares essentially measures the variation of modeling errors. In other words, it depicts how the variation in the dependent variable in a regression model cannot be explained by the model. Generally, a lower residual sum of squares indicates that the regression model can better explain the data while a higher residual sum of squares indicates that the model poorly explains the data.

The residual sum of squares can be found using the formula below:



Where:

* yi– the observed value
* ŷi– the value estimated by the regression line

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

**ANS:**Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

**4.** What is Gini–impurity index?

**ANS**: If we have C total classes and p(i) is the probability of picking a datapoint with class i, then the Gini Impurity is calculated as G=i=1∑C p(i)∗(1−p(i))

Gini Index, also known as Gini impurity, calculates the amount of 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.** Are unregularized decision-trees prone to overfitting? If yes, why?

**ANS:** 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.** What is an ensemble technique in machine learning?

**ANS**: Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging), bias (boosting), or improve predictions (stacking).

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

**ANS:Bagging** is merging same type of predictions. It decreases variance, not bias, and solves over-fitting issues in a model

**Boosting** is merging diff type of predictions.It decreases bias, not variance.

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

**ANS:** Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests

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

**ANS**: K fold cross validation is a procedure to evaluate machine learning models. The cross validation function/ method has a parameter K which refers to the number of groups that the given sample needs to be split in.

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

ANS:Hyperparameter tuning is the process of selecting optimum values of the hypermeter for learning algorithm,

Hyperparameters is used because it directly control the behaviour of the training algorithm and have a significant impact on the performance of the model is being trained.

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

ANS: **When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error.**

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

ANS:Logistic regression is not optimal for classification of nonlinear data as it can perform best when data can be linearly seperated

13. Differentiate between Adaboost and Gradient Boosting.

|  |  |
| --- | --- |
| **AdaBoost** | **Gradient** |
| AdaBoost is the first designed boosting algorithm with a particular loss function | Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem |
| minimizes the exponential loss function that can make the algorithm sensitive to the outliers | any differentiable loss function can be utilized |
| less robust to outliers | more robust to outliers |
| AdaBoost minimizes loss function related to any classification error and is best used with weak learners | Gradient Boosting is used to solve the differentiable loss function problem |

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 estimates across samples can be reduced by increasing the bias in the estimated parameters.

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.

• Gaussian RBF(Radial Basis Function) is another popular Kernel method used in SVM models for more. 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