MACHINE LEARNING ASSIGNMENT - 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?

   I think the R-squared is a better measure of goodness of fit model in regression. This statistic indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. R-squared measures the strength of the relationship between the model and the dependent variable on a convenient 0 – 100% scale. The higher the R-squared, the better the model fits our data.

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.

   TSS- The total sum of squares (TSS) measures how much variation there is in the observed data

   RSS- The residual sum of squares measures the variation in the error between the observed data and modeled values

ESS- Explained sum of square (ESS) or Regression sum of squares or Model sum of squares is a statistical quantity used in modeling of a process. ESS gives an estimate of how well a model explains the observed data for the process. It tells how much of the variation between observed data and predicted data is being explained by the model proposed. Mathematically, it is the sum of the squares of the difference between the predicted data and mean data.

ESS = (TSS) total sum of squares – (RSS) residual sum of squares

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

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. What is Gini–impurity index?

Gini impurity is an important measure used to construct the decision trees. Gini impurity is a function that determines how well a decision tree was split. Basically, it helps us to

determine which splitter is best so that we can build a pure decision tree. Gini impurity ranges values from 0 to 0.5.

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

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?

An ensemble method is a technique which uses multiple independent similar or different models/weak learners to derive an output or make some predictions. Ensemble methods usually produces more accurate solutions than a single model would.

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

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. What is out-of-bag error in 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 Random Forest Classifier to be fit and validated whilst being trained.

9. What is K-fold cross-validation?

K-fold cross-validation is defined as a method for estimating the performance of a model on unseen data. This technique is recommended to be used when the data is

scarce and there is an ask to get a good estimate of training and generalization error thereby understanding the aspects such as underfitting and overfitting.

This technique is used for hyperparameter tuning such that the model with the most optimal value of hyperparameters can be trained. It is a resampling technique without replacement. The advantage of this approach is that each example is used for training and validation (as part of a test fold) exactly once. This yields a lower-variance estimate of the model performance than the holdout method.

As mentioned earlier, this technique is used because it helps to avoid overfitting, which can occur when a model is trained using all of the data. By using k-fold cross-validation, we are able to “test” the model on k different data sets, which helps to ensure that the model is generalizable.

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

In machine learning, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.

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

This is the hyperparameter that determines the steps the gradient descent algorithm takes. Gradient Descent is too sensitive to the learning rate. If it is too big, the algorithm may bypass the local minimum and overshoot and can cause the model to converge too quickly to a suboptimal solution and due to this it can inadvertently increase the training error.

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

Logistic regression is considered a generalized linear model because the outcome always depends on the sum of the inputs and parameters. Or in other words, the output cannot depend on the product (or quotient, etc.). Non-Linear Classification refers to categorizing those instances that are not linearly separable, i.e. It is not easy to classify data with a straight line.

13. Differentiate between Adaboost and Gradient Boosting.

Gradient boosting and Adaboost are ensemble methods applied in machine learning(ML) modeling to improve the effectiveness of weak learners to increase algorithm performance. Both use a combination of weak learners to predict a target variable. However, they do that differently.

While gradient boosting trains the learners and reduces the weak learners' loss functions by training the model's residues, Ada boost focuses on training the prior miscalculated observations and alters the data distribution to improve sample weight values.

And while boosting algorithms are less prone to overfitting, they can get complex and overfit the training data.

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

It helps optimize the error in our model and keeps it as low as possible. An optimized model will be sensitive to the patterns in our data, but at the same time will be able to generalize to new data. In this, both the bias and variance should be low so as to prevent overfitting and underfitting.

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

Linear Kernel

It is the most basic type of kernel, usually one dimensional in nature. It proves to be the best function when there are lots of features. The linear kernel is mostly preferred for text-classification problems as most of these kinds of classification problems can be linearly separated.

Linear kernel functions are faster than other functions.

Linear Kernel Formula

$$F(x, xj) = sum(x.xj)$$

Here, x, xj represents the data you're trying to classify.

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.

Polynomial Kernel Formula

$$F(x, xj) = (x.xj+1)^d$$

Here '.' shows the dot product of both the values, and d denotes the degree.

F(x, xj) representing the decision boundary to separate the given classes.

Gaussian Radial Basis Function (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.

Gaussian Radial Basis Formula

$$F(x, xj) = \exp(-\text{gamma} * ||x - xj||^2)$$

The value of gamma varies from 0 to 1. You have to manually provide the value of gamma in the code. The most preferred value for gamma is 0.1.