MACHINE LEARNING ASSIGNMENT -5

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

Ans. To determine which is a better measure of the goodness of fit model in regression between R-squared and Residual Sum of Squares (RSS). We need to consider both measures in the context of regression analysis. R-squared, or the coefficient of determination, indicates the proportion of variance in the dependent variable that is predictable from the independent variables. It is obtained by comparing the fitted model to a model with no explanatory variables except for the mean of the dependent variable. As R-squared approaches 1.0,it suggests a model with a better fit.

RSS, on the other hand, measures the overall model fit by summing up the squared differences between observed and predicted values, which are known as the squared residuals. In contrast to R-squared, the lower the RSS, the better the model's predictions match the actual data.

a) R-squared , as it represents the proportion of variance explained by the model.
b) RSS, as it measures the overall model fit by summing up the squared residuals.
c) Both R-squared and RSS are equally valid measures.
d) It depends on the specific characteristics of the dataset.

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. In regression analysis, Total Sum of Squares(TSS),Explained Sum of Squares(ESS),and Residual Sum of Squares (RSS) are used to measure variation the dependent variable :

- **TSS:** The total variation in the dependent variable, calculated by squaring the differences between the observed values and the mean of the observed values.
- **ESS:** The variation in the dependent variable that is explained by the regression model, calculated by squaring the differences between the predicted values and the mean of the observed values.
- **RSS:** The variation in the dependent variable that is not explained by the regression model, calculated by squaring the differences between the observed values and the predicted values.

TSS is the sum of ESS and RSS:

TSS=ESS+RSS

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

Ans. The need of regularization in machine learning Regularization is a set of techniques used in machine learning to prevent overfitting, which occurs when a model learns noise and becomes too complex. Regularization can help: Generalize to new data,

Reduce the Impact of noisy data, prevent memorizing data training, and Balance model complexity and performance.

Regularization trades a marginal decrease in training accuracy for an increase in generalizability. It can potentially lead to model underfitting if it introduces too much bias, which can cause model variance to cease to decrease and even increase.

4. **What is Gini-impurity index?**

Ans. The Gini-impurity index ,also known as the Gini index or impurity, measures how often a randomly chosen element of a set would be incorrectly labeled. It's a measure of a dataset's impurity level or disorder, and is used in decision tree algorithms to determine the best split for classifying data.

The Gini- impurity index in calculated as:

- Gini impurity = 1—Gini
- Gini = sum of squares of success probabilities of each class.

The Gini impurity index ranges from zero to one, with a higher value indicating more pure nodes and a lower value indicating less pure nodes. A Gini impurity of 0 indicates a perfectly pure node, and the highest Gini impurity is 0.5, which indicates a completely impure node.

For example, if a simple has two possible classes, tea or coffee, the Gini impurity would be 1 minus the probability of tea squared minus the probability of coffee squared.

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

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

Ans. An Ensemble techniques in machine learning combine multiple models to improve the accuracy of results. They are ideal for regression and classification, where they reduce bias and variance to boost the accuracy of models. The underlying concept of ensemble learning is to combine the outputs of diverse models to create a more precise prediction. The algorithms within the ensemble learning are generally referred as "base models", "base learners" or "weak learners" in literature.

Simple Ensemble techniques:

Max voting: Multiple models are used to make predictions for each data point, and the predictions from the majority of the models are used as the final prediction.

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

Ans. Difference between Begging and Bosting:

Begging:

a) Combines multiple models trained on different subsets of data.
b) To reduce variance by averaging out individual model error.
c) Use Bootstrap to create subsets of data.

d) Each model serves equal weight in final decision.
e) Each model has an equal error rate.
f) Less prone to overfitting due to average mechanism.
g) Improves accuracy by reducing variance.
h) Random Forecast.
i) Best for high variance, and low bias model.

Boosting:

a) Train models sequentially, focusing on the error made by the previous model.
b) Reduces both bias and variance by correcting misclassifications of the previous model.
c) Re-weights the data based on the error from the previous model, making the next models focus on misclassified instances.
d) Models are weighted based on accuracy i.e., better accuracy models will have a higher weight.
e) Generally not prone to overfitting, but it can be if the number of the model or the iteration is high.
f) Achieves higher accuracy by reducing both bias and variance.
g) AdaBoost,XGBoost,Gradient Boosting Mechanism.
h) Effective when the model needs to be adaptive to errors, suitable for both bias and variance errors.

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

Ans. Out-of beg (OOB) error is a method used to measures the prediction error of random forests and other machine learning models that use bootstrap aggregating (begging).OOB error is the average error calculated using predictions from trees are not included in their bootstrap sample. When building a random forest model, each tree is trained using a subset of the original data, called the bootstrap. During training ,some observations are left out for each tree, which are considered “out-of-beg” (OOB). These OOB error for the entire random forest is calculated by averaging the OOB errors of the individual trees.

OOB error can be used to evaluate the performance of an ensemble model and to tune its hyperparameters. In a random forest,OOB error is the fraction of incorrect classifications over the number of OOB samples.

In scikit-learn, the OOB error can be obtained using the OOB score attribute of the random forest classifier or regressor. A negative OOB score is possible if the model is worse than the best constant predictor.

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

Ans. K-fold cross-validation is a machine learning technique that evaluates the performance of a made by dividing a dataset into K subsets, or folds, of roughly equal size. The model is than trained and evaluated K times, each time using a different fold as the validation set.

Steps for K-fold cross- validation are:

1. Shuffle the dataset.
2. Split the dataset into k subsamples.
3. In the first iteration, use the first subset as the test data.
4. Use the remaining subsets as the training data.
5. Train the model with the training data.
6. Evaluate the model using the test subset.
7. Keep the evaluation score of error rate.

The choice of K (number of folds) depends on the size and desired bias-variance trade off. Common values are K=5,K=10,and K=20. K-fold cross – validation is used for : Model selection, Parameter tuning, and Feature selection.

Cross validation can be computationally expensive, especially when the number of folds is large or when the model is complex and requires a long time to train.

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

Ans. Hyperparameter tuning is a key part of machine learning that involves finding the best hyperparameter values for a learning algorithm. These values maximize the model's performance and minimize a loss function. Hyperparameter are specific to the algorithm and are used to calculate the model parameters. Different hyperparameter values produce different model parameter values for a given data set.

Hyperparameter tuning is important because a good choice of hyperparameter can help a model meet desired metric values. However, a poor choice can lead to an endless cycle of training and optimization.

Hyperparameter optimization libraries:

- Scikit-learn
- Scikit-Optimize
- Optuna
- Hyperopt
- Ray.tune
- Talos
- Bayesian Optimization
- Metric Optimization Engine (MOE)

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

Ans. A large learning rate in gradient descent can cause a number of issues, including:

a) **Overshooting:** The algorithm may overshoot the minimum.
b) **Oscillating Performance:** Performance may oscillate over training epochs.
c) **Lower final Performance:** The final performance may be lower.
d) **Divergence:** The loss function may diverge.

e) **Exploding gradients:** Weights may increase exponentially, causing exploding gradients.
f) **Instabilities:** Instabilities may occur.
g) **Overly high loss values:** Loss values may be overly high.

The learning rate is a hyperparameter that determines how quickly or slowly a model learns. It also plays a role in controlling the convergence and divergence of the algorithm.

A learning rate is too small can also cause issues, such as slow learning or even an inability to learn at all .

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

Ans.

13. Differentiate between Adaboost and Gradient Boosting.

Ans. <u>Difference between Adaboost and Gradient Boosting:</u>

<u>Adaboost:</u>

a) The shift made by up weighting the observations that are miscalculated prior.
b) The trees are called decision stumps.
c) Every classifier has different weight assumptions to its final prediction that depend on the performance.
d) It gives values to classifiers by observing determined variance with data. Here all the week learners possess equal weight and it is usually fixed as the rate for learning which is too minimum in magnitude.
e) Maximum weighted data points are used to identify the shortcomings.
f) The exponential loss provides maximum for the samples which are fitted in worse conditions.
g) Its focus on training the prior miscalculated observations and it alters the distributions of the dataset to enhance the weight on sample values which are hard for classification.

<u>Gradient boosting:</u>

a) It identifies complex observations by huge residuals calculated in prior iterations.
b) The trees with week learners are constructed using a greedy algorithm based on split points and purity scores. The trees are grown deeper with eight to thirty-two terminal nodes. The week learners should stay a week in terms of nodes, layers, leaf nodes, and splits.
c) The classifiers are weighted precisely and their prediction capacity is constrained to learning rate and increasing accuracy.
d) It develops a tree with help of previous classifier residuals by capturing variances in data. The final prediction depends on the maximum vote of the week learners and is weighted by its accuracy.

e) Here, the gradients themselves identify the shortcomings.
f) Gradient boosting cut down the error components to provide clear explanations and its concepts are easier to adapt and understand
g) This method trains the learners and depends on reducing the loss functions of that week learner by training the residues of the model.

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

Ans. Bias-variance tradeoff is a theoretical concept in machine learning that describes the balance between a model's bias and variance. It's a fundamental concept that directly impacts a model's predictive performance.

Bias refers to the simplifying assumptions a model makes to make the target function easier to approximate. Variance is the degree to which the estimate of the target function varies when using different training data. In the bias variance tradeoff, an increase in one component tends to results in a decrease in the other. The bias-variance tradeoff helps optimize a model's error and keep it as low as possible. An optimized model is sensitive to data patterns while also being able to generalize to new data.

15. Give short description each of Linear, RBF, polynomial Kernels used in SVM?

a) **<u>Linear Kernel:</u>** If the data can be well-separated by a linear decision boundary, a linear kernel should be used. A linear Kernel is the simplest and most computationally efficient kernel function, and it works well for low-dimensional datasets with a large number of features.

b) **<u>Polynomial Kernel:</u>** If data has polynomial features or contains interaction effects between the features, a polynomial Kernel should be used. A Polynomial Kernel maps the input data to a higher-dimensional space using polynomial functions of the original features.

c) **<u>Radial basis function(RBF):</u>** If the data cannot be well-separated by a linear or polynomial decision boundary, an RBF Kernel should be used. An RBF Kernel is a popular choice for SVMs because it can capture complex nonlinear relationship in the data. However, choosing the appropriate value of the gamma hyperparameter is critical to prevent overfitting.

The choice of kernel function can significantly impact the performance of a Support Vector Machine (SVM) model. The Kernel function determines how the SVM maps the input data into a higher-dimensional space where it can be separated by a heperlane.