**MACHINE LEARNING**

**Q-1 to Q-15 are subjective answer type questions, Answer them briefl****y.**

**Q-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 and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-2. What are TSS (Total Sum od 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: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

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

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance. R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-4. What is gini\_impurity index?**

Ans. Gini impurity is a measure of impurity or randomness used in decision tree algorithms to evaluate the quality of a split. The Gini impurity index represents the probability of misclassifying a randomly chosen element from the dataset, based on the distribution of labels in the current node.

Mathematically, the Gini impurity index for a node is calculated as follows:

Gini impurity = 1 - ∑(pi)^2

where pi is the probability of an element in the node belonging to a particular class. The index ranges from 0 to 1, where 0 represents perfect purity (all elements in the node belong to the same class) and 1 represents maximum impurity (elements are evenly distributed across all classes).

When splitting a node in a decision tree, the goal is to minimize the Gini impurity index of the resulting child nodes. A split with lower Gini impurity indicates that the resulting child nodes are more homogeneous in terms of class labels and therefore more useful for making accurate predictions.

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

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

Bottom of Form

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

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

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q7: What is the difference between Bagging and Boosting techniques?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-8: what is out-of-bag error in random forests?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-9: What is K-fold cross-validation?**

Ans: K-fold cross-validation is a technique used in machine learning to evaluate the performance of a model on a dataset. It involves dividing the dataset into k equal-sized subsets, or "folds", and then training the model k times, each time using a different fold as the validation set and the remaining k-1 folds as the training set.

During each iteration, the model is trained on the training set and then evaluated on the validation set. The performance metric, such as accuracy or mean squared error, is recorded for each iteration. At the end of the k iterations, the k performance metrics are averaged to obtain an overall estimate of the model's performance on the dataset. K-fold cross-validation is useful because it provides a more reliable estimate of the model's performance than a single train-test split, especially when the dataset is small or imbalanced. It also helps to identify potential overfitting or underfitting of the model to the training data.

**Q-10: what is hyper parameter tuning in machine learning and why it is done?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-11: what issues can occur if we have a large learning rate in gradient descent?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-12: Can we use logistic regression for classification of non-linear data? If not, why?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q-13: differentiate between adaboost and gradient boosting?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q14: what is bias-variance trade off in machine learning?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.

**Q15: Give short description each of linear, RBF, polynomial kemels used in SVM?**

Ans: R-squared and Residual Sum of Squares (RSS) are both measures used to evaluate the goodness of fit of a regression model, but they capture different aspects of model performance.

R-squared (or coefficient of determination) measures the proportion of variance in the dependent variable that is explained by the independent variables included in the model. It ranges from 0 to 1, with higher values indicating a better fit. R-squared is a standardized measure that can be compared across different models, and it is useful for comparing models with the same dependent variable.

On the other hand, RSS measures the total amount of unexplained variance in the dependent variable by the model. It is a raw measure of model performance and does not take into account the scale of the dependent variable. RSS is useful for assessing the fit of a specific model but cannot be compared across different models or datasets.

In general, R-squared is a better measure of goodness of fit than RSS because it provides a standardized measure of how well the model fits the data. However, it is important to use both measures in conjunction to fully evaluate the performance of a regression model. A model with a high R-squared value may still have a large residual sum of squares if the data has a large amount of unexplained variance. Therefore, it is important to examine both measures to fully understand the quality of the model fit.