**Machine Learning Answers**

1. Both R squared and RSS are useful measures of goodness of fit In regression.

R squared tells about what percentage of total variance in the data is explained by your model. R squared is easily interpretable and directly relates to the proportion of variants captured by the model. It's sensitive to the scale of data in case of residual sum of square, It represents the total squared distance between the actual data points and the model predictions. It is not affected by the scale of data. In most of the cases both RSS and R squared. are used together to provide a comprehensive evaluation of the regression model. The choice between the both depends on the specific goals of assessing the goodness of fit.

2)In regression analysis the TSS, ESS and RSS are the three terms which represent the different components of variability in the dependent variable.

TSS

It measures the total variability in the dependent variable around its mean. It's calculated by summing the squared differences between individual Y value and the mean of Y.

ESS

It measures the variability in the dependent variable that is explained by the regression model. Calculated by summing the squared differences between the predicted values of Y and the mean of Y.

RSS

It measures the variability in the dependent variable that is unexplained by the regression model. Calculated by summing the square differences between the actual Y values and the predicted Y values.

Equation

TSS=ESS+RSS

It represents a fundamental principle in regression analysis. It states that the total variability in the dependent variable can be decomposed into two parts the variability explained by the model & the variability left unexplained.

3) In machine learning regularisation is crucial because it helped prevent overfitting which occurs when a model performs well on the training data but poor on unseen data.

Regularisation is necessary due to the following reasons

a) Combating all over fitting: This technique introduced penalties for complex models for large feature weights, Discouraging the models from overfitting to specific details

b) Improving generalizability: Regularisation helps the model capture the true underlying relationships between features and targets this led to better performance on unseen data making the model more generalizable.

c) Balancing bias and the variance: Underfitting is like having a simple model that misses important patterns while overfitting is like having a complex model that gets lost in the noise via regularisation helps find the sweet spot between their two extremes

4) Gini impurity index Also known as Gini index or simply impurity, Is a metric used in decision tree algorithms to measure the impurity or disorder of a set of data points. It tells you how likely you are to misclassify a random data point if you were to assign it a class label Based on the distribution of classes in the set. It is used in building decision trees to determine the best splitting criteria for modes. And also it helps to identify splits that result in purer child modes, Meaning they contain data points belonging mostly to the same class.

5) Yes, unregularised decision trees are highly prone to overfitting. Decision trees build themselves by making a series of decisions at each node, Choosing the split that best reduce the impurity for the specific node. Decision tree can grow very complex creating many branches and splits the account for every small detail in the training data.

6) An ensemble technique In machine learning is a strategy that combines multiple models to improve the overall performance compared to Using a single model. It's like asking several experts for this single opinion on a complex problem and then taking The combined insights to make a more informed decision.

Some for ensemble methods are:

1) Bagging

It trains multiple models independently on different subsets of the data and they combine the predictions by averaging them or taking the majority of vote

2) Boosting

It train models sequentially Where is new model focuses on learning from the errors of the previous one and combine The predictions by weighing them based on individual performance

Advantages of ensemble method

a) Improved accuracy and generalization

b) Robustness to overfitting

c) Ability to handle complex problems

7) 7) Bagging and boosting are both popular ensemble techniques, But they differ in their approach to creating the combining multiple models.

a) Training process

Bagging: Train models independently on different subsets of the data. Each model learns without considering the others.

Boosting: Train models sequentially where each new model focuses on learning from the error of the previous one.

b) Individual model focus

Bagging: Focuses on reducing the variance of individual models.

Boosting: Focuses on reducing the bias of the individual models

c) Model combination

Bagging: Combines the predictions by averaging them over taking the majority vote

Boosting: Combines predictions by the weighing them based on their individual performance.

d) Suitability:

Bagging Best for high variance and low bias

Boosting: More effective for higher bias and low variance model

e) Computational cost

Bagging: Less computationally expensive due to parallel training of models

boosting More computationally expensive due to sequential training and need error calculations