

MACHINE LEARNING

1.

R-squared because it shows how the model fits data on a scale of 0 to 1.

2.

TSS represents total variation in variable.

ESS shows portion of variation explained by the regression model.

RSS indicates unexplained portion.

The equation would be $TSS = ESS + RSS$.

3.

Regularization in ML is necessary to prevent models from becoming overly complex and fitting noise in data to avoid overfitting.

4.

The Gini impurity index metric used in decision trees for classification. It measures the degree of impurity or disorder in a dataset/

5.

Unregularized decision trees can overfit because they can memorize the training data fully.

6.

Ensemble techniques in machine learning involve combining multiple models to improve predictive performance.

7.

Bagging involves training multiple models independently on different subsets of data and counting their predictions.

8.

Out-of-bag error is error calculated on unseen data points for random forest training.

9.

K-fold cross validation divides data into K subsets for training and validation.

10.

Hyperparameter tuning adjusts model settings to optimize performance.

11.

Large learning rates in Gradient Descent can cause unstable or divergent training.

12.

Not ideal Logistic Regression assumes linear separation between classes.

13.

Adaboost focuses on adjusting weights, while Gradient Boosting minimizes errors.

14.

Bias-variance trade-off balances model simplicity with its ability to capture patterns.

15.

Linear works well for linearly separable data. RBF useful for complex and non-linear data. Polynomial: handles non-linear data by transforming it into higher dimensions.