

Worksheet-5

Statistics

1. D
2. C
3. C
4. B
5. C
6. D
7. A
8. A
9. B
10. A

Machine learning

1. R-squared is a better measure of goodness fit model in regression as it is scale-invariant, easy to interpret and accounts for number of predictors.
2. $TSS = ESS + RSS$. TSS tells us how a dataset varies wrt the mean. ESS tells us how much of the variation in the dependent variable is explained. RSS tells us how much of the variation in the dependent variable is not explained.
3. Regularization helps improve the model by eliminating the risk of overfitting.
4. Gini Impurity measures how well does a node splits the data set between the two outcomes.
5. Unregularized decision-trees tend to become excessively complex due to which they are prone to overfitting.
6. Ensemble technique combines multiple models so that the errors are leveraged due to collective intelligence.
7. Bagging techniques learn from each other independently in parallel to determine model average while boosting techniques learn sequentially and adaptively improve model predictions of learning algorithm.
8. Out-of-bag errors measure the prediction error of random forests using bagging techniques which use subsampling with replacement to create training samples for the model to learn from
9. K-fold cross-validation has a single parameter called k that refers to the number of groups that a given data sample is to be split into
10. Hyperparameter tuning is the process of selecting the optimal values for a machine learning model's hyperparameters. Hyperparameters are settings that control the learning process of the model, such as the learning rate, the number of neurons in a neural network, or the kernel size in a support vector machine
11. Gradient descent can overfit the training data if the model is too complex or the learning rate is too high
12. Non-linear problems can't be solved with logistic regression because it has a linear decision surface

13. The most significant difference is that gradient boosting minimizes a loss function like MSE or log loss while AdaBoost focuses on instances with high error by adjusting their sample weights adaptively
14. If the algorithm is too simple then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex then it may be on high variance and low bias. In the latter condition, the new entries will not perform well.
15. A linear kernel is a simple dot product between two input vectors. The RBF kernel function for two points A and B computes the similarity or how close they are to each other. the polynomial kernel looks not only at the given features of input samples to determine their similarity, but also combinations of these. In the context of regression analysis.