

# MACHINE LEARNING

## Worksheet 2 ANSWERS

1. Expert-Verified Answer. R-squared is generally a better measure of the goodness of fit for a regression model than the residual sum of squares (RSS), is a statistical measure that represents the proportion of the variance for the dependent variable that's explained by the independent variables in the model.
2.  $TSS = ESS + RSS$ , where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Squares. The aim of Regression Analysis is explain the variation of dependent variable Y.
3. Regularization is a critical technique in machine learning to reduce overfitting, enhance model generalization, and manage model complexity. Several regularization techniques are used across different types of models.
4. Def: Gini Impurity tells us what is the probability of misclassifying an observation. Note that the lower the Gini the better the split. In other words the lower the likelihood of misclassification.
5. It is easy to go too deep in the tree, and to fit the parameters that are specific for that training set, rather than to generalize to the whole dataset. This is overfitting. In other words, the more complex the model, the higher the chance that it will overfit. The overfitted model has too many features. However, the solution is not necessarily to start removing these features, because this might lead to underfitting.
6. Ensemble learning refers to a machine learning approach where several models are trained to address a common problem, and their predictions are combined to enhance the overall performance.
7. Bagging is a learning approach that aids in enhancing the performance, execution, and precision of machine learning algorithms. Boosting is an approach that iteratively modifies the weight of observation based on the last classification.
8. A method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging).
9. K-fold cross-validation splits data into k equal parts; each part serves as a test set while the others form the training set, rotating until every part has been tested.
10. Allows data scientists to tweak model performance for optimal results.
11. If the learning rate is too high, the algorithm may overshoot the minimum, and if it is too low, the algorithm may take too long to converge. Overfitting: Gradient descent can overfit the training data if the model is too complex or the learning rate is too high.
12. Logistic regression has traditionally been used to come up with a hyperplane that separates the feature space into classes. But if we suspect that the decision boundary is nonlinear we may get better results by attempting some nonlinear functional forms for the logit function.

13. gradient boosting is more robust to outliers and noise since it equally considers all training instances when optimizing the loss function. AdaBoost is faster but more impacted by dirty data since it fixates on hard examples.
14. In statistics and machine learning, the bias–variance tradeoff describes the relationship between a model's complexity, the accuracy of its predictions, and how well it can make predictions on previously unseen data that were not used to train the model.
15. **Kernel Radial Basis Function (RBF):** Same as above kernel function, adding radial basis method to improve the transformation.  
**Polynomial Kernel:** It represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.

## **STATISTICS WORKSHEET-5**

### **MCQ'S ANSWERS**

1. d) Expected
2. c) Frequencies
3. c) 6
4. b) Chisquared distribution
5. a) Binomial Distribution
6. a) Statistic
7. a) Null Hypothesis
8. a) Two tailed
9. b) Research Hypothesis
10. a) np