

## Machine Learning Worksheet 4

Q1. The value of correlation coefficient will always be

Ans: C) between -1 and 1

Q2. Which of the following cannot be used for dimensionality reduction?

Ans: C) Recursive feature elimination

Q3. Which of the following is not a kernel in Support Vector Machines?

Ans: A) linear

Q4. Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries?

Ans: D) Support Vector Classifier

Q5. In a Linear Regression problem, 'X' is independent variable and 'Y' is dependent variable, where 'X' represents weight in pounds. If you convert the unit of 'X' to kilograms, then new coefficient of 'X' will be?

Ans: C) old coefficient of 'X'  $\div$  2.205

Q6. As we increase the number of estimators in ADABOOST Classifier, what happens to the accuracy of the model?

Ans: B) increases

Q7. Which of the following is not an advantage of using random forest instead of decision trees?

Ans: B) Random Forests explains more variance in data than decision trees

Q8. Which of the following are correct about Principal Components?

Ans: B) Principal Components are calculated using unsupervised learning techniques

C) Principal Components are linear combinations of Linear Variables.

Q9. Which of the following are applications of clustering?

Ans: B) Identifying loan defaulters in a bank on the basis of previous years' data of loan accounts.

D) Identifying different segments of disease based on BMI, blood pressure, cholesterol, blood sugar levels.

Q10. Which of the following is(are) hyper parameters of a decision tree

Ans: A) max\_depth

D) min\_samples\_leaf

Q11. What are outliers? Explain the Inter Quartile Range (IQR) method for outlier detection.

Ans: In statistics, an outlier is a data point that differs significantly from other observations. An outlier may be due to variability in the measurement or it may indicate experimental error; the latter are sometimes excluded from the data set. An outlier can cause serious problems in statistical analyses.

IQR is the range between the first and the third quartiles namely Q1 and Q3:  $IQR = Q3 - Q1$ . The data points which fall below  $Q1 - 1.5 IQR$  or above  $Q3 + 1.5 IQR$  are outliers.

Q12. What is the primary difference between bagging and boosting algorithms?

Ans: Bagging attempts to tackle the over-fitting issue. Whereas boosting tries to reduce bias.

If the classifier is unstable (high variance), then we need to apply bagging. If the classifier is steady and straightforward (high bias), then we need to apply boosting.

Q13. What is adjusted R2 in linear regression. How is it calculated?

Ans: Adjusted R-squared value can be calculated based on value of r-squared, number of independent variables (predictors), total sample size. Every time you add a independent variable to a model, the R-squared increases, even if the independent variable is insignificant. It never declines

Q14. What is the difference between standardisation and normalisation?

Ans: Standardization or Z-Score Normalization is the transformation of features by subtracting from mean and dividing by standard deviation.

Standardization	Normalization
Mean and standard deviation is used for scaling.	Minimum and maximum value of features are used for scaling
It is not bounded to a certain range	Scales values between [0, 1] or [-1, 1]
It is used when we want to ensure zero mean and unit standard deviation.	It is used when features are of different scales.

Q15. What is cross-validation? Describe one advantage and one disadvantage of using cross-validation

Ans: Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model.

Advantage: It is used to protect against overfitting in a predictive model, particularly in a case where the amount of data may be limited

Disadvantage: The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation.