Machine Learning worksheet 1 Answers

In Q1 to Q7, only one option is correct, Choose the correct option:

1. The value of correlation coefficient will always be:

A) between 0 and 1 B) greater than -1

C) between -1 and 1 D) between 0 and -1

Ans :- C) between -1 and 1

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

A) Lasso Regularisation B) PCA

C) Recursive feature elimination D) Ridge Regularisation

Ans:- B) PCA

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

A) linear B) Radial Basis Function

C) hyperplane D) polynomial

Ans:- C) hyperplane

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

A) Logistic Regression B) Naïve Bayes Classifier

C) Decision Tree Classifier D) Support Vector Classifier

Ans:- D) Support Vector Classifier

5. 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?

(1 kilogram = 2.205 pounds)

A) 2.205 × old coefficient of ‘X’ B) same as old coefficient of ‘X’

C) old coefficient of ‘X’ ÷ 2.205 D) Cannot be determined

Ans:- A) 2.205 × old coefficient of ‘X’

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

A) remains same B) increases

C) decreases D) none of the above

Ans:-

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

A) Random Forests reduce overfitting

B) Random Forests explains more variance in data then decision trees

C) Random Forests are easy to interpret

D) Random Forests provide a reliable feature importance estimate

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

In Q8 to Q10, more than one options are correct, Choose all the correct options:

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

A) Principal Components are calculated using supervised learning techniques

B) Principal Components are calculated using unsupervised learning techniques

C) Principal Components are linear combinations of Linear Variables.

D) All of the above

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

C) Principal Components are linear combinations of Linear Variables.

9. Which of the following are applications of clustering?

A) Identifying developed, developing and under-developed countries on the basis of factors like GDP, poverty

index, employment rate, population and living index

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

C) Identifying spam or ham emails

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

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

C) Identifying spam or ham emails

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

A) max\_depth B) max\_features

C) n\_estimators D) min\_samples\_leaf

Ans:- A) max\_depth

D) min\_samples\_leaf

Q10 to Q15 are subjective answer type questions, Answer them briefly.

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

Ans:- . Outliers are individual values that fall outside of the overall pattern of a data set.

The interquartile range rule is useful in detecting the presence of outlier. The interquartile range is calculated in much the same way as the range. subtract the first quartile from the third quartile:

IQR = Q3 – Q1.

The interquartile range shows how the data is spread about the median.

The interquartile range can be used to detect outliers. This is done using these steps:

1.Calculate the interquartile range for the data.

2. Multiply the interquartile range (IQR) by 1.5 (a constant used to discern outliers).

3. Add 1.5 x (IQR) to the third quartile. Any number greater than this is a suspected outlier.

4. Subtract 1.5 x (IQR) from the first quartile. Any number less than this is a suspected outlier.

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

Ans:- Bagging and Boosting are similar in that they are both ensemble techniques, where a set of weak learners are combined to create a strong learner that obtains better performance than a single one.

Bagging and Boosting get N learners by generating additional data in the training stage. N new training data sets are produced by random sampling with replacement from the original set. By sampling with replacement some observations may be repeated in each new training data set.

* In the case of Bagging, any element has the same probability to appear in a new data set. However, for Boosting the observations are weighted and therefore some of them will take part in the new sets more often:
* While the training stage is parallel for Bagging (i.e., each model is built independently), Boosting builds the new learner in a sequential way
* In Bagging the result is obtained by averaging the responses of the N learners (or majority vote). However, Boosting assigns a second set of weights, this time for the N classifiers, in order to take a weighted average of their estimates. In the Boosting training stage, the algorithm allocates weights to each resulting model. A learner with good a classification result on the training data will be assigned a higher weight than a poor one
* Bagging will rarely get a better bias. However, Boosting could generate a combined model with lower errors as it optimises the advantages and reduces pitfalls of the single model.
* By contrast, if the difficulty of the single model is over-fitting, then Bagging is the best option than boosting.

13. What is adjusted R2 in logistic regression. How is it calculated?

Ans:- The adjusted R2 square is a modified version of r-squared that has been adjusted for the no. of predictors in the model. It increases only if the new terms improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance.

It compares the explanatory power of regression models that contains different no. of predictors.

It can be calculated in terms of sum of squares.

R2 Adjusted = 1- ((1-R2)(N-1)/(N-p-1))

Where

R2 = sample r-square

P = no. of predictors

N = total sample size

14. What is the difference between standardisation and normalisation?

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