MACHINE LEARNING

Worksheet set - 1

In Q1 to Q11, only one option is correct, choose the correct option:

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B

Ans: A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?
A) Linear regression is sensitive to outliers
B) linear regression is not sensitive to outliers
C) Can't say
D) none of these

Ans: A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?
A) Positive
B) Negative
C) Zero
D) Undefined

Ans: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?
A) Regression
B) Correlation
C) Both of them
D) None of these

Ans: C) Both of them

5. Which of the following is the reason for over fitting condition?
A) High bias and high variance
B) Low bias and low variance
C) Low bias and high variance
D) none of these

Ans: C) Low bias and high variance

6. If output involves label then that model is called as:
A) Descriptive model

B) Predictive model
C) Reinforcement learning
D) All of the above

Ans: B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation
B) Removing outliers
C) SMOTE
D) Regularization

Ans: D) Regularization

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation
B) Regularization
C) Kernel
D) SMOTE

Ans: D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
A) TPR and FPR
B) Sensitivity and precision
C) Sensitivity and Specificity
D) Recall and precision

Ans: A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True
B) False

Ans: False

11. Pick the feature extraction from below:
A) Construction bag of words from a email
B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Ans: A), B), C)

In Q12, more than one options are correct, choose all the correct options

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
A) We don't have to choose the learning rate.
B) It becomes slow when number of features is very large.
C) We need to iterate.
D) It does not make use of dependent variable.

Ans: A), B)

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?
Ans:

Regularization is a method of regression, that regularizes or shrink the co-efficient estimates to zero. This method discourages learning a more complex or flexible model to avoid the risk of overfitting.

Over fitting results in the lack of accuracy in data model. To avoid this, the regularization method is used for performing data operations in a simplified or less complex manner.

14. Which particular algorithms are used for regularization?
Ans:

There are three types of regularization. Namely:

- Ridge Regression
- Lasso Regression
- Dropout

The ridge and lasso regression are classified into L2 and L1 regression algorithms.

In the Ridge regression, the original loss function is modified by adding normalized weights. In neural networks the normalized weights are in the form of squares
In the lasso regression, it uses absolute weight values for normalization.

Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.

15. Explain the term error present in linear regression equation?
Ans:

The term error is the difference between the predicted value and the actual value. In other it defines the certainty or accuracy of the formula used in predicting the missing dependent values based on the independent variables.

******End of Machine Learning Work sheet 1******