MACHINE LEARNING

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

Answer : A

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

Answer : A

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

Answer : B

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

Answer : B

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

Answer : C

6. If output involves label then that model is called as:
   A) Descriptive model B) Predictive modal
   C) Reinforcement learning D) All of the above

Answer : B

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

Answer : D

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

Answer : D

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

Answer : A

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

Answer : B

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

Answer : 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.

Answer : A & B

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

13. Explain the term regularization?

Explanation : Technique to reduce the error by fitting the function appropriately on the data set and avoid overfitting.

Overfitting is a situation when a machine learning model is constraints to data set and not able to perform well on data set.

14. Which particular algorithms are used for regularization?

Explanation : L1(Lasso-Least absolute shrinkage and selection operator) : Penalize the model based on the sum of magnitude of the coefficient.

$\lambda = \sum |\beta_j|$

λ → Shrinkage Factor

(It ignore those column who don't have any relationship with dependent variable)

L2(Ridge) : Penalize the model based on the square of the sum of magnitude of the Coefficient**.**

$\lambda = \sum |\beta^2_j|$

(It give much less importance near to zero to those column who don't have any relationship with dependent variable)

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

Explanation : The margin of error within a statistical model. It refers to sum of deviation within the regression line, which provides an explanation for the difference between theoretical value of the model and actual observed results.

$Y = \alpha X + \beta p + €$

α, β --→ Constant parameter

X,p ---→ Independent variable

€------→ Error Term