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: Option A – Least Square Error

   - **Linear Regression is the process of finding a line that best fits the data points available on the plot, so that we can use it to predict output values for given inputs.**

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: Option A – Linear regression is sensitive to outliers

   - **The solution of the regression line will change due to outliers in most of the cases. In general, sum of squared errors is sensitive to outliers. So 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

   Answer: Option B – Negative

   - **When a line rises from left to right, the slope is a positive number. When a line falls from left to right, the slope is a negative number.**

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: Option B – Correlation

   - **The relationship is symmetric in case of correlation but in case of regression it is not symmetric**

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: Option C - Low bias and high variance

   - **Under fitted models experience high bias and less variance within their predictions and over fitting models experience low bias and high variance within their predictions**

MACHINE LEARNING

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: Option B - Predictive model

- **If output involves label then that model is called as predictive model and if output does not involve label then that model is called as descriptive model**

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

Answer: Option D – Regularization

- **There are two main regularization techniques, namely Ridge Regression and Lasso Regression**

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

Answer: Option D – SMOTE

- **A primary technique used in oversampling is SMOTE (Synthetic Minority Over-sampling TEchnique) to overcome with imbalance dataset**

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: Option A – TPR and FPR

- **It is a probability curve that plots the TPR against FPR at various threshold values and essentially separates the 'signal' from the 'noise'.**

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

Answer: Option A – True

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: Option A - Construction bag of words from a email

MACHINE LEARNING

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: Option B – It becomes slow when number of features is very large.

Option C – We need to iterate.

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

13. Explain the term regularization?

- Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid over fitting.
- In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables. It maintains accuracy as well as a generalization of the model.
- It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of over fitting.

14. Which particular algorithms are used for regularization?

- Ridge Regression
- LASSO (Least Absolute Shrinkage and Selection Operator) Regression
- Elastic-Net Regression

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

- Within a linear regression model tracking a stock's price over time, the error term is the **difference between the expected price at a particular time and the price that was actually observed**. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.