ASSIGNMENT – 39

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) 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
Answer: 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
Answer: 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

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) 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
Answer: B) Predictive model

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

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

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: C) Sensitivity and Specificity and 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
Answer: B) 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
Answer: B) Apply PCA to project high dimensional data

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) 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.

MACHINE LEARNING
Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?
Answer: Regularization in Machine Learning is used to avoid overfitting / underfitting of a model. It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting / underfitting.

14. Which particular algorithms are used for regularization?
 Answer: There are two algorithms for Regularization:

1) Ridge Regression: Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty so that we can get better long-term predictions.In Statistics, it is known as the L-2 norm.

    In this technique, the cost function is altered by adding the penalty term (shrinkage term), which multiplies the lambda with the squared weight of each individual feature. Therefore, the optimization function(cost function) becomes:

    $$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

2) Lasso Regression: Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator. In Statistics, it is known as the L-1 norm.

It is similar to the Ridge Regression except that the penalty term includes the absolute weights instead of a square of weights. Therefore, the optimization function becomes:

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{P} X_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{P} |\beta_j|$$

Cost function

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

Answer: The Regression line does not pass through all the data points on the Scatter plot exactly unless the Correlation Coefficient is ±1. In general, the data are scattered around the regression line. The vertical amount by which the line misses a data point is called a Residual, it is the error in estimating the value of Y for that data point from its value of X using the regression line. The Root Mean Square of the residuals has a simple relation to the Correlation Coefficient and the Standard Deviation (SD) of Y: It is $\sqrt{(1 - r^2)}$ X SD(Y)