Machine Learning Assignment

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 and Receiver Operator Characteristic (ROC) 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

10. In AUC and Receiver Operator Characteristic (ROC) curve for the better model AUC (area under the curve) should be less.

(A)True (B) False

Answer:- False

11. Pick the feature extraction from below:

A) Construction bag of words from an email

B) Apply PCA to project high dimensional data

C) Removing stops words

D) Forward selection

Answer:- A

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)

Q (13 – 15):- are subjective answer type questions. Answer them briefly.

13. Explain the term regularization?

Answer:- When we use regression models to train some data, there is a good chance that the model will overfit the given training data set. Regularization helps sort overfitting problem by restricting the degrees of freedom of a given equation i.e. simply reducing the number of degrees of a polynomial function by reducing their corresponding weights. In a linear equation, we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependent variable. So, regularization constraints the weights of such features to avoid overfitting.

To regularize the model, a shrinkage penalty is added to the cost function. Different types of regularization in regression:

- LASSO
- RIDGE
- ELASTICNET (less popular)

Regularization helps to reduce the variance of the model, without a substantial increase in the bias (Bias occurs when an algorithm has limited flexibility to learn the true signal from the dataset). If there is variance in the model that means that the model won't fit well for dataset different than training data. The tuning parameter λ controls this bias and variance tradeoff. When the value of λ is increased up to a certain limit, it reduces the variance without losing any important properties in the data. Thus, the selection of good value of λ is the key. The value of λ is selected using cross-validation methods. A set of λ is selected and cross-validation error is calculated for each value of λ and that value of λ is selected for which the cross-validation error is minimum.

14. Which particular algorithms are used for regularization?

Answer:- LASSO (least Absolute Shrinkage and Selection Operator) Regression is also known as L1 form. LASSO regression penalizes the model based on the sum of magnitude of the coefficients.

Mathematical equation of Lasso Regression: Residual Sum of Squares + λ * (Sum of the absolute value of the magnitude of coefficients)

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

Where,

- λ denotes the amount of shrinkage.
- λ = 0 implies all features are considered and it is equivalent to the linear regression where only the residual sum of squares is considered to build a predictive model
- λ = ∞ implies no feature is considered i.e, as λ closes to infinity it eliminates more and more features
- The bias increases with increase in λ
- variance increases with decrease in λ

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

Answer:- Linear regression most often uses mean-square error (MSE) and is similar to mean absolute error (MAE) but noise is exaggerated and larger errors are "punished. It is harder to interpret than MAE as it's not in base units, however, it is generally more popular.

To calculate the error of the model. MSE is calculated by:

1. measuring the distance of the observed y-values from the predicted y-values at each value of x;
2. squaring each of these distances;
3. Calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that result in the smallest MSE. A regression line always has an error term because, in real life, independent variables are never perfect predictors of the dependent variables. Rather the line is an estimate based on the available data. So the error term tells you how certain you can be about the formula.