MACHINE LEARNING

1. The computational complexity of linear regression is:

A) $O(n2.4)$ B) $O(n)$

C) $O(n2)$ D) $O(n3)$

Answer: - B) $O(n)$

2. Which of the following can be used to fit non-linear data?

A) Lasso Regression B) Logistic Regression

C) Polynomial Regression D) Ridge Regression

Answer: - C) Polynomial Regression

3. Which of the following can be used to optimize the cost function of Linear Regression?

A) Entropy B) Gradient Descent

C) Pasting D) None of the above.

Answer: - B) Gradient Descent

4. Which of the following method does not have closed form solution for its coefficients?

A) extrapolation B) Ridge

C) Lasso D) Elastic Nets

Answer: - C) Lasso

5. Which gradient descent algorithm always gives optimal solution?

A) Stochastic Gradient Descent B) Mini-Batch Gradient Descent

C) Batch Gradient Descent D) All of the above

Answer: - B) Mini-Batch Gradient Descent

6. Generalization error measures how well a model performs on training data.

A) True B) False

Answer: - A) True

7. The cost function of linear regression can be given as $J(w0, w1) = 1/2m \sum m\ i=1\ (w0 + w1x(i) - y(i))2mi=1$. The half term at start is due to:

A) scaling cost function by half makes gradient descent converge faster.

B) presence of half makes it easy to do grid search.

C) it does not matter whether half is there or not.

D) None of the above.

Answer: - C) it does not matter whether half is there or not.

8. 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) Correlation

9. 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 are 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 are very large.

C) We need to iterate.

10. Which of the following statement/s are true if we generated data with the help of polynomial features with 5 degrees of freedom which perfectly fits the data?

A) Linear Regression will have high bias and low variance.

B) Linear Regression will have low bias and high variance.

C) Polynomial with degree 5 will have low bias and high variance.

D) Polynomial with degree 5 will have high bias and low variance.

Answer: - A) Linear Regression will have high bias and low variance.

*) Polynomial with degree 5 will have low bias and low variance.

11. Which of the following sentence is false regarding regression?

A) It relates inputs to outputs.

B) It is used for prediction.

C) It discovers causal relationship.

D) No inference can be made from regression line.

Answer: - C) It discovers causal relationship.

D) No inference can be made from regression line.

12. Which Linear Regression training algorithm can we use if we have a training set with millions of features?

Answer: - There are various algorithms used to ensure that we pick out the most impacting feature since we have millions of features, we can afford to remove the features that have a very low to zero contribution while building a machine learning model. Few of the algorithms that can be used are: Lasso regression where with the help of lasso we can shrink the features by marking the non-contributors to a zero, VIF (Variable Inflation Factor) helps to detect Multicollinearity visible through a heat map and can check if a feature can be neglected. We need to be cautious to make multiple tests and check through various proofs before deciding on removing features from a particular data set to avoid the curse of dimensionality. We also have the wrapper method that allows us to simply take up on the subset/sample of the whole population train the model using less time and computational resource eventually allowing us to work on larger data sets without hampering productivity. Then we have batch gradient descent, stochastic gradient descent and mini-batch gradient descent methods than allows us to again use chunks of data set and work upon them ensuring we can cover up the entire data set containing millions of features in it.

13. Which algorithms will not suffer or might suffer, if the features in training set have very different scales?

Answer: - The normal equations method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales. Features scaling is required for the various gradient descent algorithms. feature scaling will help gradient descent coverage quicker. The cost function will have the shape of an elongated bowl, so the Gradient Descent Algorithms will take a long to converge. To solve this, you should scale the data before training the model.