MACHINE LEARNING

In Q1 to Q8, only one option is correct, Choose the correct option:

1. The computational complexity of linear regression is:
   A) $(n^{2.4})$ B) (n)
   C) (n^2) D) (n^3)

   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- A)

6. Generalization error measures how well a model performs on training data.
   A) True B) False

   Answer- B) False

7. The cost function of linear regression can be given as $J(w_0, w_1) = \frac{1}{2n}\sum_{i=1}^{m}(w_0 + w_1^{(i)} - {}^{(i)})^2$.

   The half term at start is due to:

MACHINE LEARNING

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)

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- C)

In Q9 to Q11, more than one options are correct, Choose all the correct options:

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),B) and C)

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) and D)

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) and D)

Q12 and Q13 are subjective answer type questions, Answer them briefly.

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

Answer- Choosing the normal equations method is not recommended in this case because it is computationally inefficient and the main cause of computational complexity is because of inverse operation on an (n x n) matrix.

When you have a training set with millions of features, gradient descent algorithms like batch gradient descent, stochastic gradient descent or mini-batch gradient descent can be used. The best bet is stochastic gradient descent and mini-batch gradient descent as they do not need to load the entire dataset into the memory for 1 step of gradient descent. If you have enough memory to load the whole data then batch gradient descent will also work fine.

MACHINE LEARNING

13. Which algorithms will not suffer or might suffer, if the features in training set have very different scales?
Answer- If the features in the training set have very different scales, then the normal equations method remains unaffected as it does not require normalizing the features.

Whereas, in various gradient descent algorithms, feature scaling is required. Feature scaling will help gradient descent converge quicker so Gradient Descent Algorithms might suffer if the features in training set have very different scales.