MACHINE LEARNING

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

1. The computational complexity of linear regression is:

   A) $O(n^{2.4})$ B) $O(n)$

   C) $O(n^2)$ D) $O(n^3)$

   Ans 1 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

   Ans 2 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

   Ans 3. 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

   Ans 4 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

   Ans 5 D) All of the above

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

   A) True B) False

   Ans 6 B)False

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

   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.

Ans 7 A)scaling cost function by half makes gradient descent converge faster.

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

Ans 8 C) Both of them

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.

   Ans 9 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.
    E) Ans 10 A) Ans Linear Regression will have high bias and low variance.
    And D) Polynomial with degree 5 will have high bias and low variance

MACHINE LEARNING

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.

Ans 11 C) It discovers causal relationship.

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?

Ans 12) **If you have a training set with millions of features you can use Stochastic Gradient Descent or Mini-batch Gradient Descent, and perhaps Batch Gradient Descent if the training set fits in memory. But you cannot use the Normal Equation because the computational complexity grows quickly (more than quadratically) with the number of features.**

Since there are lots of features, we cannot use Normal Equations (it will be very, very computationally expensive). Instead we can use **Gradient Descent.**

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

Ans 13) **If the features in your training set have very different scales, the cost function will have the shape of an elongated bowl, so the Gradient Descent algorithms will take a long time to converge. To solve this you should scale the data before training the model. Note that the Normal Equation will work just fine without scaling.**

The Gradient Descent suffers from features of different scales, because the model will take a longer time to reach the global maximum. We can always scale the features to eliminate this problem.