MACHINE LEARNING

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

1. The computational complexity of linear regression is:

Ans: - C) O (n 2)

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

Ans: - C) Polynomial Regression

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

Ans: - B) Gradient Descent

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

Ans: - C) Lasso

5. Which gradient descent algorithm always gives optimal solution?

Ans: - D) All of the above

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

Ans: - B) False

7. The cost function of linear regression can be given as $J(w0, w1)$ = 1 2m ∑ ($w0$ + $w1x$ (i) − y (i)) m 2 i=1 .

The half term at start is due to:

Ans: - D) None of the above

8. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans: - B) Correlation

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?

Ans: - 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?

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

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

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

Ans: - 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: - You could use batch gradient descent, stochastic gradient descent, or mini-batch gradient descent. SGD and MBGD would work the best because neither of them need to load the entire dataset into memory in order to take 1 step of gradient descent. Batch would be ok with the caveat that you have enough memory to load all the data.

The normal equations method would not be a good choice because it is computationally inefficient. The main cause of the computational complexity comes from inverse operation on an (n x n) matrix.

O n2. 4 to O n3

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

Ans: - 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.