MACHINE LEARNING

1. The computational complexity of linear regression is:

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

C) $O(n^{2})$ **D)** $O(n^{3})$

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: - (B) Logistic 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: - (D) All of the above

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

A) True **B) False**

Answer: - (B) True

7. The cost function of linear regression can be given as

$J(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^{m}(w_0 + w_1 x^i - y^i)^{\,2}$**. 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.

Answer: - (D) None of the above

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

(C) Polynomial with degree 5 will have low bias and high 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: - The Normal Equations method would not be a good choice because the computational complexity grows quickly (more than quadratically) with the number of features.

But we can use the Gradient Descent i.e. Batch Gradient Descent, Stochastic Gradient Descent, or Mini-Batch Gradient Descent. Stochastic Gradient Descent and Mini Batch Gradient Descent 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. But Mini Batch Gradient Descent would be the fastest and good for optimization.

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 in training set, therefore it remains unaffected by different scales.

- If the features in the training set have very different scales, the cost function will have the shape of a drawn-out bowl, therefore the Gradient Descent algorithms will yield a long time to converge. So, to resolve this we have to scale the data before training the model.