

MACHINE LEARNING

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

1. The computational complexity of linear regression is:

Ans. **$O(n^3)$**

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

Ans. **Polynomial Regression**

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

Ans **Gradient Descent**

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

Ans. **Lasso**

5. Which gradient descent algorithm always gives optimal solution?

Ans. **All of the above**

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

Ans. **False**

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:

Ans. **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?

Ans. **Both of them (Regression & 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

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.**

(D) No inference can be made from regression line

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 . **Batch gradient descent, stochastic gradient descent or mini-batch gradient descent. It does not need to load the entire dataset into memory for taking 1st step of gradient descent. Batch gradient descent is used when it have enough memory to load all data. But Normal equations method cannot be used because computational complexity grows very quickly with number of features.**

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

Ans. **The normal equations method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales.**

Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker