

**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 = Batch Gradient Descent

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

Ans = True

**7. The cost function of linear regression can be given as  $J(w_0, w_1$**

**) =**

**1**

**$2m$**

**$\sum (w_0 + w_1 x$**

**$(i) - y$**

**$(i)$**

**)**

**$m^2$**

**$i=1$**

**.**

**The half term at start is due to:**

Ans = 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?**

Ans = Correlation

**9. Which of the following is true about Normal Equation used to compute the coefficient of the Linear**

**Regression?**

Ans= We don't have to choose the learning rate.

It becomes slow when number of features are very large.

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=Linear Regression will have high bias and low variance

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

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

Ans=It discovers causal relationship

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

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