MACHINE LEARNING

1. The computational complexity of linear regression is:

A) $O(n2.4)$

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

C) Polynomial Regression

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

B) Gradient Descent

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

C) Lasso

5. Which gradient descent algorithm always gives optimal solution?

B) Mini-Batch Gradient Descent

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

A) True

7. The cost function of linear regression can be given as $J(w_0, w_1) = 1/2m \sum_{i=1}^{m} (w_0 + w_1x^{(i)} - y^{(i)})^2$. The half term at start is due to:

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

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.

D) It does not make use of dependent variable.

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.

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

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

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?

If we have a training set with millions of features, we can use Stochastic Gradient Descent or Mini-batch GradientDescent and we can also use Batch Gradient Descent if the training set fits in memory. But we cannot use the Normal Equation because the computational complexity grows quickly with the number of features.

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

If the features in the 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 we should scale the data before training the model, but the Normal Equation will work just fine without scaling.