MACHINE LEARNING

1. The computational complexity of linear regression is:

B) $O(n)$

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?

A) Stochastic Gradient Descent

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

B) False

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

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.

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.

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

Stochastic gradient descent (SGD) is one of the methods which can perform redundant computations for large datasets. It recomputes gradients for similar examples before each

parameter update. SGD with this redundancy by performing one update at a time. Due to this it is much faster and can be used to learn online. It performs update with high variance.

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

Feature scaling is used when the features in the data have ranges which vary widely. The gradient descent algorithms may suffer since the feature scaling and convergence are dependent. This may affect the optimization. The algorithms such as Random Forest/Gradient boost decision tree do not require feature scaling which does not affect them.