MACHINE_LEARNING_WORKSHEET-1

1. The computational complexity of linear regression is:

Answer- A) $0(n^{2.4})$

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

Answer – C) Polynomial Regression

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

Answer – B) Gradient Descent

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

Answer –C) Lasso

5. Which gradient descent algorithm always gives optimal solution?

Answer – A) Stochastic Gradient Descent

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

Answer – A) True

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:

Answer – 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?

Answer –B) Correlation

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

Answer – 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?

Answer –D) Polynomial with degree 5 will have high bias and low variance.

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

Answer - C) It discovers causal relationship.

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

Answer – We can use batch gradient descent, stochastic gradient descent, or mini-batch gradient descent. Stochastic gradient and mini-batch gradient would be best because neither of them need to load the entire dataset into memory.

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

Answer –The Gradient Descent suffers from features of different scales, because the model will take a longer time to reach the global maximum. We can always scale the features to eliminate this problem.