Machine Learning Worksheet 1

Q.1. The computational complexity of linear regression is:

Ans- B) 0(n)

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

Ans- C) Polynomial Regression

Q.3. Which of the following can be used to optimize the cost function of linear regression?

Ans- B) Gradient descent

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

Ans- C) Lasso

Q.5. Which gradient descent algorithm always gives optimal solution?

Ans- A) Stochastic Gradient Descent

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

Ans- A) True

Q.7. The half term at start of the cost function of linear regression is due to:

Ans- C) It does not matter whether half is there or not.

Q.8. Which of the following have symmetric relation between dependent variable and independent variable?

Ans- B) Correlation

Q.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 is very large. C) No need to iterate. D) It does not make use of dependent variable.

Q.10. Which of the following statement 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. C) Polynomial with degree 5 will have low bias and high variance.

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

Ans- C) It discovers causal relationship.

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

Ans- Since there are lots of features, we cannot use Normal Equations (it will be very, very computationally expensive). Instead we can use Gradient Descent.

Q.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.