MACHINE LEARNING WORKSHEET 1

Question 1 : The computational complexity of linear regression is :

Answer : B) O(n)

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

Answer : C) Polynomial Regression

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

Answer : B) Gradient Descent

Question 4 : Which of the following methods does not have closed from solution for its coefficients ?

Answer : C) Lasso

Question 5 : Which gradient descent algorithm always gives optimal solution ?

Answer : A) Stochastic Gradient Descent

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

Answer : A) True

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

Answer : D) None of the above.

Question 8 : Which of the following will have symmetric relation between dependent variable and independent variable ?

Answer : B) Correlation

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

Answer : A) We don’t have to choose the learning rate.
B) It becomes slow when number of features are very large.
C) We need to iterate.

Question 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 : A) Linear Regression will have high bias and low variance.
C) Polynomial with degree 5 will have low bias and high variance.

Question 11 : Which of the following sentence is false regarding regression ?

Answer : C) It discovers causal relationship.
D) No inference can be made from regression line.

Answer in brief

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

Answer : 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.

Stochastic Gradient Descent has the fastest training iteration since it considers only one training instance at a time, so it is generally the first to reach the vicinity of the global optimum (OR Mini- batch Gradient Descent with a very small mini-batch size). However, only Batch Gradient Descent will actually converge, given enough training time. As mentioned, Stochastic Gradient and Mini Batch Gradient will bounce around the optimum, unless we gradually reduce the learning rate.

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

Answer : If the features in 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.

In order to overcome this flaw of miscellaneous data types or structures, we can use Standard Scaler (which will transform the data into a single format). It will disperse the data between plus or minus one Standard Deviation and Mean equals to zero.

Although the Normal Equation will work just fine without scaling.

****