MACHINE LEARNING WORKSHEET-1

ANSWER-: Question 1 to Question 8:

1. 0(n)
2. Logistic Regression
3. Gradient Descent
4. Lasso
5. Stochastic Gradient Descent
6. True
7. Scaling cost function by half makes gradient descent coverage faster.
8. Both of them

Question 9 to Question 11:

9: We don’t have to choose the learning rate

10: Polynomial with degree 5 will have high bias and low variance.

11: It discovers causal relationship.

QUESTION 12 TO QUESTION 13:

12 ANSWER: Linear Regression training algorithm can see have training set with millions of features you could use batch gradient descent, stochastic gradient descent, or mini gradient descent.

SGD and MBGD would work the best because neither of them need to load the entire dataset into memory in order to take 1 step od gradient descent. Batch would be ok with the cavest that you have enough memory to load all the data.

The normal equations method would not be a good choice because it computationally inefficient. The main cause of the computational complexity comes from inverse operation on an (n x n) matrix.

13 ANSWER: The features in training set have very different scales are:

Machine learning is like making mixed fruit juice. If we want to get the best mixed juice, we need to mix all fruit not by their size but based on their right proportion. We just need to remember apple and strawberry are not the same unless we make them similar in some context to compare their attribute. Similarly, in many machine learning algorithm to bring all features in the same standing, we need to do scaling doesn’t impact the model just because of their large magnitude.

Features scaling in machine learning is one of the most critical steps during the pre- processing of data before creating a machine learning model. Scaling can make a difference between a weak machine learning model a better one. The most common techniques of features scaling are normalization and standardization.

Normalization is used when we want to bound our values between two number, typically between [0,1] or [-1,1].

While Standardization transforms the data to have zero mean and a variance of 1, they make our data unitless.

THE END