Machine Learning Worksheet 1

1.b

2.c

3.b

4.c

5.d

6.a

7.c

8.b

9.a,b

10.a,d

11.c,d

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

We can use batch gradient descent, stochastic gradient descent, or mini-batch gradient descent. Stochastic gradient decent and mini-batch gradient descent would work the best because neither of them need to load the entire dataset into memory in order to take 1 step of gradient descent. Batch would be ok with the caveat that you have enough memory to load all the data.

Batch gradient descent, Stochastic gradient descent or mini-batch gradient descent. It does not need to load the entire dataset into the memory.

For taking 1st step of gradient descent. Batch gradient descent is used when it has enough memory to load all the data.

But normal equations method cannot be used because computational complexity grows very quickly with number of features.

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

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's scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent coverage quicker. The cost function will have the shape of an elongated bowl, So the Gradient Descent Algorithms will take a long to coverage. To solve this, you should scale the data before training the model.