**Assignment 5**

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

Ans:

Stochastic Gradient Descent and Mini-Batch Gradient Descent can be used for large datasets. As neither of them will need to load all the records onto the memory for calculating the gradient descent.

2. Can Gradient descent get stuck in a local minimum when training a logistic regression model?

Ans:

No, Gradient descent follows a convex function or a bowl shaped function. Meaning there will be only one global minimum.

3. Do all Gradient Descent algorithms lead to the same model, provided they run for the same no. of epochs?

Ans:

No, Gradient Descent algorithm depends on the initialization in case of Mini-batch and stochastic GD.

4. Suppose you are using a Polynomial Regression and you plot the learning curves and you notice there is a large gap between training and validation error. What is the problem? How to solve it?

Ans:

Possibly due to Overfitting of the model with training data. Common ways to fix Overfitting are,

-Regularization

-Feature selection

-Reducing model complexity

5. Suppose the features in your training dataset are in different scales. Which algorithms will suffer from this? How to handle this situation?

Ans:

Gradient Descent algorithm will be affected with different scales in features. But this can be avoided by using Feature Scaling techniques. (Like Normalization and Standardization).