**Assignment 3**

1. Which Linear Regression algorithm we can use if we 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 the Gradient Descent Algorithm get stuck in a local minimum when training a linear 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 if they are running for the same no of epochs?

Ans:

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

4. If you are doing a batch gradient descent and you are monitoring the validation error at every epoch. If the validation error is constantly increasing what can be the problem? How to fix that?

Ans,

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

-Regularization

-Feature selection

-Reducing model complexity