

# Assignment 3

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

Linear Regression having training sets with millions of features can use **batch gradient descent, stochastic gradient descent, or mini-batch gradient descent**.

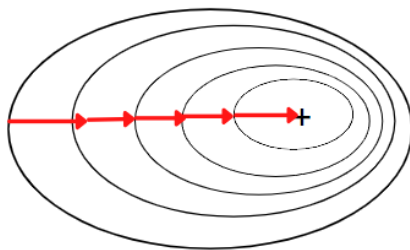
2. Can the Gradient Descent Algorithm get stuck in a local minimum when training a linear regression model?

Gradient descent produces a convex shaped graph which only has one global optimum. Therefore, **it cannot get stuck in a local minimum**.

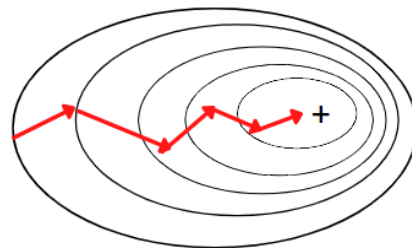
3. Do all Gradient Descent Algorithms lead to the same model if they are running for the same no of epochs?

No. The issue is that **stochastic gradient descent and mini-batch gradient descent** have randomness built into them. This means that they can find their way to nearby the global optimum, but they generally don't converge

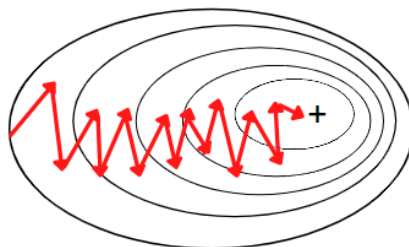
**Batch Gradient Descent**



**Mini-Batch Gradient Descent**



**Stochastic Gradient Descent**



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?

If the validation error consistently goes up after every epoch, then one possibility is that the learning rate is too high and the algorithm is diverging.

If the training error also goes up, then this is clearly the problem and you should reduce the learning rate.