Gradient Descent:

-> Gradient descent is an iterative machine learning optimization algorithm to reduce the cost function so that we have models that makes accurate predictions.

-> It helps to find out the Global Minimum.

-> We randomly initialize all the weights for a neural network to a value close to zero but not zero.

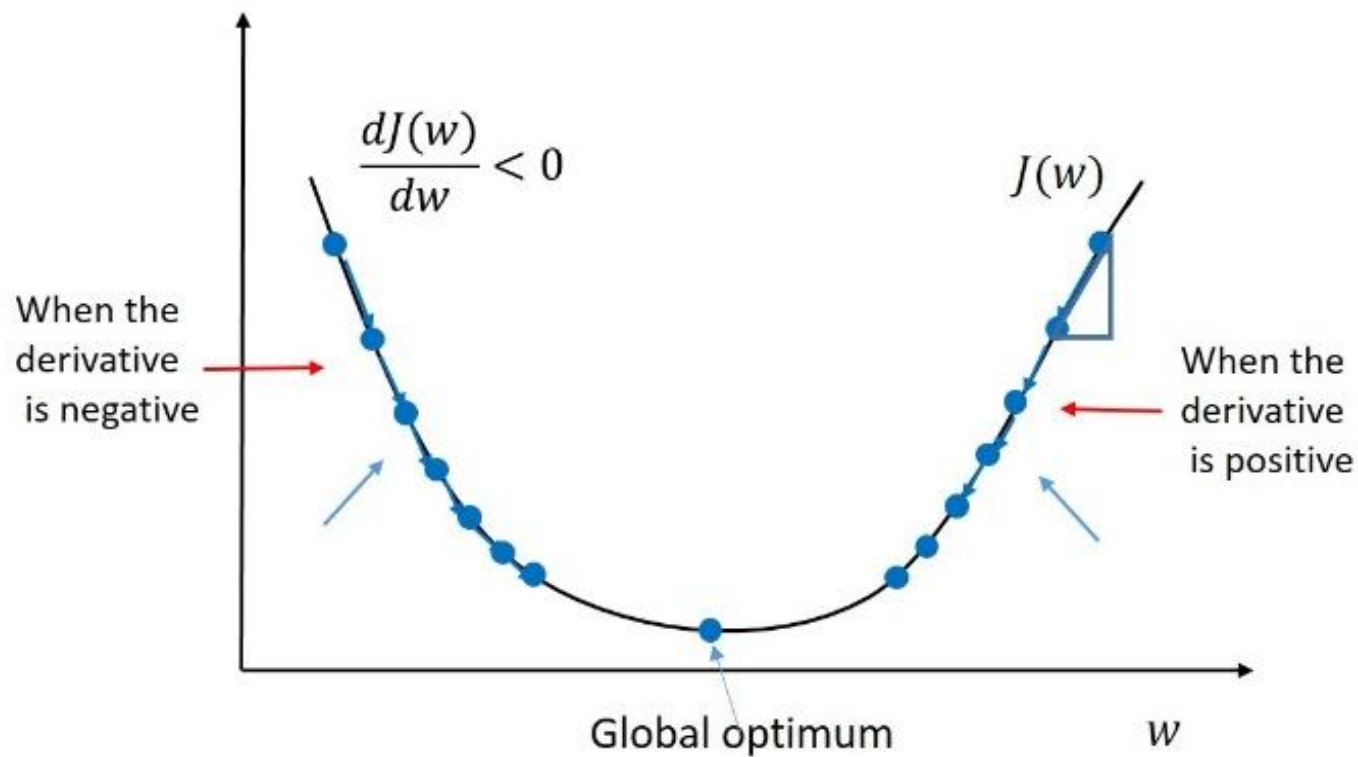## Cost Function/Loss function

-> Cost function(C) or Loss function measures the difference between the actual output and predicted output from the model.
-> Cost Function is the average of error of n-sample in the data and Loss Function is the error for individual data points.

Types of cose function/ Error:

1. MSE(Mean Square Error)

2. MAE (Mean absolute error)

$$\frac{dJ(w)}{dw} < 0$$

$J(w)$

When the derivative is negative

When the derivative is positive

Global optimum

w

Different types of Gradient descents are :

Batch Gradient Descent

-> Batch gradient descent uses the entire dataset to calculate each iteration of gradient descent
-> If the dataset is huge and contains millions or billions of data points then it is memory as well as computationally intensive.

Stochastic Gradient Descent :

-> In stochastic gradient descent we use a single datapoint or example to calculate the gradient and update the weights with every iteration.

-> Learning is much faster than batch gradient descent.

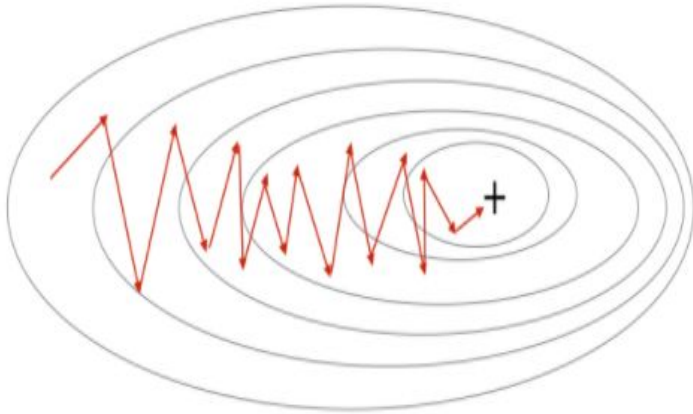-> As we frequently update weights, Cost function fluctuates heavily.
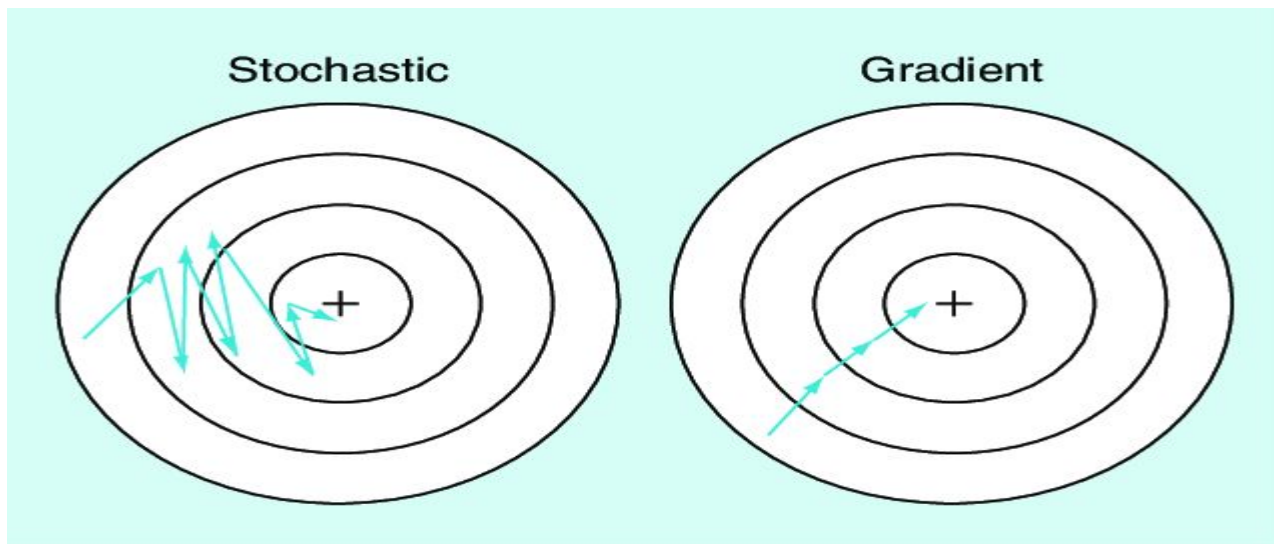
Mini batch Gradient Descent:

-> Here instead of single training example, mini-batch of samples is used.

-> Mini batch gradient descent is widely used and converges faster and is more stable.

Stochastic Gradient Descent   Mini-Batch Gradient Descent