#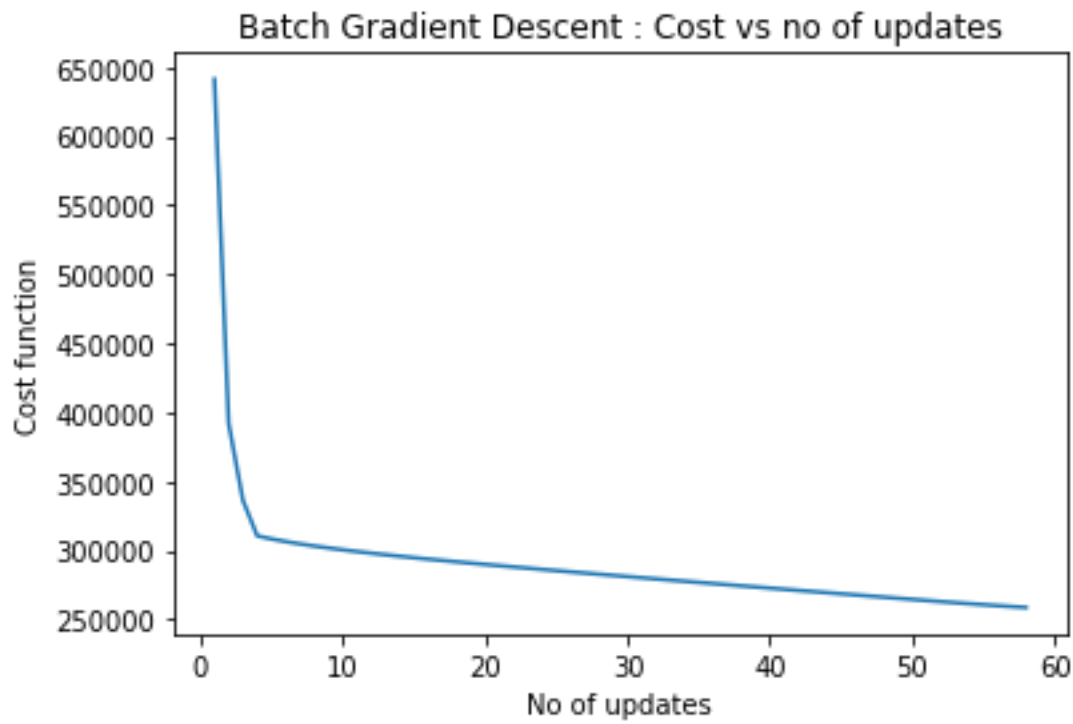 Total Time taken for convergence by each of the gradient descent techniques. Interpretation of plot and convergence times.


Batch Gradient Descent : Cost vs no of updates

Interpretation: We take the entire dataset into consideration in a single step. Hence it is called as batch and one whole batch is used ina single iteration . The graph of cost vs No of updates is a smooth decreasing curve. Takes the least time to converge.

```
Total time taken for convergence for batch gradient descent 6.338583500008
099
```

Mini Batch Gradient Descent : Cost vs Iteration
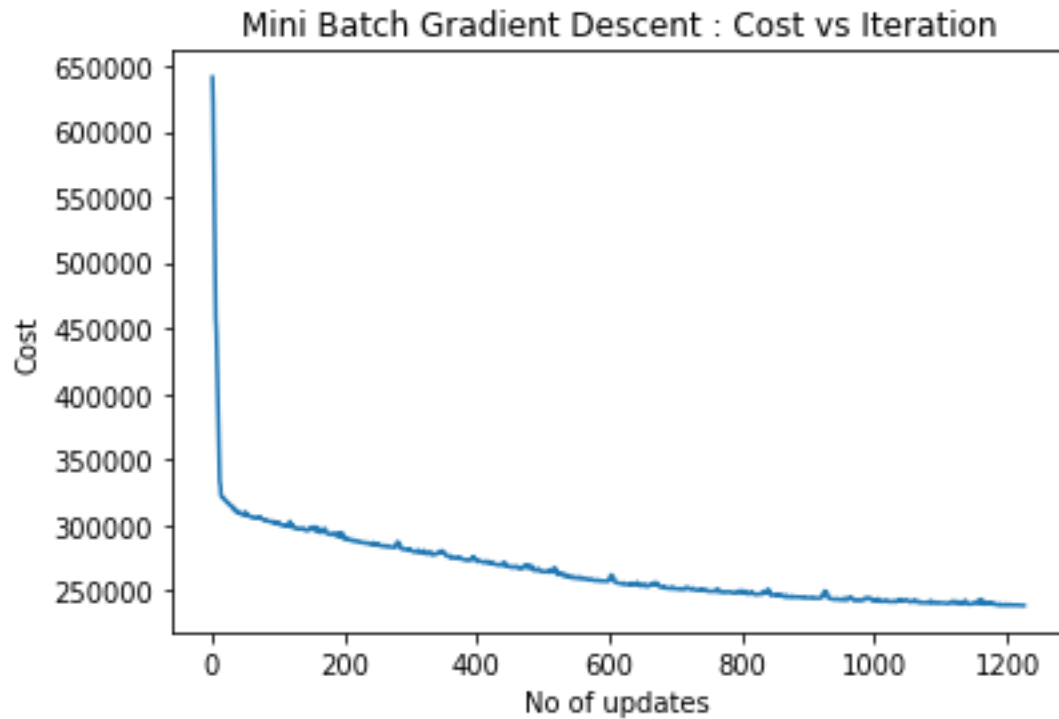
`Mini Batch` Interpretation : With respect to a subset of the training set ie in small batches, we update the parameters and compute the gradient. The batch size can be chosen by the prog rammer. Higher the batch size, lower the noise as it resembles the batch gradient descent. The graph is little more noisy than batch gradient descent.

```
Total time taken for convergence by mini batch 52.393720699998084
```

Stochastic Gradient Descent : Cost vs No of updates
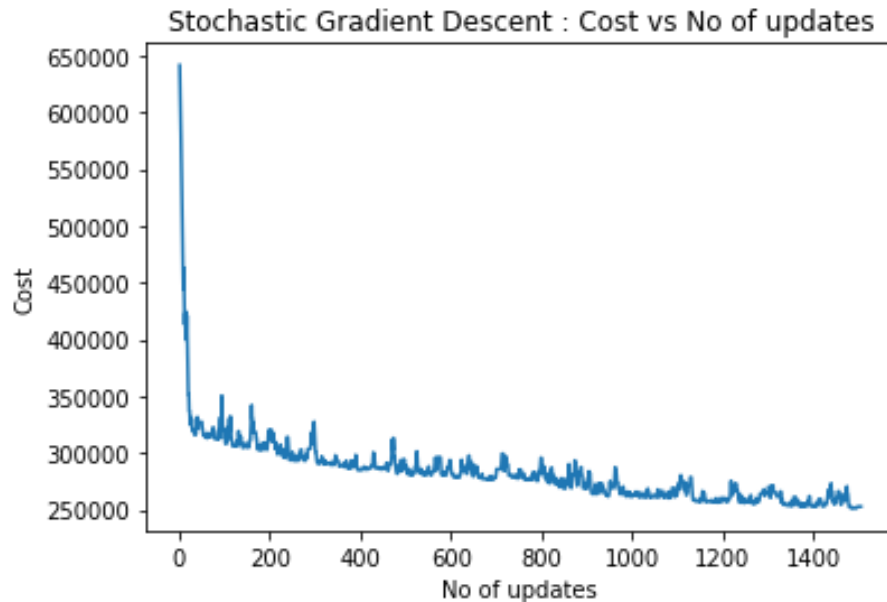
Interpretation : With respect to single training example, we update the parameters and compute the gradient. At each training instant, we update the coefficients. The updates are more noisy as we update at every single instance. Thus there is more fluctuation in the cost than batch and mini batch but overall there will be a decrease. Takes the longest time to converge.

```
Total time taken for convergence by stochastic gradient descent 142.448995
200044
```

Overall comparison between three gradient descents: