Stochastic gradient descent is a stochastic approximation of the Gradient Descent method to identify the global minimum by iterations. It is popularly used for large scale datasets. In Gradient Descent, the cost gradient is calculated based on the complete training dataset after each pass. *The cost function Q(w) can be written as Q(w)=and the cost gradient can be written as -ɳ∇Q­i(w)=- ɳ.* When there are large datasets, GD is very costly since we run through all samples in the training dataset to do a single update of the value. In GD, the weights are updated slower and it takes longer to converge to a global minimum cost due to its asymptotic rate of convergence. In stochastic Gradient Descent(SGD) however weights are computed using single sample hence is significantly faster and starts improving itself from the first sample.

Pseudocode

* Select an initial vector of parameters w and learning rate *ɳ*
* Repeat until an approximate minimum is obtained:
  + Randomly shuffle examples in the training set
  + For i=1,2,…….,*n*. do:
    - *w=w- ɳ∇Q­i(w)*

|  |  |
| --- | --- |
|  |  |

Here *Q­i(w) is the value of loss function*