Reference for introduction:

<https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/>

<https://nbviewer.jupyter.org/github/Likahler/Pdf/blob/master/ML_Andrew/LargeDataset.pdf>

* In Bigdata, no. of training data is very high (so need to be scattered among the processors) but no. of features is generally low w.r.t to m
* The model is run in aqua hpce

Even though 200 epochs are enough in our case, in order to show the paralleling power (speedup)…1000 epochs are taken into account (in real life 1000 epochs are generally used…)

Reduce the explanation on stochastic gradient descent and add a reference ofr it instead

**What is MPI?**

[**https://www.mcs.anl.gov/research/projects/mpi/**](https://www.mcs.anl.gov/research/projects/mpi/)

Plots to be included:

* Loss v/s epoch
* Speedup v/s threads v/s N=1000, 10000, etc…