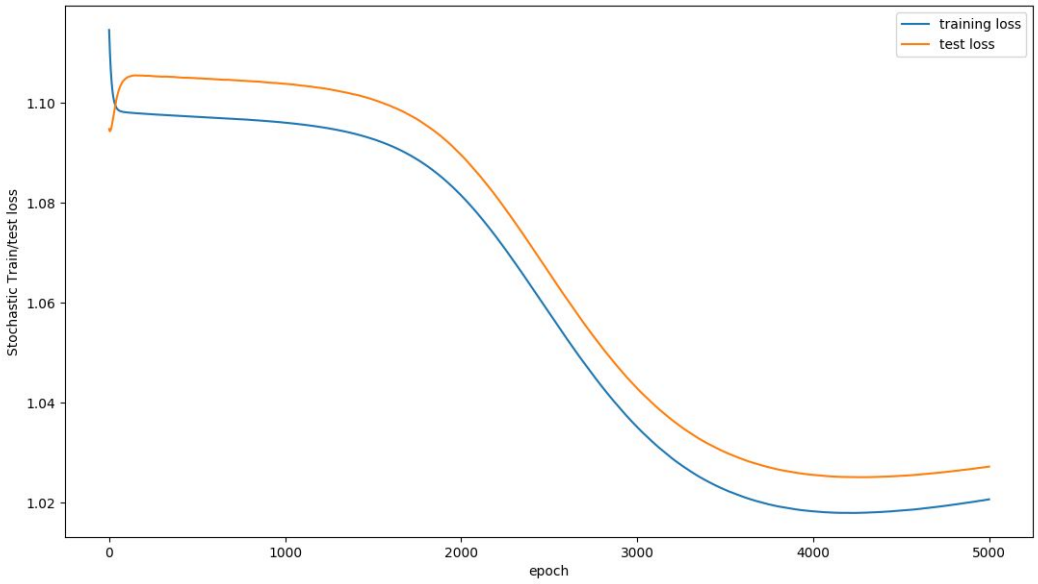
Deep Learning

HW - 2 Report

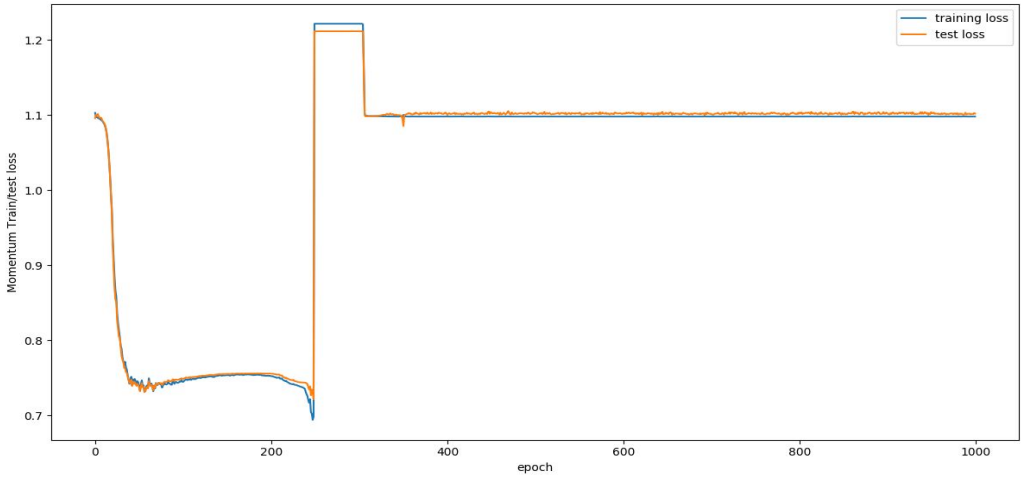
- Siddhartha Mishra
- ES15BTECH11018

Plots of train and test loss vs epoch for all the optimization methods mentioned :-

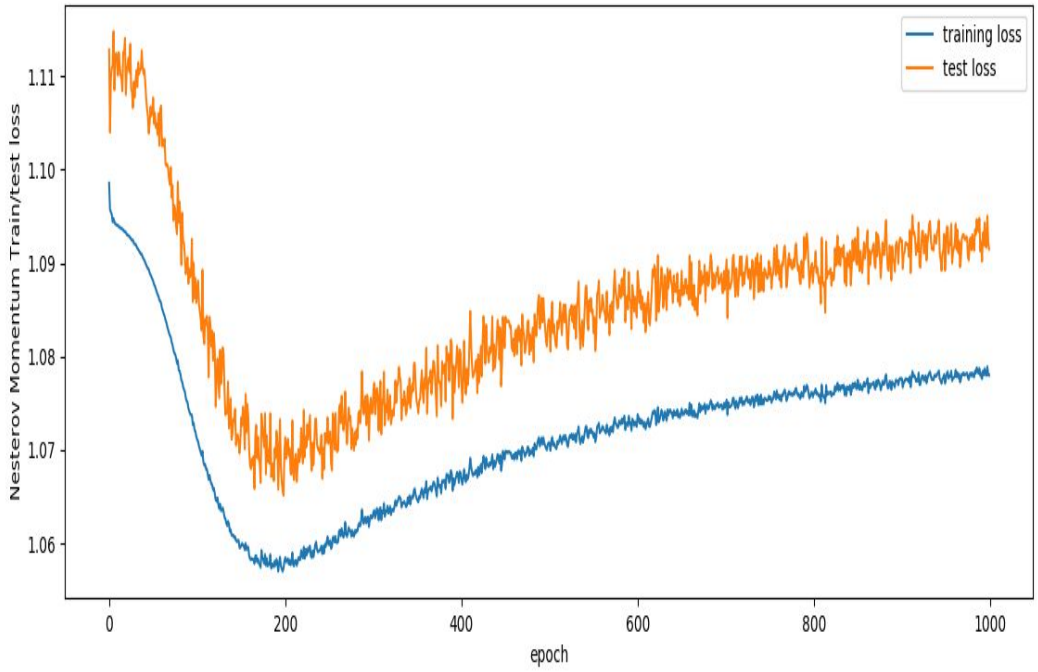
Stochastic



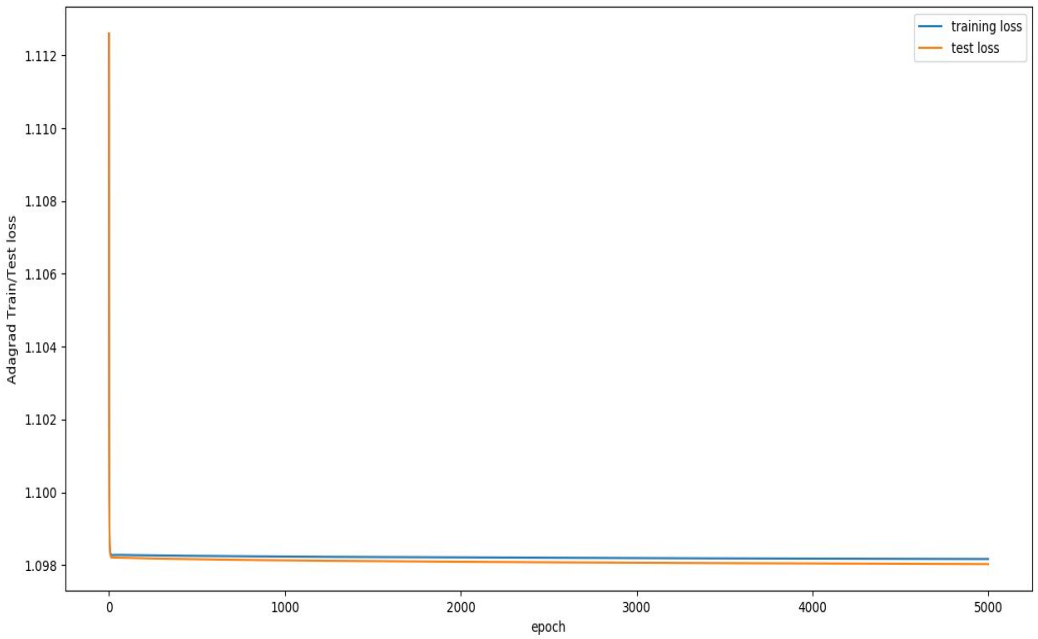
Momentum



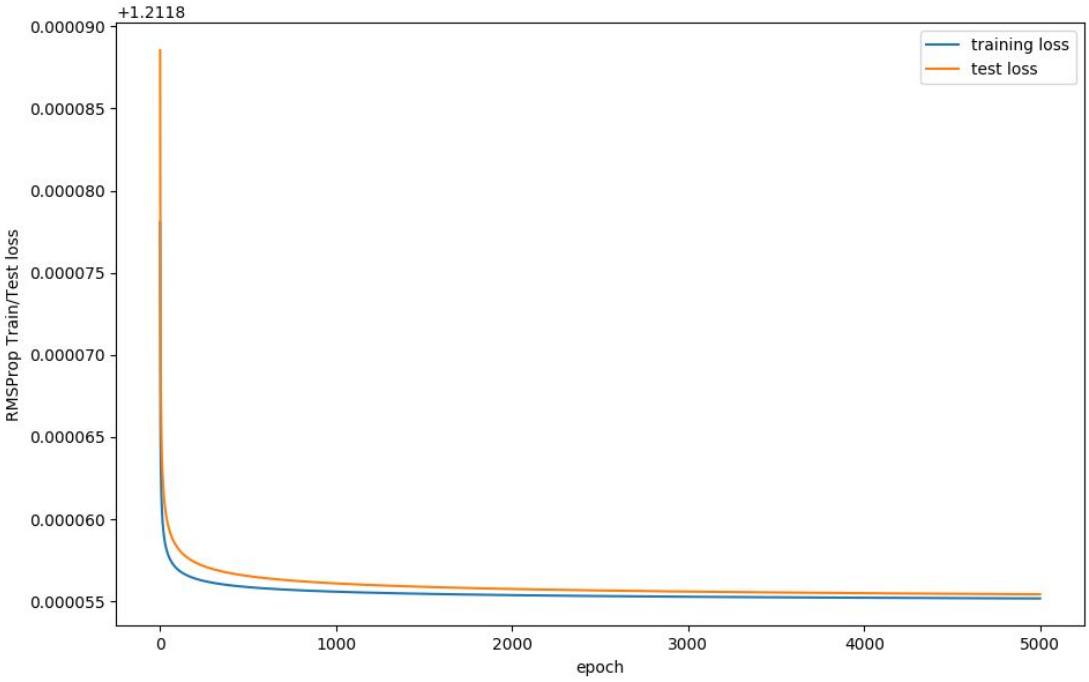
Nesterov Momentum



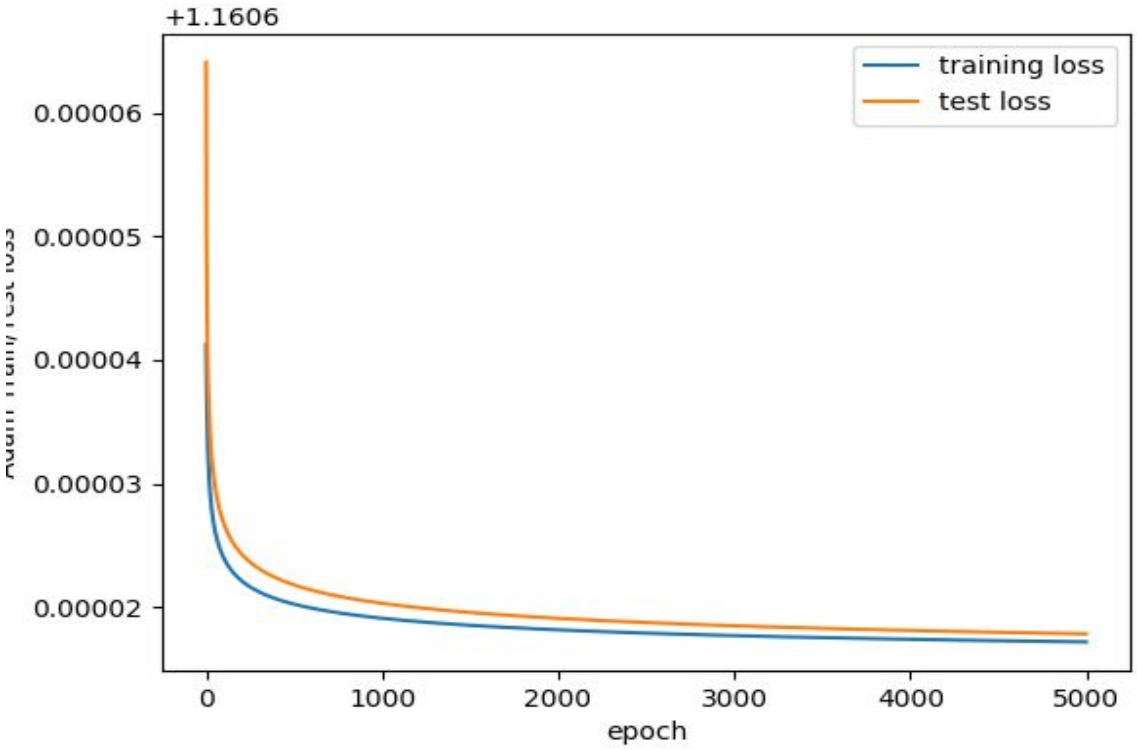
Adagrad



RMSProp



Adam



INFERENCES:

- Stochastic's loss goes down smoothly but after a point it diverges since we do not decrease the learning rate.
- Momentum and Nesterov Momentum are accelerated and go down to minima in lesser number of epochs , but diverge faster after the same points.
- Adagrad and RMSProp go down really slow and smoothly, but since the data is really small , the carefullness is leading to worse performance than stochastic.
- Adam performs better than Adagrad/RMSProp because of the bias correction term is not giving any bias towards 0.