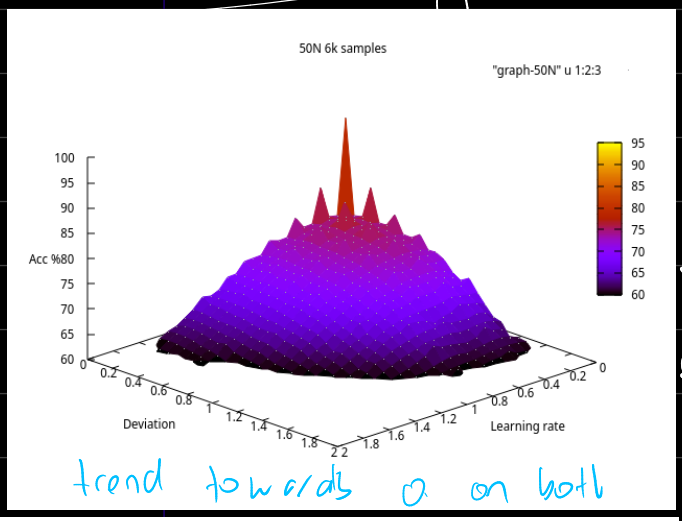
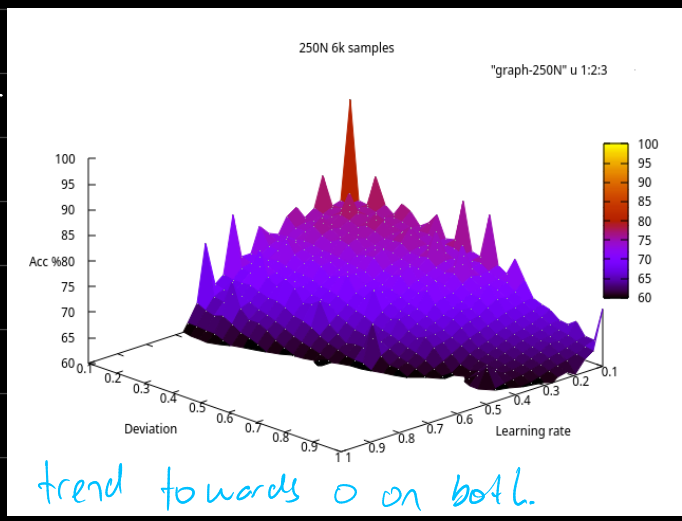
Zhilkevge

Note Coloured graphs only give average of data but are easier to see.

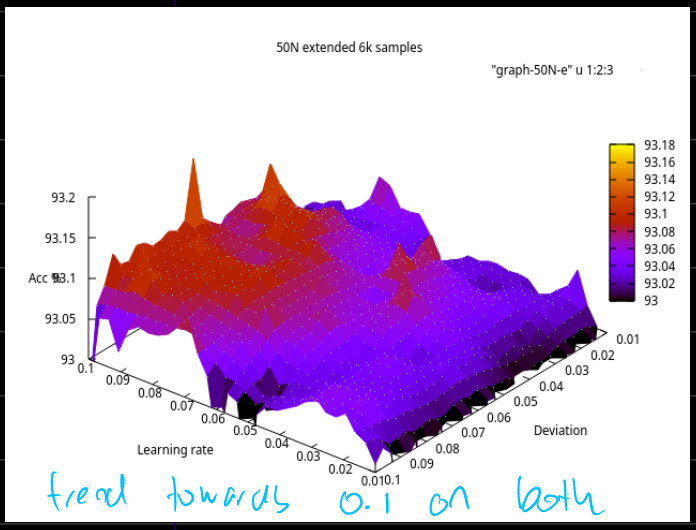


trend towards 0 on both

0-2 0.1 step Control 0-1 0.1 step

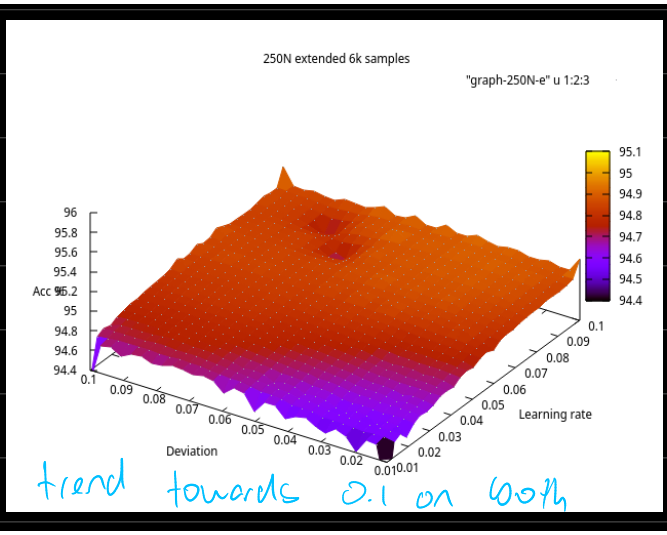


trend towards 0 on both.

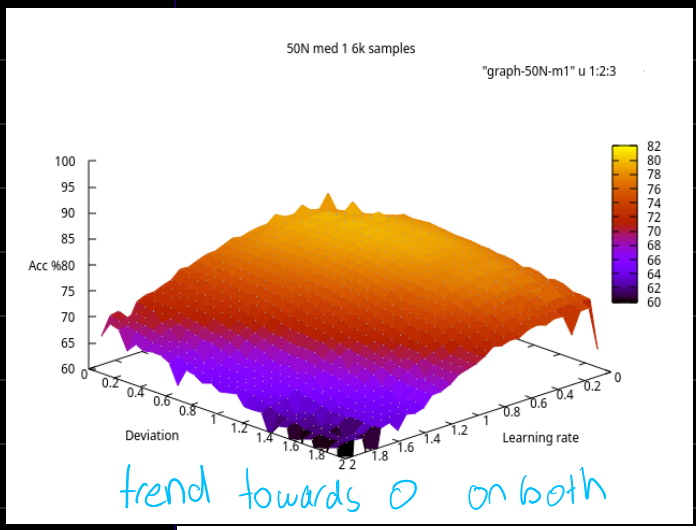


trend towards 0.1 on both

Extend 0-0.1 0.01 step

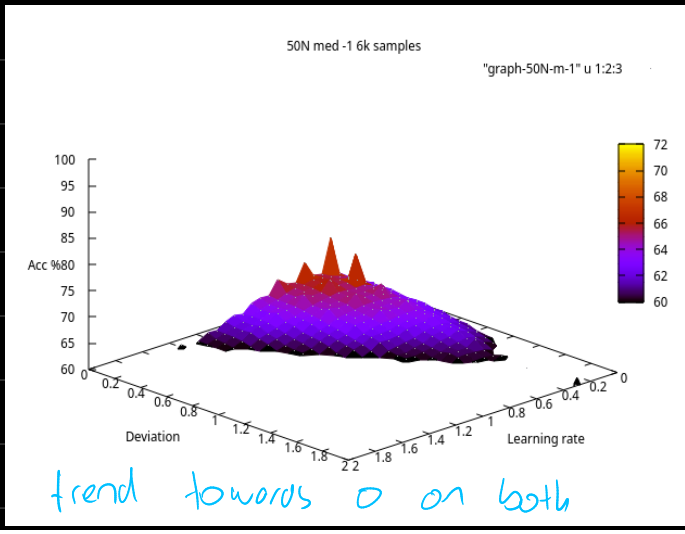


trend towards 0.1 on both



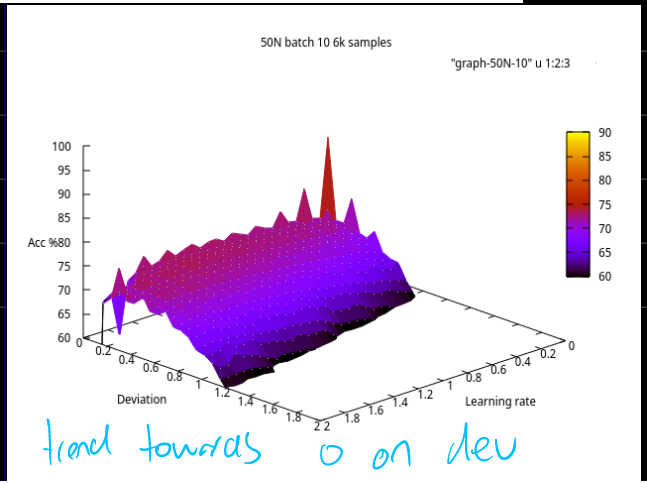
trend towards 0 on both

Median 0-2 0.1 step

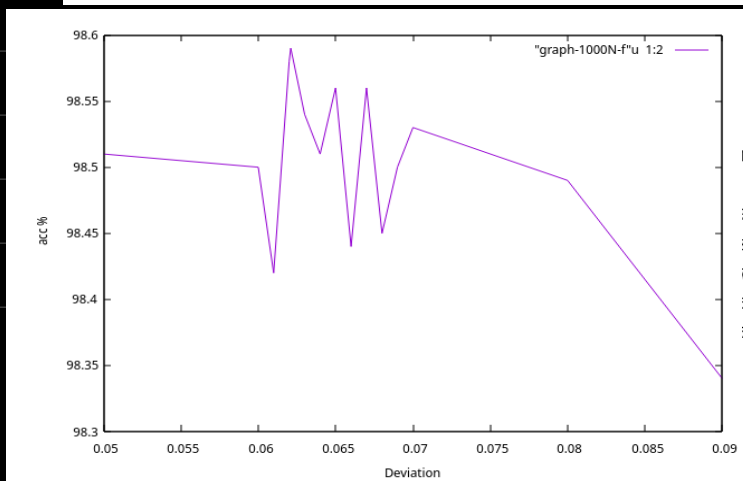


trend towards 0 on both

batch 0-2 0.1 step



trend towards 0 on dev



Strategy.

When searching I wanted to get an idea on how the meta parameters interact with each other and to do that I plotted the changes onto a 3d graph of deliction vs learning rate vs accuracy.

after seeing the trend and finding the parameters that result in highest values (control was best) I would refine the search into the direction of the trend. Then I would repeat process until I don't have time.

I started with 0-2 and 0.1 step due to wating a wide range and not too much time taken.

Q1) My final learning rate was 0.1, I have arrived on this value by following my strategy of refining the search but didn't have time to refine it more.

Q2) The 2 major impacts were calculation time and accuracy. The time increase is due to it needing to train a larger network while accuracy is due to being able to 'fit' the data better.

Q3) My final deviation was 0.623 I have arrived at it due to my strategy and it giving me the highest accuracy during testing. Could refine it more but this is close to the limit unless learning rate changes.

Q4) yes increasing the samples did increase the accuracy but the closer they were to the highest the less effect it had with the very top only increasing by ~3%. This is because the network gets a larger sample to train on and therefore it will be able to adapt to new data better due to experiencing a wider range of info at training

Q5) on-line performed better as seen on the graphs. This shows that our data had large variety as batch performs better when there is only one type of data. It also shows that the test data differs from the training such that batched data has difficulty generelizing on the training data.