Semester 2 Week 3 meeting notes

|  |
| --- |
| Short discussion of progress during semester 2 week 2  Briefly Guy mentions his current neural network progress  Guy gets better results with lower learning rates  With high learning rates you get immediate improvement from untrained network but it flattens out and gets very noisy.  He starts training with small batch size (100 or 1000 points per batch), train until loss flattens out  Then increasing batch, starting with Nadam, but eventually Nadam becomes too noisy to he changes the optimizer to SGD, but keeps relatively small learning rate.  Increase the learning rate a small amount |
| Discussion on SGD and Nadam, beginning with how Nadam performs when the learning rate is too high and SGD is slower but with more sensible behaviour.  Advice: later on in training you don’t want large learning rates and small batch sizes. Also SGD seems to perform well later on.  From a small batch size and small learning rate, you will move quickly but eventually you will be exploring the noise of the data due to each batch not being representative of the whole data. Therefore, you increase the batch size, and so that it doesn’t take forever you also need to DELICATELY increase the learning rate. In order to converge quickly but keep stability |
| Discussion of Dex  Basics of how uncertainty maps to dex  Deriving the equation to map uncertainty to dex |

Students To Do:

Guy to do:  
- provide link to adaptive batch sizes paper