To start creating my own all-encompassing, automated grid search script I needed to generate plots from Mos original scripts. I simply plotted the Mean Squared Error % (MSE) for a range of learning rates, all other parameters were kept to what Mo originally set them to.

Full\_range:  
This covered all data points originally used. A very erratic output.  
  
Small\_subset  
10 data points from the 0.0001 to 0.0009. It just looks like an exponential decay.

Large\_subset  
10 data points from 0.001 to 0.009. Similar results to Fig. 9 of Mo’s thesis, page 34. A clear minimum can be identified at 0.006 giving the same conclusion as Mo.

Higher\_res\_large\_subset  
Data from the range 0.004 to 0.008 but with 41 data points to get a high-resolution output over the optimum identified at 0.006. However, the output from this is confusing and does not follow what was expected. It appears that the MSE value calculated is not independent of the range of data used. I need to clarify how MSE is calculated, if there is any relative comparison to the other MSE calculated, then that could explain the anomaly I am seeing. Yet I don’t believe this to be the case as I assumed each evaluation of an individual learning rate should be independent of each other.

The MSE is evaluated by the ADAM optimiser program, so I could look at that to see how it’s linked.

Design questions:  
Grid Search and random search – High dimensionality for a fully automated, single nested loop process. Manually, can do one at a time but requires initial intuition to pick other parameters for the first two runs.

Two versions of grid:   
High dimension – A single big loop over all configurations  
Separated Two-stage – Find Hidden Neurons and EPOCHs first using some default/manually inputted estimates, then find Learning rate, batch size, Hidden layers and Activation functions.

Technical questions:  
Does running multiple consoles utilise separate CPU cores?