Problem 1

The eight figures below illustrate the effect of different learning rates for different noise conditions, averaged over 20 simulations each. In Figures 1-4, it seems the learning rate is too high – because for the noise = .5 condition, the network should do at least at chance (get one out of ten right, or 90% error), but the network doesn't even do that well. In Figure 5 and 6, we begin to see an average of 90% error for the noisiest condition. However, here it takes longer for the e.g. noise = .1 condition to reach asymptote. So the optimal learning rate depends on a speed-accuracy trade-off, and on what noise conditions are expected.

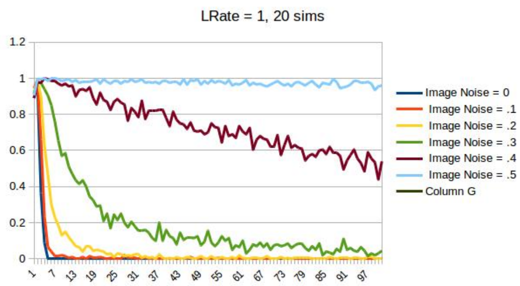

Figure 1: learning rate = 1.0

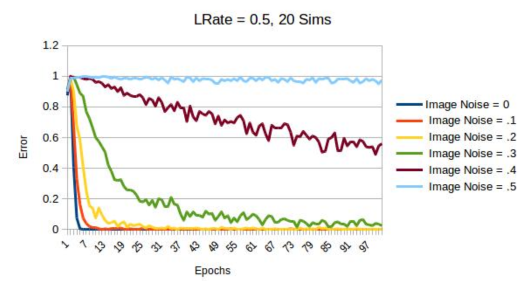

Figure 2: learning rate = 0.5

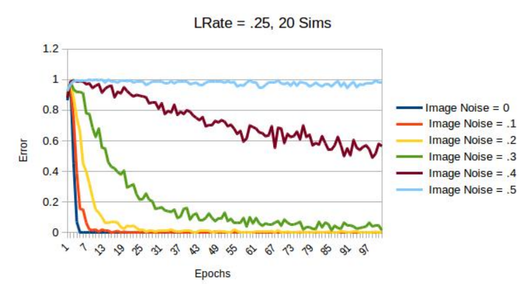

Figure 3: learning rate = 0.25

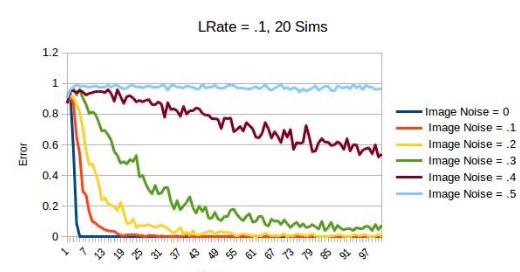

Figure 4: learning rate = 0.1

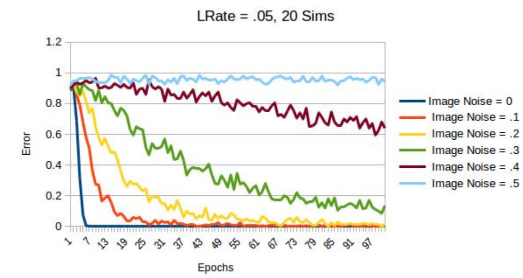

Figure 5: learning rate = 0.5

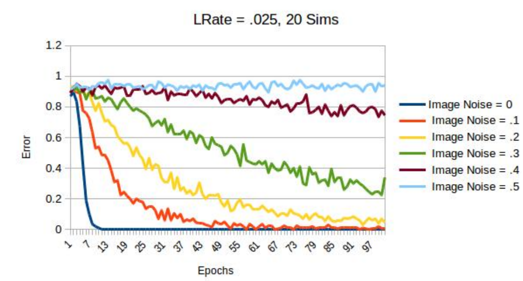

Figure 6: learning rate = 0.025

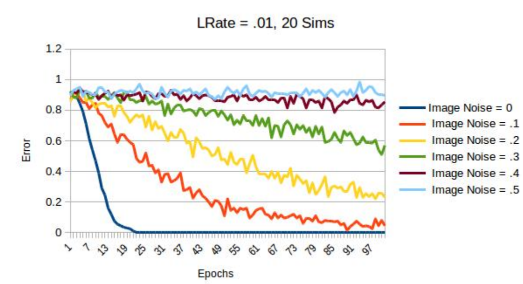

Figure 7: learning rate = 0.01

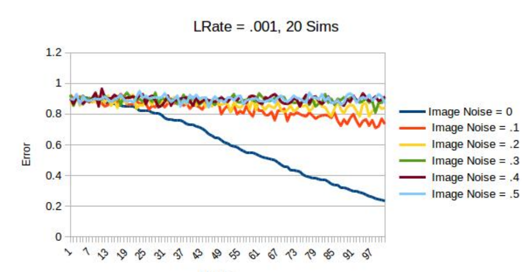

Figure 8: learning rate = 0.001

Problem 2

Figures 9 through 11 illustrate the use of different learning rates for Problem 2, averaged over 20 simulations per learning rate. It can be seen in Figure 10 that most improvement is accomplished during the first training epoch. Figure 11 then illustrates that a learning rate of 0.001 results in a little over 8% error after 100 training epochs, while a learning rate of 0.5 ultimately results in about 14% predictive error (almost double the error). So a learning rate of .001 seems ultimately to be the best here.

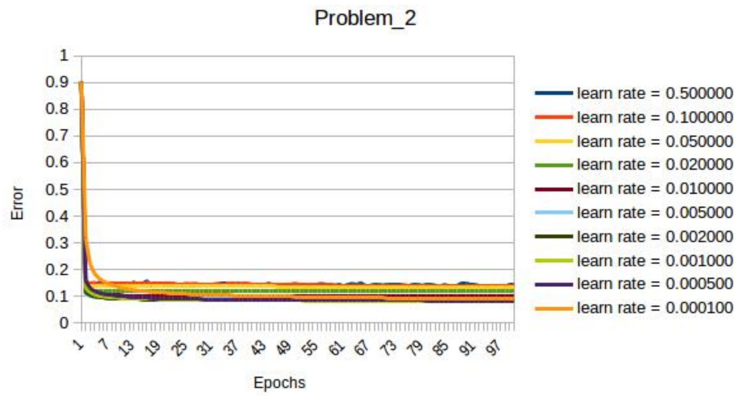

Figure 9: error over 105 training epochs

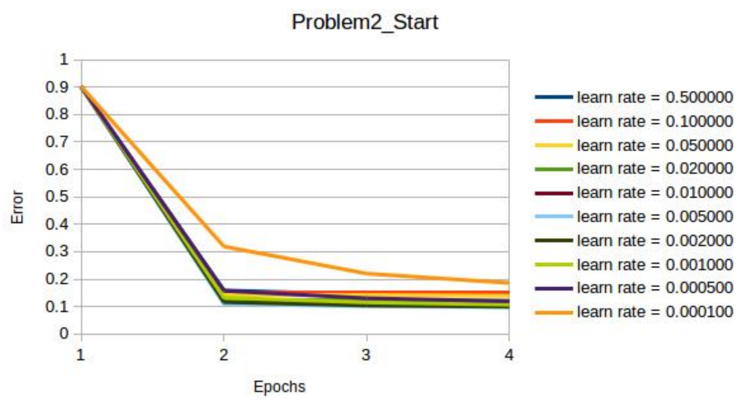

Figure 10: first five epochs

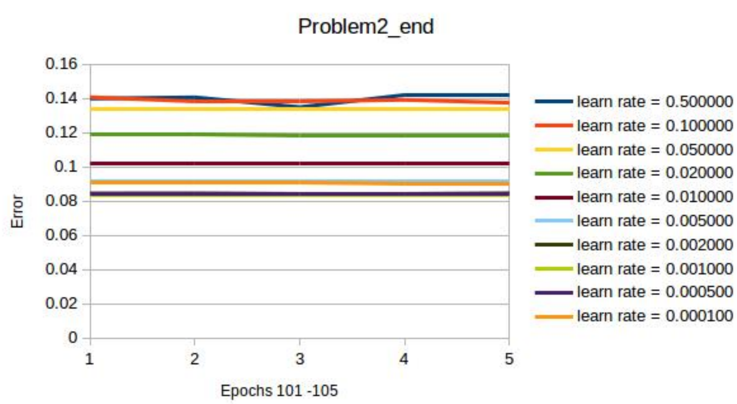

Figure 11: last five epochs