This documents FixTouch as of 5/24/2015

This program learns how to decide whether to move its eyes to increase resolution before moving its hand to try to touch an object. The program assumes that the control over the hand and eye are perfect but that there is scalar variability in the ability to see the correct target location. In particular the target location tends to be both perturbed and smeared so that there is likely to be error that grows with distance from fixation.

Another key assumption is that there is a high time cost to making a hand movement. The reward is only obtained after the hand touches the object, but if you miss you have to try again and that costs time. Thus it is better to look before you leap.

The weight updates are updating an estimate of the discounted future reward for the action in a given state, where the state is the fuzzy input pattern based on the target location. Perhaps prior hand position should also be considered. In any case this promotes generalization proportional to distance since the fuzziness is greater for greater distances.

My explorations of this revealed some tradeoffs. At first I initialized eye and hand expected reward to the same very wide Gaussian. This seemed to result in a bias toward selecting hand movements that meant that learning the value of eye movements was slow. So I tried a few things, including learning just based on the sign of the reward prediction error (based on the idea that the prediction error for eye movements is always going to have a big discount due to the delay and I didn’t want this to slow down the learning too much – not sure if this makes sense). I also tried building in relative optimism for eye movements, which may or may not help. I also explored the advantages and disadvantages of annealing the softmax scale factor – i.e. keeping it constant at 1 or letting it gradually increase over 4096 runs to a value of about 5. I’m not sure this had a big effect

In general I did not explore in detail the time course of reward rate improvement. Graphing the running average of this or of the time to correct choice might have been interesting. It might also have been interesting to explore egreedy policies rather than a softmax policy.

Reading part of Sutton and Barto was helpful in advancing my thinking about many of the issues raised above.