

Network number	Classification error	Epochs before end
1	0.5128	210
2	0.5237	158
3	0.4924	375

Figure 1: Classification error and number of epochs for each network

As you can see in table 1 network 1 and network 2 show a similar classification error, while network 2 was finished 50 episodes earlier. This is because adding an additional hidden layer significantly increases the probability of overfitting. Overfitting can be mitigated by tweeking the regularization, for example L2. As you can see in network 3, simply tweaking L2 regularization increased the number of epochs highly, showing that overfitting happened significantly later. The network was able to learn longer and deliver a better classification error, without a change of the networks layout.