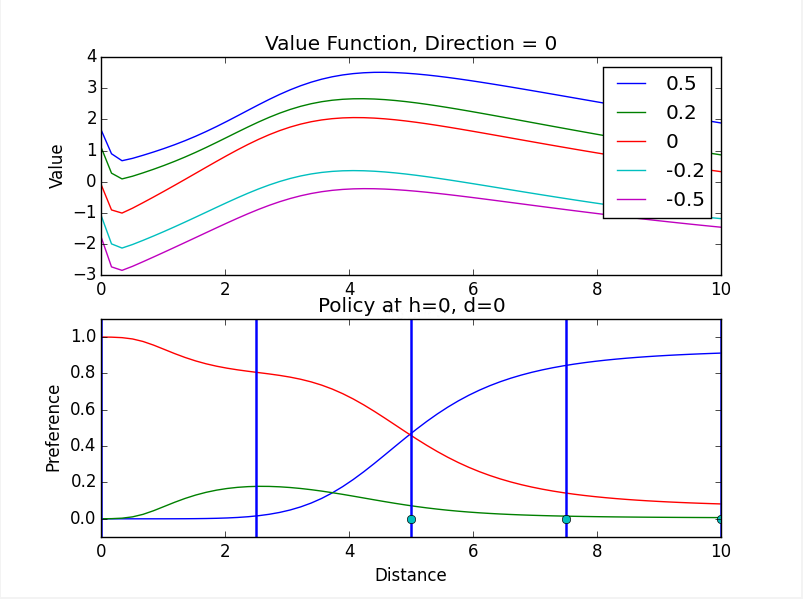
**Remove text, add information:**

Under the centroid method, the UAV circles the estimated thermal center at a fixed radius. We wanted to develop a more flexible orbiting strategy. We tried to do this three different ways, all using machine learning: table-based Q-learning, neural fitted Q-learning, and dynamic programming with neural interpolation. Table-based Q-learning discretizes the state space, and estimates the value of each possible action in each state space chunk. We used this method successfully in a low dimensional setting, but we had difficulty scaling it up. Neural-fitted Q-learning uses a neural network to stores value estimates, but we had difficulty properly incorporating new information into the neural network. Dynamic programming with neural interpolation uses a system model to simulate interactions. To allow for a continuously varying policy, we store the resulting value and policy estimates using neural networks.

RESULTS.

Using dynamic programming with neural interpolation, the learning agent was able to learn a plausible and flexible policy, which adjusted in response to thermal position and shape. This was carried out in an idealized environment, with the following state variables: distance from center of thermal, height of UAV, and direction of UAV. Current work is focused on testing the developed algorithms in the more sophisticated CRRCSim simulator.