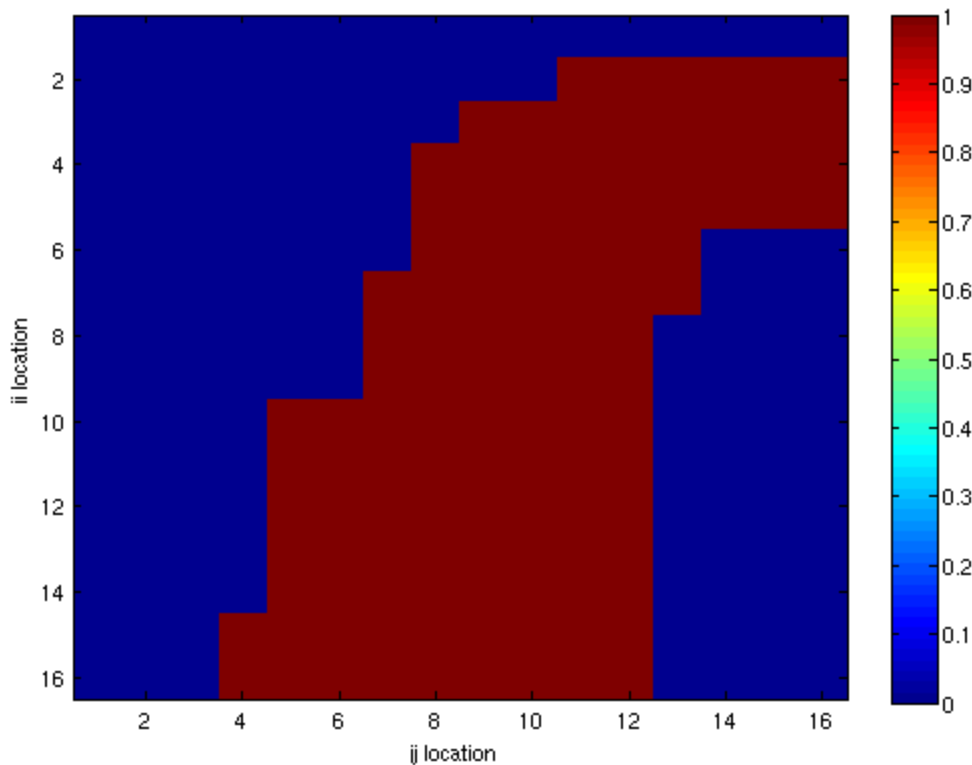
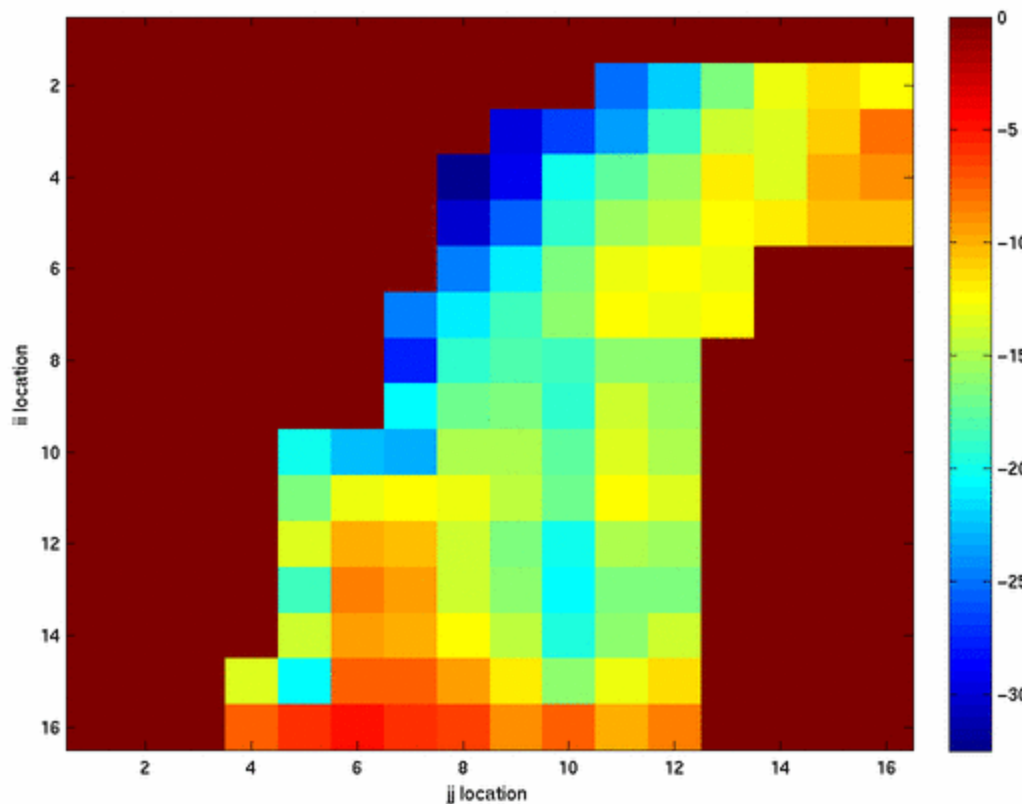


Here we present some very simple results with the racetrack example. We first constructed a sample racetrack similar to the ones that are presented in the book. The allowable spots where our "car" can be found are



We next implement a Monte Carlo learning algorithm to estimate the action value function of a vehicle driving on this track. From that we can estimate the greedy policy that maximized this action value function. As a means of visualizing the solution computed we average out the possible actions and velocities to obtain an state value function that depends on position only. When we plot this we obtain



Interpreting the values on this graph as the cost to go from the given state to the exit of the race track we see the intuitive fact that states closer to the exit are cheaper while those further from the exit are more expensive. These experiments are not complete because some of the states may have had very few samples pass through them and consequently the state value function might not be estimated very accurately. Obviously, additional analysis could be done to assess the quality of our solution.

---

[\*John Weatherwax\*](#)