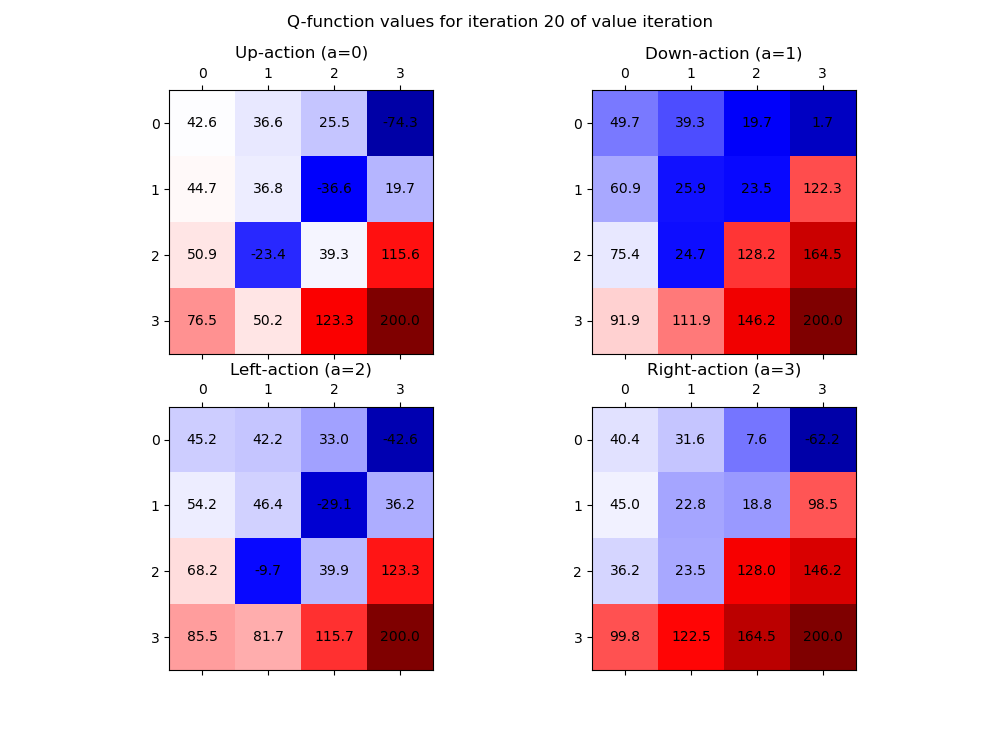
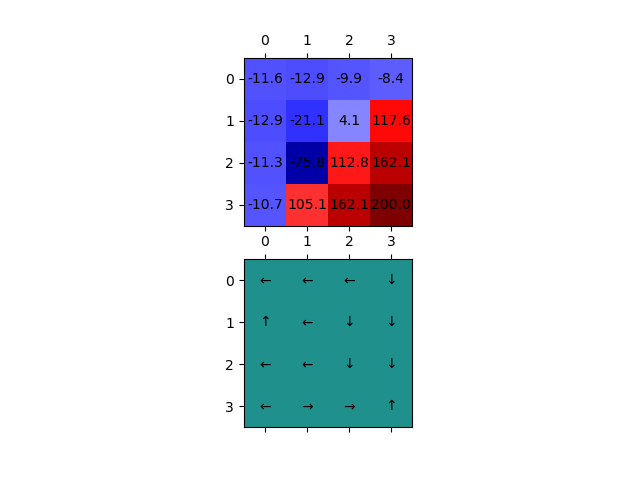
CS 5789 A1 Writeup

Name: Xingze Li NetID: xl834 15/Feb/2023

**Value iteration section (20 iteration):**

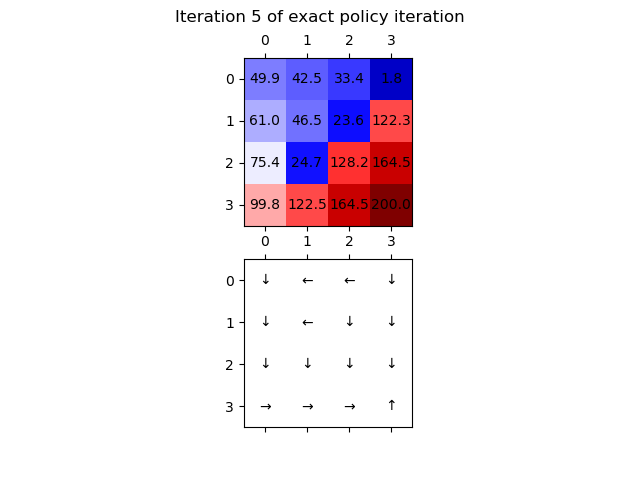


max policy and corresponding value function:

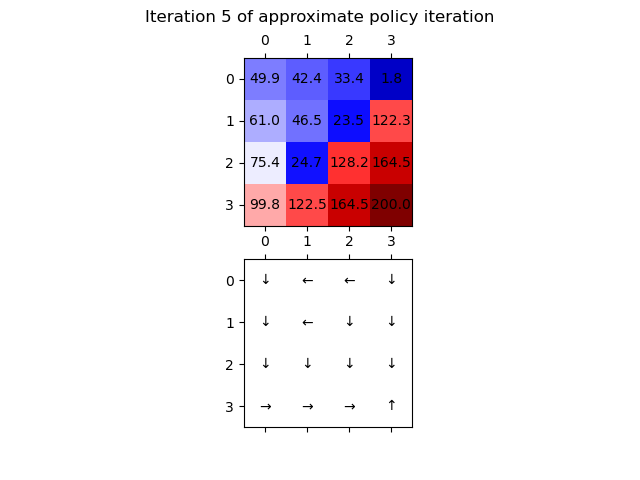


**Policy iteration section:**

Exact iteration:



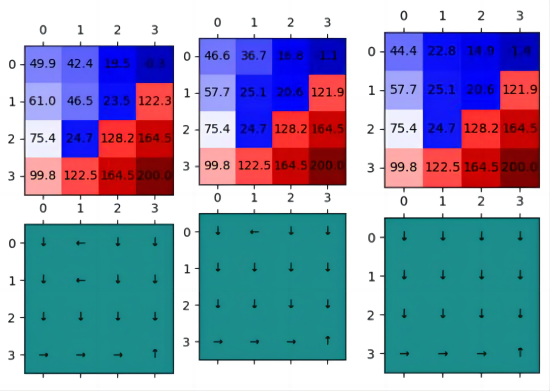
Approx iteration:



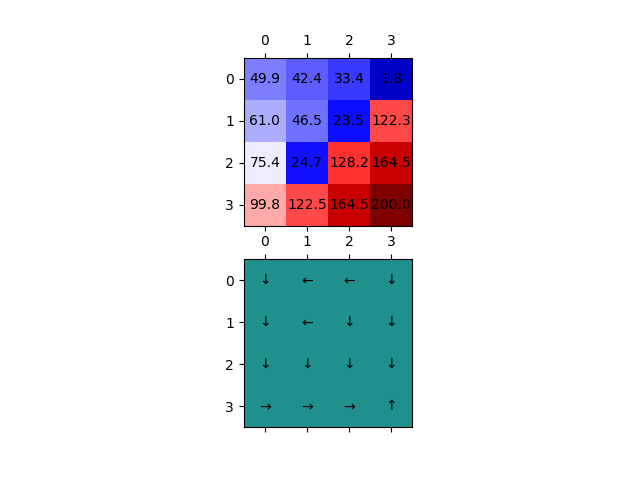
**IL section:**

With N = 75 datapoints,

The learned policy and corresponding value:



The expert policy and corresponding value:



Comparison: The expert policy is more likely to choose go left when encountering region with reward -80. In terms of value function, if policy doesn`t choose to turn left when stays at left side of -80 reward region. Its surrounding region`s value will be lower.