**ECE232E - Project 3**

**Reinforcement learning and**

**Inverse Reinforcement learning**

*Zhechen Xu (805030074), Yufei Hu (404944367), Qi Zeng (204946904), Liangkun Zhao (204947003)*

*May 20, 2018*

# **Part 2: Reinforcement learning (RL):**

**Q1:**

blank

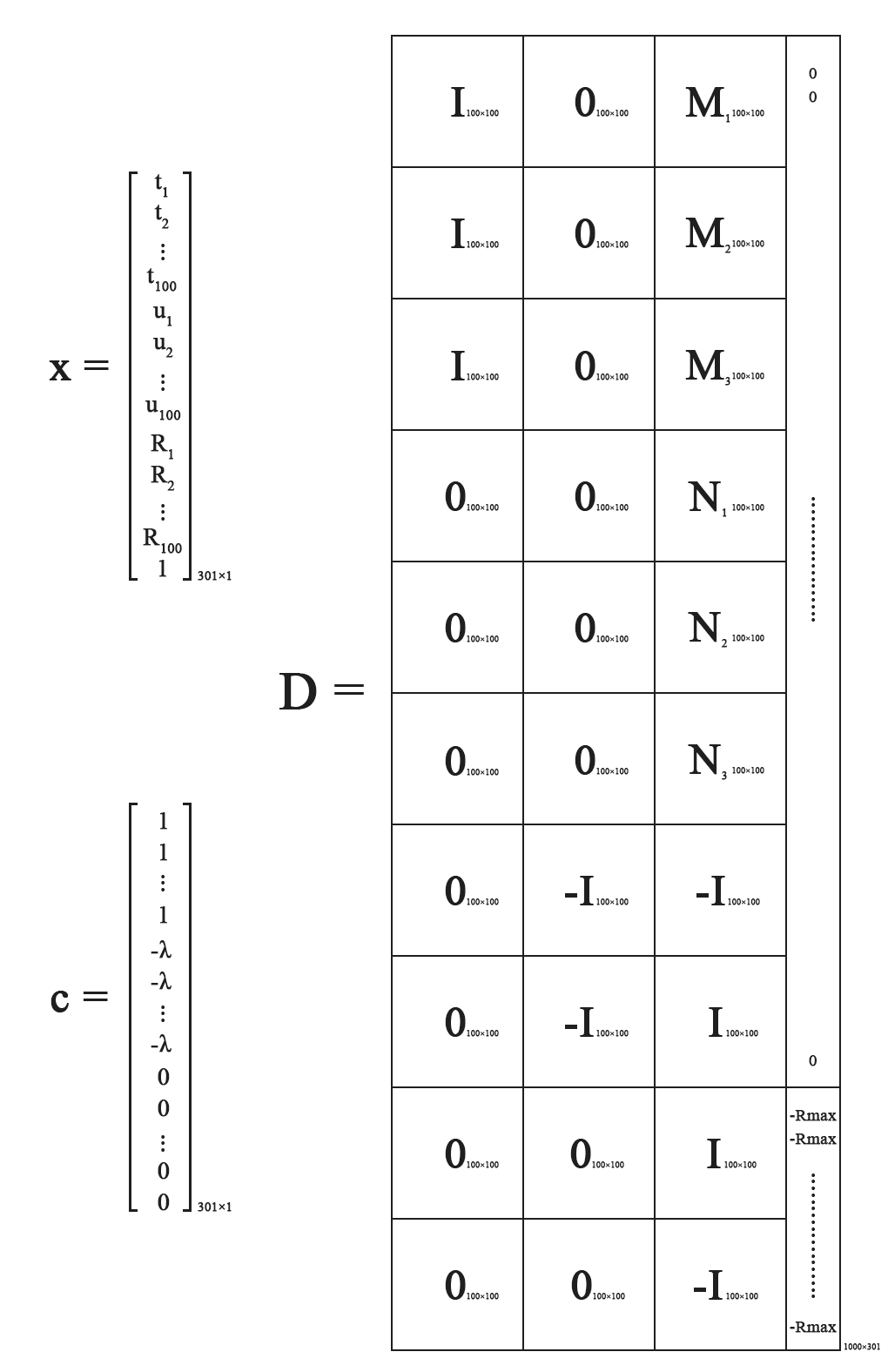
# **Part 3: Optimal policy learning using RL algorithms:**

**Q2:**

blank

# **Part 4: Inverse Reinforcement learning (IRL):**

**Q10:**



**Figure 1:** Reformulation of the original LP formula

**x**, **c**, and **D** are clearly defined in the figure above. In c, there are 100 ones at the top, followed by 100 *–λ* and 101 zeros. In D, **I** refers to an identity matrix and **0** refers to an all-zero matrix with their dimensionalities denoted in the subscript. **Mi** is defined as follows:

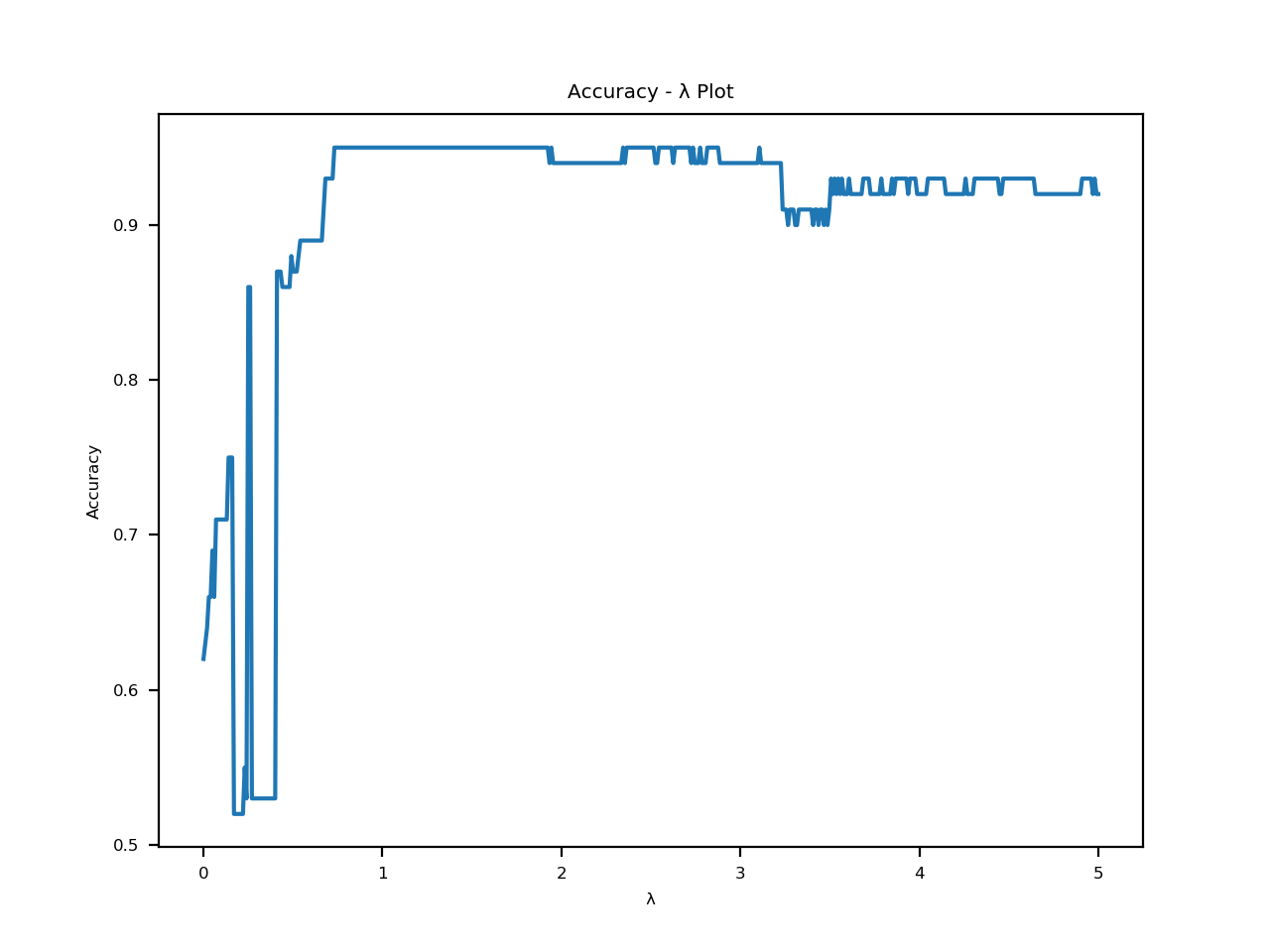


where j represents an entire row of matrix **M**.

Similarly, **Ni** is defined as:



**Q11:**

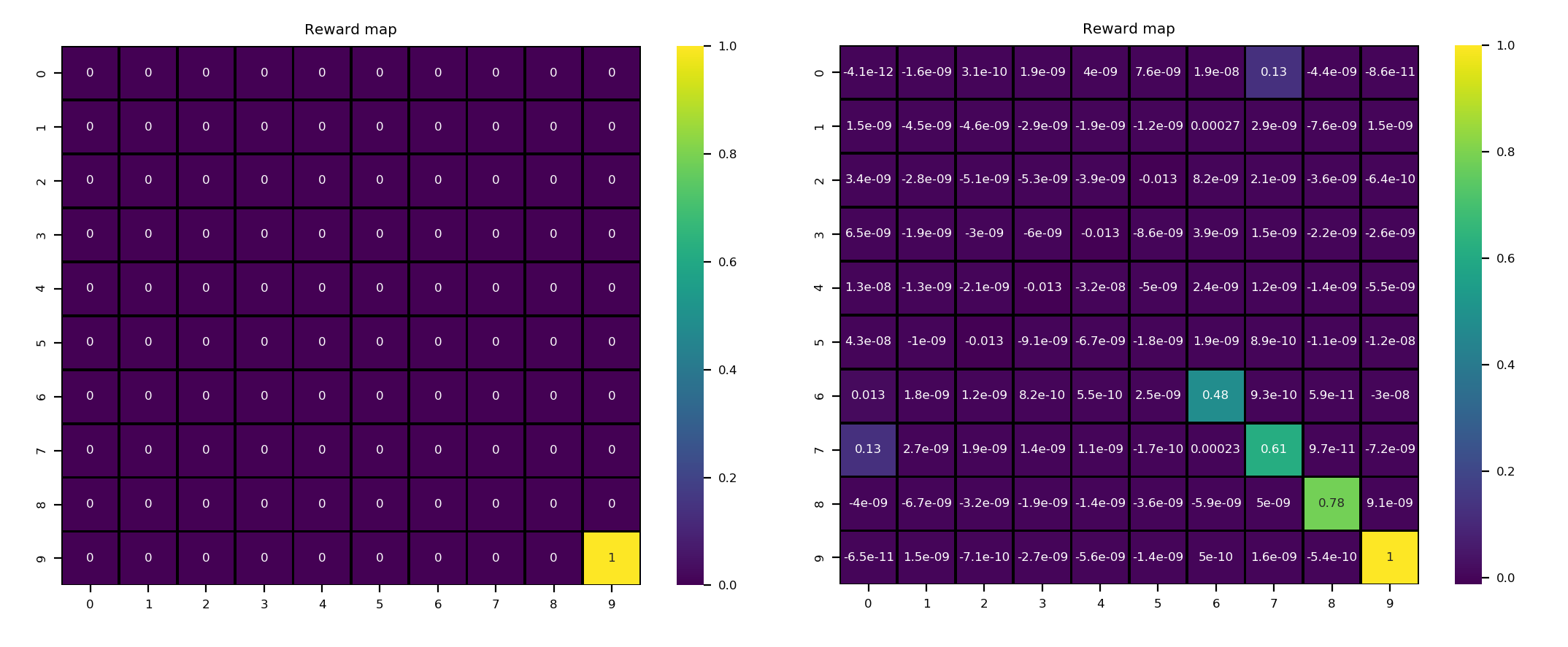


**Figure 2:** Accuracy along with different λ using reward function 1

**Q12:**

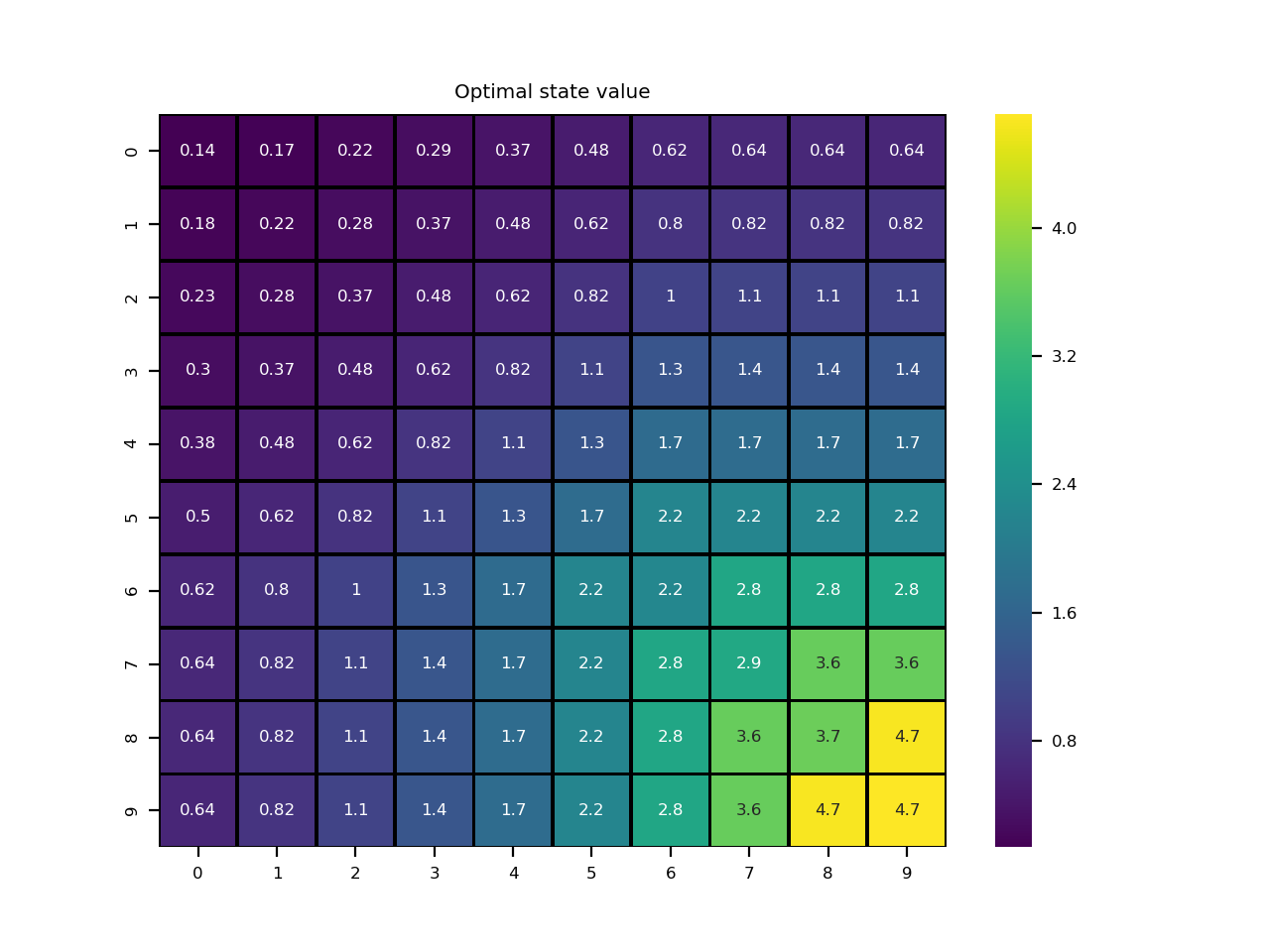


**Q13:**



**Figure 3:** The figure on the left is the heat map of the ground truth reward function 1; the figure on the right is the heat map of the extracted reward function 1.

**Q14:**

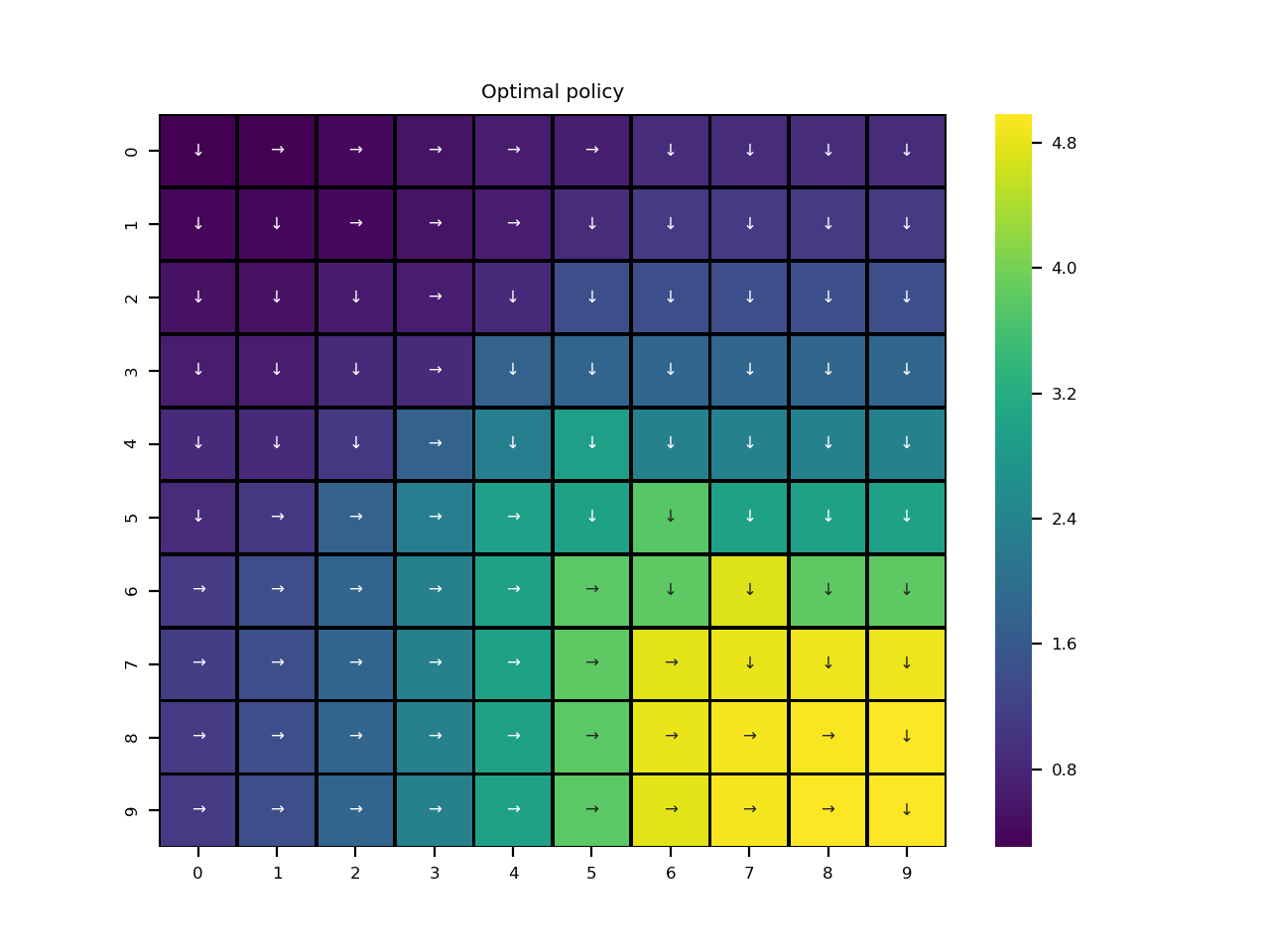


**Figure 4:** Heat map of the optimal state values based on extracted reward function 1

**Q15:**

njnj

**Q16:**

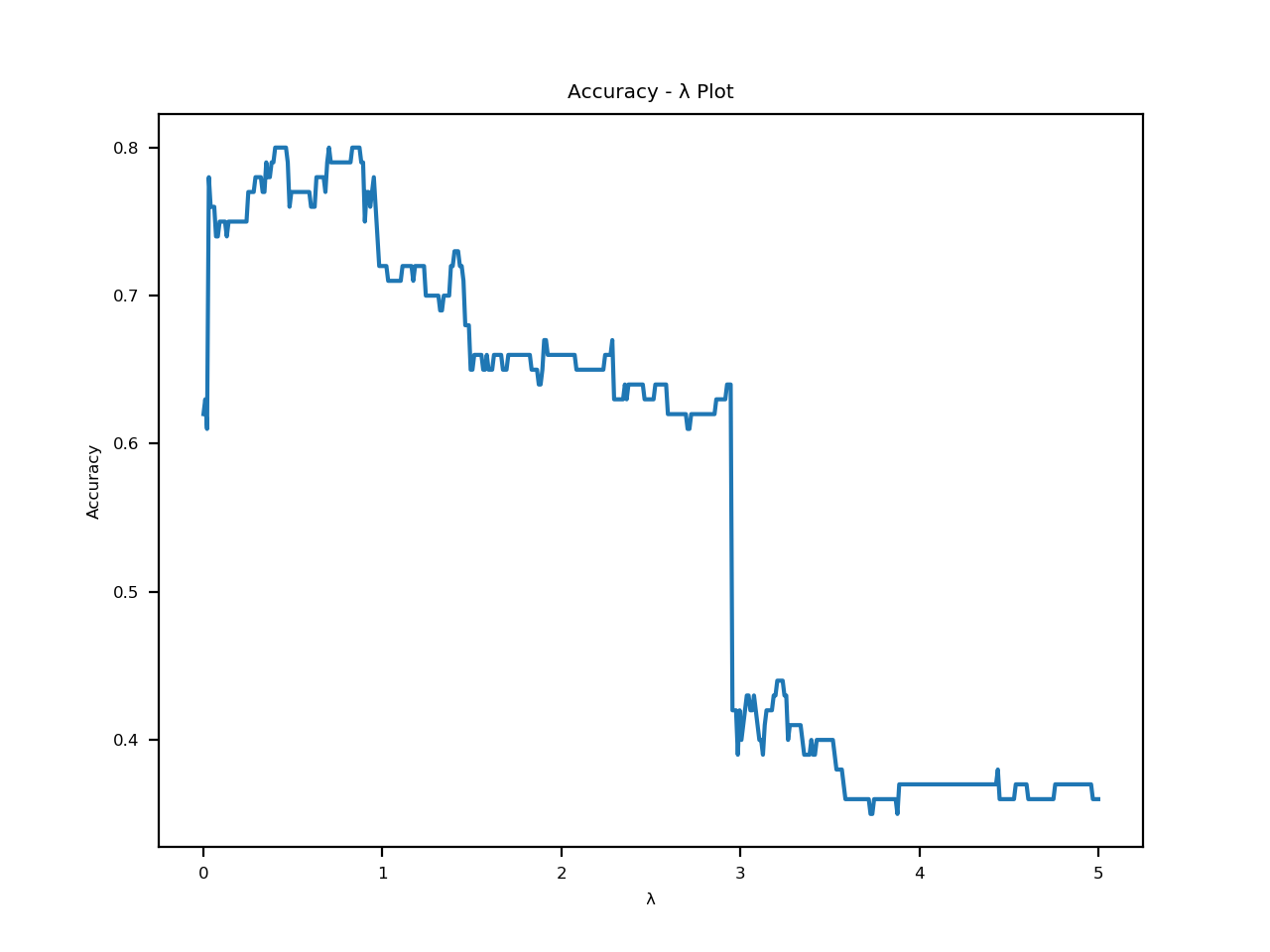


**Figure 5:** Extracted optimal policy based on extracted reward function 1

**Q17:**

According to the figure above, our extracted optimal policy is almost the same to the expert policy. This can be verified by the fact that the maximum accuracy for our extracted policy is 96%. Most of the actions point to the bottom-right grid which is exactly the most rewarding position. Their difference is too small to be observed.

**Q18:**

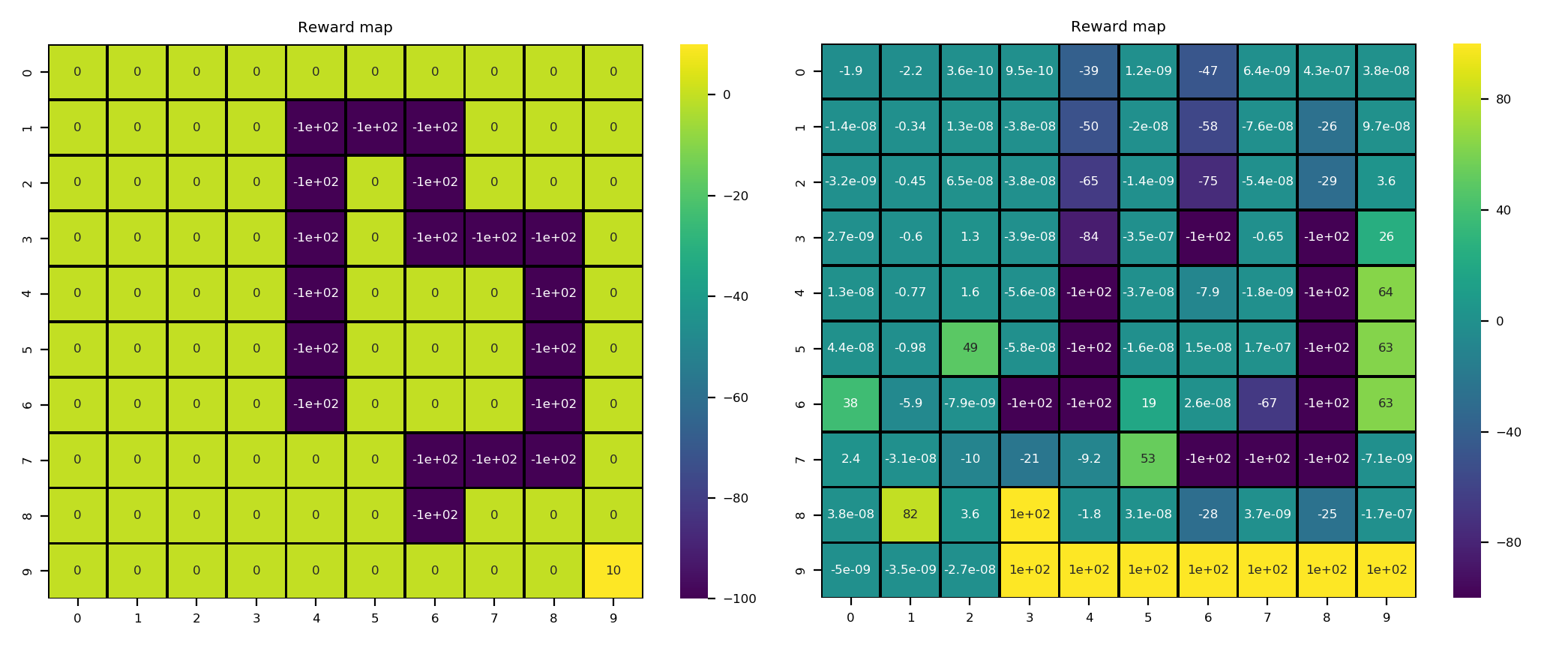


**Figure 6:** Accuracy along with different λ using reward function 1

**Q19:**

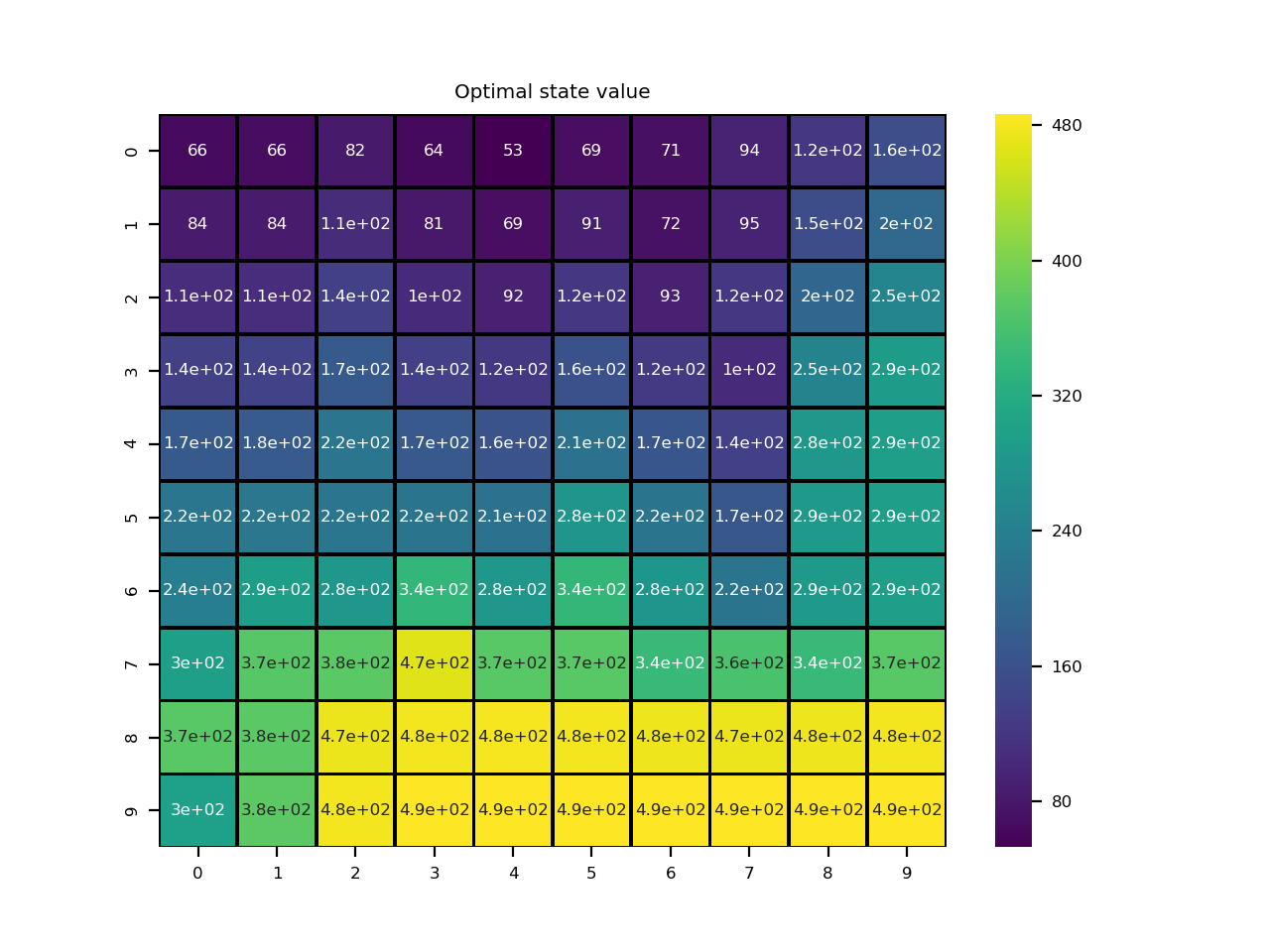


**Q20:**



**Figure 7:** The figure on the left is the heat map of the ground truth reward function 2; the figure on the right is the heat map of the extracted reward function 2.

**Q21:**

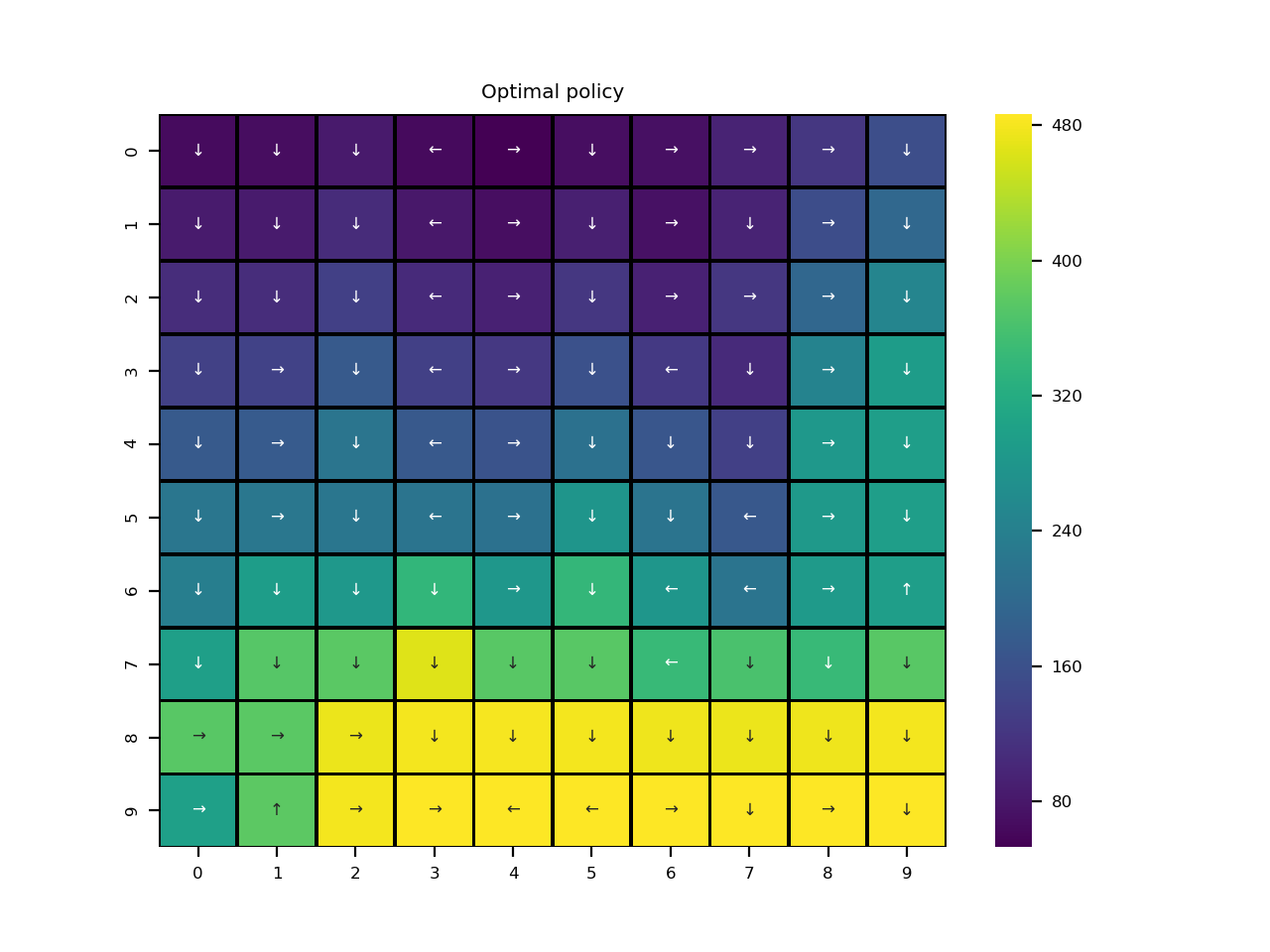


**Figure 8:** Heat map of the optimal state values based on extracted reward function 2

**Q22:**

njnj

**Q23:**



**Figure 9:** Extracted optimal policy based on extracted reward function 2

**Q24:**

njnj

**Q25:**

njnj