50.007 Machine Learning

Simriti Bundhoo 1006281

Homework 2

1

$X = (0.6, 0.8), (0.8, 0.6), (0.8, 0.6)$

Assume a point z where $z = (a, b)$ where a is the x-coordinate and b is the y-coordinate.

(a) Using the formula for Euclidean distance:

$$||x - y||_2 = (\sum_j |x_j - y_j|^2)^{1/2}$$

$$\begin{aligned} d(x, z) &= \sum_{x \in X} d(x, z) \\ &= \sqrt{(0.6 - a)^2 + (0.8 - b)^2 + (0.8 - a)^2 + (0.6 - b)^2 + (-0.8 - a)^2 + (0.6 - b)^2} \end{aligned}$$

$$\frac{\partial f}{\partial a} = \frac{6a - 1.2}{2\sqrt{3a^2 - 1.2a + 3b^2 + 3 - 4b}} = 0$$

$$\frac{\partial f}{\partial b} = \frac{3b - 2}{\sqrt{3a^2 - 1.2a + 3b^2 + 3 - 4b}} = 0$$

$6a - 1.2 = 0$
$a = 0.2$

$3b - 2 = 0$
$b = 0.666666 \approx 0.67$

$z = (0.2, 0.67)$

(b) Using the formula for the squared Euclidean distance:

$$||x - y||_2 = (\sum_j |x_j - y_j|^2)$$

$$\begin{aligned} d(x, z) &= \sum_{x \in X} d(x, z) \\ &= (0.6 - a)^2 + (0.8 - b)^2 + (0.8 - a)^2 + (0.6 - b)^2 + (-0.8 - a)^2 + (0.6 - b)^2 \\ &= 3a^2 - 1.2a + 3b^2 + 3 - 4b \end{aligned}$$

$$\frac{\partial f}{\partial a} = 6a - 1.2 = 0$$

$$\frac{\partial f}{\partial b} = 3b - 2 = 0$$

$a = 0.2$
$b = 0.666666 \approx 0.67$
$z = (0.2, 0.67)$

(c) Using the formula for the Manhattan distance:

$$||x - y||_1 = \sum_j |x_j - y_j|$$

$$\begin{aligned} d(x,z) &= \sum_{x \in X} d(x,z) \\ &= |(0.6 - a) + (0.8 - b) + (0.8 - a) + (0.6 - b) + (-0.8 - a) + (0.6 - b)| \\ &= | - 3a - 3b + 2.6| \end{aligned}$$

As the partial derivatives of both a and b have no solution, there are no critical points where the minimum or maximum can occur within the range of the variables. This means that the minimum or maximum must occur at the boundary.

Since there are no specific constraints mentioned, the minimum or maximum may be unbounded.

$$z = (a \in \mathbb{R}, b \in \mathbb{R})$$

The following plots (from Wolfram Alpha) may be used to determine values a,b.

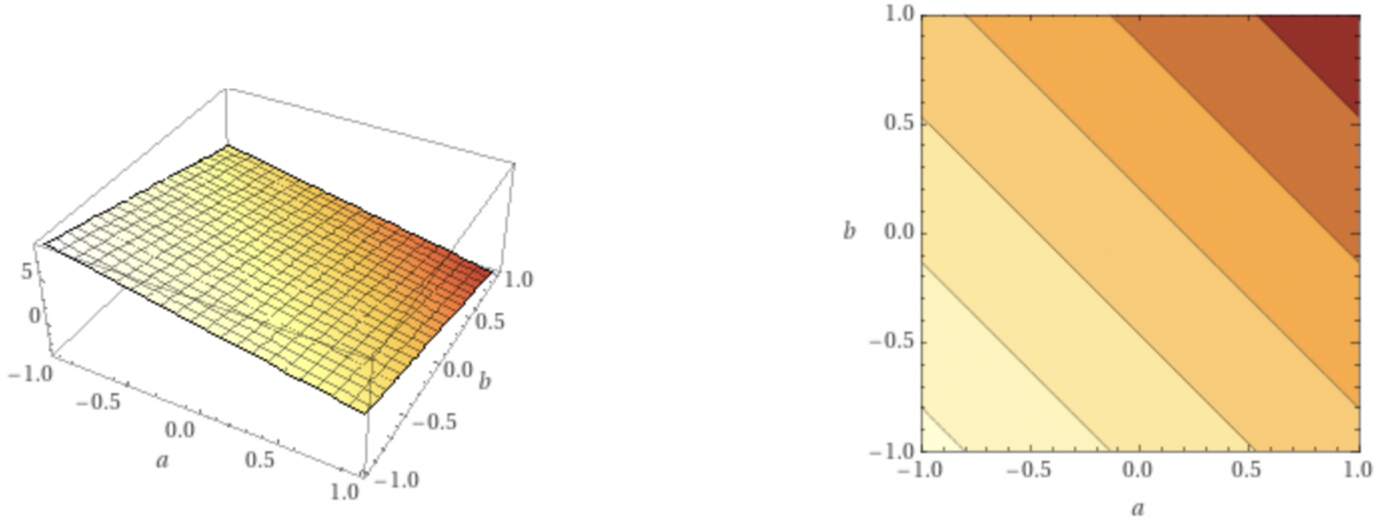

Figure 1: 3D Plot and Contour Plot

2

Refer to the attached .ipynb file for the code and the report (it is generated upon running the code).

3

(a)

	s = Uncertain	s = Certain	s = Lose	s = Win
A	0.1	1.0	0.2	2.0
G	-1.6	0.8	-1.2	2.0

Using the Q-value iteration equation:

$$Q^*_{i+1}(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma max_{a'} Q^*_i(s',a')]$$

As this is the very first iteration, $Q_0(s,a) = 0$ and $R(s,a,s') = R(s')$, therefore the equation becomes:

$$Q^*_{i+1}(s,a) = \sum T(s,a,s') \times R(s')$$

$$\begin{aligned} Q(Uncertain, A) &= (0.9 \times 0.0) + (0.1 \times 1.0) \\ &= 0.1 \end{aligned}$$

$$\begin{aligned} Q(Uncertain, G) &= (0.9 \times -2.0) + (0.1 \times 2.0) \\ &= -1.6 \end{aligned}$$

$$\begin{aligned} Q(Certain, A) &= 1.0 \times 1.0 \\ &= 1.0 \end{aligned}$$

$$\begin{aligned} Q(Certain, G) &= (0.3 \times -2.0) + (0.7 \times 2.0) \\ &= 0.8 \end{aligned}$$

$$\begin{aligned} Q(Lose, A) &= (0.8 \times 0.0) + (0.2 \times 1.0) \\ &= 0.2 \end{aligned}$$

$$\begin{aligned} Q(Lose, G) &= (0.8 \times -2.0) + (0.2 \times 2.0) \\ &= -1.2 \end{aligned}$$

$$\begin{aligned} Q(Win, A) &= (1.0 \times 2.0) \\ &= 2.0 \end{aligned}$$

$$\begin{aligned} Q(Win, G) &= (1.0 \times 2.0) \\ &= 2.0 \end{aligned}$$

(b)

	s = Uncertain	s = Certain	s = Lose	s = Win
$\pi^*(s)$	(Uncertain, A)	(Certain, A)	(Lose, A)	(Win, G)

Assume action G is taken for draws. Using the policy equation:

$$pi^*(s) = argmax_a Q^*(s,a)$$

$pi^*(Uncertain) = (Uncertain, A)$

$pi^*(Certain) = (Certain, A)$

$pi^*(Lose) = (Lose, A)$

$pi^*(Win) = (Win, G)$

(c)

	s = Uncertain	s = Certain	s = Lose	s = Win
V_1^* (s)	0.1	1.0	0.2	2.0

Using the optimal value equation:

$$V^*(s) = argmax_a Q^*(s,a)$$

$V^*(Uncertain) = 0.1$ $\quad$ $V^*(Lose) = 1.2$

$V^*(Certain) = 1.0$ $\quad$ $V^*(Win) = 2.0$

(d)

	s = Uncertain	s = Certain	s = Lose	s = Win
A	0.15	1.5	0.3	3.0
G	-1.55	1.3	-1.1	3.0

Using the Q-value iteration equation:

$$Q^*_{i+1}(s,a) = \sum_{s'} T(s,a,s')[R(s') + \gamma V^*(s')]$$

$$\begin{aligned} Q(Uncertain, A) &= 0.9 \times (0.0 + (0.5 \times 0.1)) + 0.1 \times (1.0 + (0.5 \times 0.1)) \\ &= 0.15 \end{aligned}$$

$$\begin{aligned} Q(Certain, A) &= 1.0 \times (1.0 + (0.5 \times 1.0)) \\ &= 1.5 \end{aligned}$$

$$\begin{aligned} Q(Lose, A) &= 0.8 \times (0.0 + (0.5 \times 0.2)) + 0.2 \times (1.0 + (0.5 \times 0.2)) \\ &= 0.3 \end{aligned}$$

$$\begin{aligned} Q(Win, A) &= 1.0 \times (2.0 + (0.5 \times 2.0)) \\ &= 3.0 \end{aligned}$$

$$\begin{aligned} Q(Uncertain, G) &= 0.9 \times (-2.0 + (0.5 \times 0.1)) + 0.1 \times (2.0 + (0.5 \times 0.1)) \\ &= -1.55 \end{aligned}$$

$$\begin{aligned} Q(Certain, G) &= 0.3 \times (-2.0 + (0.5 \times 1.0)) + 0.7 \times (2.0 + (0.5 \times 1.0)) \\ &= 1.3 \end{aligned}$$

$$\begin{aligned} Q(Lose, G) &= 0.8 \times (-2.0 + (0.5 \times 0.2)) + 0.2 \times (2.0 + (0.5 \times 0.2)) \\ &= -1.1 \end{aligned}$$

$$\begin{aligned} Q(Win, G) &= 1.0 \times (2.0 + (0.5 \times 2.0)) \\ &= 3.0 \end{aligned}$$

(e)

	s = Uncertain	s = Certain	s = Lose	s = Win
$\pi^*(s)$	(Uncertain, A)	(Certain, A)	(Lose, A)	(Win, G)

$pi^*(Uncertain) = (Uncertain, A)$ $\quad$ $pi^*(Lose) = (Lose, A)$

$pi^*(Certain) = (Certain, A)$ $\quad$ $pi^*(Win) = (Win, G)$

(f)

	s = Uncertain	s = Certain	s = Lose	s = Win
V_2^* (s)	0.15	1.5	0.3	3.0

$V^*(Uncertain) = 0.15$ $V^*(Lose) = 0.3$

$V^*(Certain) = 1.5$ $V^*(Win) = 3.0$

4

(a) Using the equation for policy iteration:

$$V^{\pi_i}_{k+1} = \sum_{s'} T(s, \pi_i(s), s')[R(s, \pi_i(s), s') + \gamma V^{\pi_i}_k(s')]$$

As $V^{\pi_i}_k(A) = 0.000$, the equation becomes:

$$V^{\pi_i}_{k+1} = \sum_{s'} T(s, a, s')R(s, a, s')$$

$$\begin{aligned} V^{\pi_i}_{k+1}(A) &= (0.8 \times 0.0) + (0.2 \times 2.0) + (0.4 \times 1.0) + (0.6 \times 0.0) \\ &= 0.8 \end{aligned}$$

(b) Using the following equation:

$$Q^{\pi_i}_{\infty}(s, a) = \sum_{s'} T(s, a, s')[R(s, a, s') + \gamma V^{\pi}_{\infty}(s')]$$

$$\begin{aligned} Q^{\pi_i}_{\infty}(A, clockwise) &= 0.8 \times (0.0 + (0.5 \times -0.203)) + 0.2 \times (2.0 + (0.5 \times -0.203)) \\ &= 0.2985 \end{aligned}$$

(c)

$$\begin{aligned} Q^{\pi_i}_{\infty}(A, counterclockwise) &= 0.4 \times (1.0 + (0.5 \times -0.203)) + 0.6 \times (0.0 + (0.5 \times -0.203)) \\ &= 0.2985 \end{aligned}$$

(d) As the Q-values from part(a) and part(b) are the same, it indicates that there is no significant difference in the expected returns among the actions. Therefore, any of the actions can be chosen as part of the updated policy.

5

Refer to the attached .ipynb file for the code. The final Q-values are printed at the end of the code.