Homework 7 Solution

Deep Learning

The reward matrix R and the initial Q are

$$R = \begin{bmatrix} -1 & -1 & -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 & -1 & 100 \\ -1 & -1 & -1 & 0 & -1 & -1 \\ -1 & 0 & 0 & -1 & 0 & -1 \\ 0 & -1 & -1 & 0 & -1 & 100 \\ -1 & 0 & -1 & -1 & 0 & 100 \end{bmatrix}, \qquad Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Episode 1: At room 1, go to room 5 (10 points)

$$\begin{aligned} Q(1,5) &= R(1,5) + 0.8 \max\{Q(5,1), Q(5,4), Q(5,5)\} \\ &= 100 + 0.8 \max\{0,0,0\} \\ &= 100 \ . \end{aligned}$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Episode 2: At room 2, go to room 3 (10 points)

$$\begin{aligned} Q(2,3) &= R(2,3) + 0.8 \max\{Q(3,1), Q(3,2), Q(3,4)\} \\ &= 0 + 0.8 \max\{0,0,0\} \\ &= 0 \ . \end{aligned}$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Episode 3: At room 3, go to room 4 (10 points)

$$\begin{aligned} Q(3,4) &= R(3,4) + 0.8 \max\{Q(4,0), Q(4,3), Q(4,5)\} \\ &= 0 + 0.8 \max\{0,0,0\} \\ &= 0 \ . \end{aligned}$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Episode 4: At room 4, go to room 0 (10 points)

$$\begin{aligned} Q(4,0) &= R(4,0) + 0.8 \max\{Q(0,4)\} \\ &= 0 + 0.8 \max\{0\} \\ &= 0 \ . \end{aligned}$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Episode 5: At room 5, go to room 1 (10 points)

$$\begin{aligned} Q(5,1) &= R(5,1) + 0.8 \max\{Q(1,3), Q(1,5)\} \\ &= 0 + 0.8 \max\{0,100\} \\ &= 80 \ . \end{aligned}$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 80 & 0 & 0 & 0 & 0 \end{bmatrix}$$

7 points for calculating the associated entry of Q at each episode and **3 points** for updating the Q matrix. Using the Bellman equation correctly **3 points**.