PAI exam 2014, Problem 7

Reminder: Q-learning update

$$Q(x,a) \leftarrow (1-\lambda)\, Q(x,a) + \lambda \left[r + \gamma \max_{a'} Q(x',a') \right]$$

(using λ instead of α)

Episode 1

A1, Right, B1, 0

$$Q(A1, Right) \leftarrow (1-\lambda)\, Q(A1, Right) + \lambda \left[r + \gamma \max_{a'} Q(B1, a') \right]$$

$$Q(A1, Right) \leftarrow 0.5 \times 0 + 0.5 \left[0 + 0.5 \times 0 \right]$$

$Q(A', Right) \leftarrow 0$

All $Q(x,a)$ are initialized to 0, and most rewards are 0. Therefore, most transitions will actually not update the Q values. We will only write down the ones that do a non-zero update.

$$Q(C1, Right) \leftarrow 0.5 \times 0 + 0.5 \left[80 + 0.5 \times 0 \right]$$

$$\Rightarrow \boxed{Q(C1, Right) \leftarrow 40}$$

Episode 2

$$Q(D3, Up) \leftarrow 0.5 \times 0 + 0.5 \left[100 + 0.5 \times 0 \right]$$

$$\Rightarrow \boxed{Q(D3, Up) \leftarrow 50}$$

Episode 3

$$Q(B3, Left) \leftarrow 0.5 \times 0 + 0.5[25 + 0.5 \times 0]$$

$$\rightarrow \boxed{Q(B3, Left) \leftarrow \frac{25}{2}}$$

Episode 4

$$Q(B2, Down) \leftarrow 0.5 \times 0 + 0.5[0 + 0.5 \times \overbrace{\max_{a'} Q(B3, a')}^{25/2 \,=\, 12.5}]$$

$$\Rightarrow \boxed{Q(B2, Down) \leftarrow \frac{25}{8}}$$

$$Q(C3, Up) \leftarrow 0.5 \times 0 + 0.5[-80 + 0.5 \times 0]$$

$$\Rightarrow \boxed{Q(C3, Up) \leftarrow -40}$$

Final answers:

$$Q(B1, Right) = 0$$

$$Q(B2, Down) = \frac{25}{8}$$

$$Q(C1, Right) = 40$$

$$Q(C3, Right) = 0$$