78572 - Telma Correia
78958 - Gonçalo Rodrigues

Homework 2

1.a)
We chose to represent each state as Vertical (V), Horizontal (H), Diagonal (D) and Simultaneous (S).

- The first (V) represents that the wolf and the hare are in the same column;
- The second (H) represents that the wolf and the hare are in the same row;
- The third (D) represents that the wolf and the hare are in different rows and columns;
- Finally, the fourth (S) represent that the wolf and the hare are in the same row and column.

We chose this representation because the actions and probabilities do not depend on the actual positions.
We also noticed that the actions “up” and “down” represent the exact same thing, as well as “left” and “right”.

X = {V, H, D, S}
A = {up, down, left, right, stay}

1.b)

Pstay
[0.6 0.0 0.2 0.2]
[0.0 0.6 0.2 0.2]
[0.2 0.2 0.6 0.0]
[0.2 0.2 0.0 0.6]

Pup = Pdown
[0.28 0.16 0.04 0.52]
[0.16 0.28 0.52 0.04]
[0.04 0.52 0.28 0.16]
[0.52 0.04 0.16 0.28]

Pright = Pleft
[0.28 0.16 0.52 0.04]
[0.16 0.28 0.04 0.52]
[0.52 0.04 0.28 0.16]
[0.04 0.52 0.16 0.28]

Cost(x,a) = 0 if x = S; 1 otherwise
That is:
[1 1 1 1 1]
[1 1 1 1 1]
[1 1 1 1 1]
[0 0 0 0 0]

1.c)
Cost-to-go function with π= wolf always goes up and ϒ=0.99

$$J^{\pi}(x) = E_{\pi}[\sum_{t=0}^{inf} \gamma^t C_t | x_0 = x]$$

To compute the state distribution for t = 1, we use xP, x being a row vector representing the initial state and P is Pup (transition matrix for action Up) since for this policy the wolf always chooses to go up. For any given t, the distribution is given by xP^t

As we use Pup for the transition matrix we also use the Cost(x,Up) as a column vector instead of the whole matrix and call it C. The estimated cost of a state distribution is then given by multiplying the distribution by C.

So we have:

$$J^{\pi}(x) = [\sum_{t=0}^{inf} \gamma^t x P^t C]$$

For example, for x = H:

$$J^{\pi}(H) = E_{\pi}[\sum_{t=0}^{inf} \gamma^t C_t | x_0 = H] = 1 + \gamma((0.16 + 0.28 + 0.52) * 1 + 0.04 * 0) + \gamma^2(\ldots) + \ldots$$

$$J^{\pi}(x) = xC + xP^1 C\gamma + \ldots = x(C + P^1 C\gamma + \ldots) = x(I + P^1\gamma + \ldots)C$$

$$Let\ A = (I + P^1\gamma + \ldots)\ so:$$

$$J^{\pi}(x) = xAC$$

$$A = (I + P^1\gamma + P^2\gamma^2 + \ldots) = I + \gamma P(I + P^1\gamma + \ldots) = I + \gamma PA$$

$$A = I + \gamma PA \Leftrightarrow A - \gamma PA = I \Leftrightarrow (I - \gamma P)A = I \Leftrightarrow A = (I - \gamma P)^{-1}$$

$$J^{\pi}(x) = x(I - \gamma P)^{-1} C$$

J = [74.79200157]
[75.65490583]
[75.57662126]
[73.97647134]
where row 1 is state V, row 2 is H, row 3 is D and row 4 is S