TPC2

1.

a) The dragon and the princess do not move, the knight can move to six different positions, thus, there are 6 different states:

$$X = \{1, 2, 3, 4, 5, 6\}$$

The action space for the MDP is the knight move to LEFT, RIGHT, UP and DOWN:

$$A = \{LEFT, RIGHT, UP, DOWN\}$$

b) The transition probabilities for each action are:

Probability for LEFT action:

Pl:

	0	1	2	3	4	5
0	0.8	0.1	0	0.1	0	0
1	0.6	0.2	0.1	0	0.1	0
2	0	0.6	0.3	0	0	0.1
3	0.1	0	0	0.8	0.1	0
4	0	0.1	0	0.6	0.2	0.1
5	0	0	0.1	0	0.6	0.3

Probability for RIGHT action:

Pr:

	0	1	2	3	4	5
0	0.3	0.6	0	0.1	0	0
1	0.1	0.2	0.6	0	0.1	0
2	0	0.1	0.8	0	0	0.1
3	0.1	0	0	0.3	0.6	0
4	0	0.1	0	0.1	0.2	0.6
5	0	0	0.1	0	0.1	0.8

Probability for UP action:

Pu:

	0	1	2	3	4	5
0	0.8	0.1	0	0.1	0	0
1	0.1	0.7	0.1	0	0.1	0
2	0	0.1	0.8	0	0	0.1
3	0.6	0	0	0.3	0.1	0
4	0	0.6	0	0.1	0.2	0.1
5	0	0	0.6	0	0.1	0.3

Probability for DOWN action:

Pd:

	0	1	2	3	4	5
0	0.3	0.1	0	0.6	0	0
1	0.1	0.2	0.1	0	0.6	0
2	0	0.1	0.3	0	0	0.6
3	0.1	0	0	0.8	0.1	0
4	0	0.1	0	0.1	0.7	0.1
5	0	0	0.1	0	0.1	0.8

For the cost function we assume that the cost for the states where there's no dragon or princess is 0.1, when reaching the state where there is a dragon, the knight receives a penalty, of cost 1, and no cost at all at the state where princess is. Resulting in:

C:

	0	1	2	3
0	0.1	0.1	0.1	0.1
1	0.1	0.1	0.1	0.1
2	0.1	0.1	0.1	0.1
3	0.1	0.1	0.1	0.1
4	1	1	1	1
5	0	0	0	0

c) To compute the cost-to-go function associated with the policy which the knight always do the action UP, we solve the linear system:

$$\boldsymbol{J}^{\pi} = \boldsymbol{c}_{\pi} + \gamma \mathbf{P}_{\pi} \boldsymbol{J}^{\pi}$$

Where $\pi = UP$.

Solving the equation:

$$\boldsymbol{J}^{\pi} = (\mathbf{I} - \gamma \mathbf{P}_{\pi})^{-1} \boldsymbol{c}_{\pi}$$

Resulting:

```
[1.28062328 1.48282118 1.22285691 1.39022903 2.48145131 1.21051144]
```

Grupo nº 37

João Silvestre, nº 80996

João Pestana, nº 79515