# Solutions to the exercises on reinforcement learning and dialogue system evaluation

# (IN4080 course)

## Question 1

1.1



##

1.2



## Question 2

We need to compute two Q-values:

 $Q\left(HumanNotEngaged, SayHi\right)$

$$=R\left(HumanNotEngaged, SayHi\right)+0.9 \left(0.5\*max\_{a\_{1}}Q\left(HumanNotEngaged, a\_{1}\right)+0.5\*max\_{a\_{2}}Q\left(HumanEngaged, a\_{2}\right)\right)$$

$$=-1+0.9 \left(0.5\*max\_{a\_{1}}Q\left(HumanNotEngaged, a\_{1}\right)+0.5\*10\right)$$

$$=3.5+0.45\* max\_{a\_{1}}Q\left(HumanNotEngaged, a\_{1}\right)$$

$$Q\left(HumanNotEngaged, SayHiWithGestures\right)$$

$$=R\left(HumanNotEngaged, SayHiWithGestures\right)+0.9 \left(0.3\*max\_{a\_{1}}Q\left(HumanNotEngaged, a\_{1}\right)+0.7\*max\_{a\_{2}}Q\left(HumanEngaged, a\_{2}\right)\right)$$

$$=-2+0.9 \left(0.3\*max\_{a\_{1}}Q\left(HumanNotEngaged, a\_{1}\right)+0.7\*10\right)$$

$$=4.3+0.27\* max\_{a\_{1}}Q\left(HumanNotEngaged, a\_{1}\right)$$

So, for the first iteration, starting with Q-values initialized to 0, we get:

$$Q\_{t\_{1}}\left(HumanNotEngaged, SayHi\right)=3.5$$

$$Q\_{t\_{1}}\left(HumanNotEngaged, SayHiWithGestures\right)=4.3$$

For the second iteration, we get:

$$Q\_{t\_{2}}\left(HumanNotEngaged, SayHi\right)=3.5+0.45\*4.3=5.44 $$

$$Q\_{t\_{2}}\left(HumanNotEngaged, SayHiWithGestures\right)=4.3+0.27\*4.3=5.46 $$

We continue:

$$Q\_{t\_{3}}\left(HumanNotEngaged, SayHi\right)=3.5+0.45\*5.46=5.96 $$

$$Q\_{t\_{3}}\left(HumanNotEngaged, SayHiWithGestures\right)=4.3+0.27\*5.46=5.77$$

$$Q\_{t\_{4}}\left(HumanNotEngaged, SayHi\right)=3.5+0.45\*5.96=6.18 $$

$$Q\_{t\_{4}}\left(HumanNotEngaged, SayHiWithGestures\right)=4.3+0.27\*5.96=5.91 $$

$$Q\_{t\_{5}}\left(HumanNotEngaged, SayHi\right)=3.5+0.45\*6.18=6.28 $$

$$Q\_{t\_{5}}\left(HumanNotEngaged, SayHiWithGestures\right)=4.3+0.27\*6.18=5.97$$

So the correct answer at the end of iteration fine would be 6.28 for Q(HumanNotEngaged, SayHi) and 5.97 for Q(HumanNotEngaged, SayHiWithGestures).