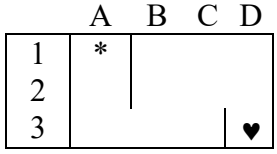
Consider Q-learning of the following classical maze (15 points)



The object is to get from the asterix to the heart. States are maze positions (e.g., [1,A]). Actions are Down (**v**), Up (^), Right (>) and Left (<). Assume you start with the Q values listed in the table below.

	1,A	**2,A**	**3,A**	**1,B**	**2,B**	**3,B**	**1,C**	**2,C**	**3,C**	**1,D**	**2,D**	**3,D**
V	.2	.2	0	0	0	0	.1	0	0	.1	.1	0
^	0	0	0	0	.1	.1	0	0	0	0	0	0
>	0	0	.2	.1	.2	.2	.1	.2	0	0	0	0
<	0	0	0	0	0	0	0	0	0	0	0	0

Transition probabilities are deterministic and based solely on the maze. Reward is 0 in all states but moving to [3,D] gets you a reward of 1. Assume $\gamma = 0.9$

Using Q-learning and the policy indicated by the above table, move the asterix through **ten** time steps. Assume that each episode starts at 1,A. Please provide the following information in your answer booklets. If a chosen move is random, indicate with an "*"

Step	*Action taken*	*s'* (subsequent state)	*Q(s,a) to update and value*
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			