1.MDP

	Trajectory 1	Trajectory 2	Trajectory 3	Trajectory 4	Trajectory 5	Trajectory 6
S_0	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]	[0, 0]
V	-2.0	-2.32	-2.47	-3.77	-4.89	-5.09
a_0	UP	UP	RIGHT	RIGHT	UP	RIGHT
S_1	[1, 0]	[1, 0]	[0, 1]	[0, 1]	[1, 0]	[0, 1]
V	-2.0	-3.6	-2.0	-3.6	-4.88	-4.07
a_1	UP	UP	RIGHT	RIGHT	UP	RIGHT
S_2	[2, 0]	[2, 0]	[0, 2]	[0, 2]	[2, 0]	[0, 2]
V	-2.0	-3.6	-2.0	-2.34	-4.88	-3.70
a_2	RIGHT	RIGHT	UP	RIGHT	RIGHT	UP
S_3	[2, 1]	[2, 1]	[1, 2]	[0, 3]	[2, 1]	[1, 2]
V	-2.0	-3.6	-2.28	-2.08	-3.54	-3.06
a_3	RIGHT	RIGHT	RIGHT	UP	RIGHT	UP
S_4	[2, 2]	[2, 2]	[1,3]	[1,3]	[2, 2]	[2, 2]
V	-2.0	-1.52	-1.0	-1.0	-1.66	-1.74
a_4	RIGHT	RIGHT			RIGHT	RIGHT
S_5	[2, 3]	[2, 3]			[2, 3]	[2, 3]
V	+1.0	+1.0			+1.0	+1.0

2.RL

RL Exercise

Observed Data	Weights after seeing data $W_{AD}=1$ $W_{AS}=W_{BD}=W_{BS}=0$
Initial Sensors = D=0, S=2 Action: A Reward = -2 Final Sensors: D=1, S=0	let $S=(D=0, S=2)$ $S'=(D=1, S=0)$ So $Q(s,A) = W_{AD} \times D + W_{AS} \times S = 1\times0 + 0\times2 = 0$ ⇒ $Q(s',A)=W_{AD}\times D + W_{AS}\times S = 1\times1 + 0\times0 = 1$ $Q(s',B)= W_{BD}\times f_{BD} + W_{BS}\times f_{BS} = 0\times1 + 0\times0 = 0$ $\max(Q(s',A), Q(s'B))=1$ $\Delta = (-2+1\times1) - Q(s,a)$ update $= -2-1-0 = -1$ $W_{AD} \leftarrow 1+0.5\times(-1)\times0$, $W_{AD}=1$ $W_{AD}=1$ $W_{AS} = 0 + 0.5\times(-1)\times2 = -1$, $W_{BD}=0$, $W_{BS}=0$
Initial Sensor: D=1, S=0 Action: B Reward = 0 Final Sensors: D=1, S=0	let $s=(D=1, S=0)$ $s'=(D=1, S=0)$ So $Q(s,B) = W_{BD}\times f_{BD} + W_{BS}\times f_{BS} = 0\times1 + 0\times0 = 0$ $Q(s',A)= W_{AD}\times f_{AD} + W_{AS}\times f_{AS} = 1\times1 + (-1)\times0 = 1$ $Q(s',B)= W_{BD}\times f_{BD} + W_{BS}\times f_{BS} = 0\times1 + 0\times0 = 0$ ⇒ $\max Q(s',a')=1$, $\Delta = (0+1\times1)-0 = 1$ update $W_{BD} = 0+0.5\times1\times1 = 0.5$, $W_{BS}=0+0.5\times1\times0=0$ $W_{AD}=1$, $W_{AS}=-1$
$S=(D=1, S=1)$	$Q(s,A)= W_{AD}\times f_{AD} + W_{AS}\times f_{AS} = 1\times1 + (-1)\times1=0$ # So action B would $Q(s,B)= W_{BD}\times f_{BD} + W_{BS}\times f_{BS} = 0.5\times1 + 0\times1 = 0.5$ be preferred