COMP-424: Artificial intelligence

Homework 4 Solutions

Answer 1: Hidden Markov Models

Part a:

High	Low
0.6	0.4

Transition	Low	High
Low	0.4	0.6
High	0.7	0.3

Emission	Buy	Sell	Keep
Low	0.5	0.2	0.3
High	0.1	0.6	0.3

Part b:

	Buy	Keep	Sell
High	0.6*0.1=0.06	(0.06*0.3 + 0.2*0.6)*0.3 = 0.0414	(0.0414*0.3 + 0.0366*0.6)*0.6 = 0.020628
Low	0.4*0.5=0.2	(0.2*0.4 + 0.06*0.7)*0.3 = 0.0366	(0.0366*0.4 + 0.0414*0.7)*0.2 = 0.008724

0.020628 + 0.008724 = 0.029352

Part c:

P(High | High) P(High | Buy, Keep, Sell) + P(High | Low) P(Low | Buy, Keep, Sell)
= 0.3 * (0.020628/0.029352) + 0.6 * (0.008724/0.029352)
= 0.389

Part d:

	Buy	Keep	Sell
High	0.6*0.1=0.06	**Max** (0.06*0.3 + **0.2*0.6**) *0.3 = 0.036	**Max** (0.036*0.3 + **0.024*0.6**) *0.6 = **0.00864**
Low	0.4*0.5=0.2	**Max** (**0.2*0.4** + 0.06*0.7) *0.3 = 0.024	**Max** (0.024*0.4 + **0.036*0.7**) *0.2 = 0.00504

Low -> Low -> High

Answer 2: Utility

Part a
P(A) = 0.1*0.8*0.9 + 0.2*0.8*0.1 + 0.3*0.2*0.9 + 0.5*0.2*0.1 = 0.152
EU(with insurance) = -200 * 0.152 - 10 * 0.848 = -38.88
EU(without insurance) = -400 * 0.152 - 0 * 0.848 = -60.8
Price = -38.88 + 60.8 = 21.92

Part b
Pr(A=1 | S=1,D=0) = 0.3
Pr(A=0 | S=1,D=0) = 0.7

EU(with insurance) = -200*0.3– 10*0.7 = -67
EU(without insurance) = -400*0.3– 0*0.7 = -120
Price = 53

Part c
Pr(A=1 | D=1) = 0.2 * 0.8 + 0.5 * 0.2 = 0.26
Pr(A=0 | D=1) = 0.74

EU(with insurance) = -200 * 0.26– 10*0.74= -59.4
EU(without insurance) = -400 * 0.26– 0*0.74= -104
Price = 44.6

Part d
EU(with new insurance) = -200*0.152*0.8 - 400*0.152*0.2 – 10*0.848 = -44.96
Price = -44.96 + 60.8 = 15.84

Answer 3: Markov Decision Processes

Part a
Number of policies: 4^6

Part b
Solve system of linear equations to get,
V(S1)=154.19, V(S2)=175.6, V(S3)=200, V(S4)=-64.9, V(S5)=-87.8, V(S6) = -100

Part c
π (S1)=Right, π (S2)=Right, π (S3)=Down, π (S4)=Down, π (S5)=Up, π (S6) = Up

Part d

V(S1)= 25.92/0.1681=154.19, V(S2)=72/0.41=175.6, V(S3)=200, V(S4)= 10.1717 /0.068921 = 147.58, V(S5)=25.92/0.1681=154.19, V(S6) = 67/0.41 = 163.4

Part e
Yes, Optimal value function represents the maximum attainable values at each state

$$V^*(s) = \max_a E_{\pi^*}\{\sum_{k=0}^{\infty} \gamma^k r_{t+1} | s_t = s, a_t = a\}$$

Thus it has to be unique as long as the policy is optimal.

Part f
π (S1)=Right, π (S2)=Right, π (S3)=Down, π (S4)=Right, π (S5)=Up, π (S6) = Up

Part g
The optimal policies are not unique,
π (S1)=Right, π (S2)=Right, π (S3)=Right, π (S4)=Right, π (S5)=Up, π (S6) = Up
In fact by letting go the assumption of alphabetic order you can get more optimal policies.

Part h
Change reward for S2 to anything between 0 < R < 20, as long as direction of reward difference remains the same as the optimum policy.
You could also scale the rewards by the same amount.

Question 4: Bandits

Part a

Q(t=0)	Q(t=1)	Q(t=2)	Q(t=3)	Q(t=4)	Q(t=5)
1	1	1	1	0	0
1	1	1	1	1	1
3	2	2	1.5	1.5	1
1	1	1	1	1	1
1	1	0	0	0	0

Part b
T=1 - No

T=2 - Yes
T=3 - No
T=4 - Yes
T=5 - No