Assignment 3 Temporal Models - Reinforcement Learning

Fatima AlSaadeh
fatima.alsaadeh@rutgers.edu
fya7

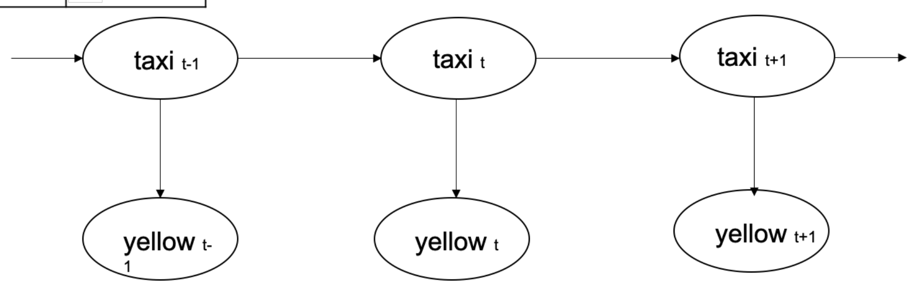
Ashley Dunn
ashley.dunn@rutgers.edu
apd109

May 8, 2020

1 Problem 1

1)

$taxi_{t-1}$	$P(taxi_t)$
t	.25
f	.5



$taxi_{t-1}$	$P(Yellow_t)$
t	.75
f	.1

P(taxi)	P(not taxi)
.4	.6

with transition model conditional probability :

	$taxi_{t+1}$	Not $taxi_{t+1}$
$taxi_t$	.25	.75
not $taxi_t$	.5	.5

and evidence model conditional probability :

	yellow t	not yellow t
taxi t	.75	.25
not taxi t	.1	.9

2)

$$P\left(X1 = taxi|\, e1 = yellow\right) =$$

$$\frac{P\left(e1 = yellow|\, x1 = taxi\right)\sum_{xt-1} P\left(Xt = taxi|\, xt-1\right) * P\left(Xt-1|\, e0 : t-1\right)}{\sum_{xt} P\left(et = yellow|\, xt\right) * P\left(Xt|\, Xt-1\right) * P\left(Xt-1|\, e0 : t-1\right)}$$

$$P\left(e1 = yellow|\, x1 = taxi\right) = .75$$

$$\sum_{xt-1} P\left(Xt = taxi|\, xt-1\right) * P\left(Xt-1|\, e0 : t-1\right) =$$

$$.25 * .4 + .5 * .6 = .4$$

$$\sum_{xt} P\left(et = yellow|\, xt\right) * P\left(Xt|\, Xt-1\right) * P\left(Xt-1|\, e0 : t-1\right)$$

$$(.75 * (.25 * .4 + .5 * .6)) + (.1 * (.75 * .4 + .5 * .6)) = .36$$

$$P\left(X1 = taxi|\, e1 = yellow\right) = \frac{.75 * .4}{.36} = .833$$

$$P\left(X1 = nottaxi|\, e1 = yellow\right) = 1 - .833 = .167$$

3)

$$P(X2 = taxi) = \sum_{xt} P\left(Xt+1 = taxi|\, xt\right) * P\left(Xt|\, et = yellow\right) =$$

$$P\left(X2 = taxi|\, x1 = taxi\right) * P\left(X1 = taxi|\, e1 = yellow\right) +$$

$$P\left(X2 = taxi|\, x1 = nottaxi\right) * P\left(X1 = nottaxi|\, e1 = yellow\right)$$

$$((.25 * .833)) + (.5 * .167)) = .29175$$

4)

$$P\left(X1 = taxi|\, e2 = yellow\right) =$$

$$\frac{P\left(e2 : 2 = yellow|\, X1\right) P\left(X1|\, e1 = yellow\right)}{\sum_{X1} P\left(e2 : 2 = yellow|\, X1\right) * P\left(X1|\, e1 : 1\right)}$$

$$P\left(e2 : 2 = yellow|\, X1\right) =$$

$$\sum_{x2} P\left(X2|\, x1 = taxi\right) * P\left(e2 = yellow|\, X2\right) =$$

$$.25 * .75 + .75 * .1 = .2625$$

$$P\left(X1 = taxi|\, e1 = yellow\right) = .833$$

$$\sum_{x2} P\left(X2|\, x1 = nottaxi\right) * P\left(e2 = yellow|\, X2\right) =$$

$$.5 * .75 + .5 * .1 = .425$$

$$\sum_{X1} P\left(e2:2 = yellow|\, X1\right) * P\left(X1|\, e1:1\right) =$$

$$.2625 * .833 + .425 * .167 = .2896$$

$$P\left(X1 = taxi|\, e2 = yellow\right) = \frac{.2625 * .833}{.2896} = .7566$$

5)
Watching the cars from above we know the possibility of two taxi or two yellow cars followed by each other is not as high as knowing that only the current car is taxi and yellow, which means having a future observation updated our current belief on the probability that the car at time 1 is a taxi.

6)
First as we find that for t=0 the probability of not taxi is higher than a taxi with a value of .6 we pick the state of not taxi for time 0 having the prior distribution.
next step for time t=1 :
P(taxi at t-1)*$P(taxi|taxi) * P(yellow|taxi)$= .4*.25*.75 = 0.075
P(not taxi at t-1)*$P(taxi|nottaxi) * P(yellow|taxi)$= .6*.5*.75 = 0.225
max(0.075,0.225) = 0.225

P(taxi at t-1)*$P(taxi|nottaxi) * P(yellow|nottaxi)$= .4*.75*.1 = 0.03
P(not taxi at t-1)*$P(nottaxi|nottaxi) * P(yellow|nottaxi)$= .6*.5* .1 = 0.03
max(.03,.03) = .03
max(0.225,.03) = 0.225
having the probability for taxi higher so t1 in the sequence is taxi

next step for time t=2 using previous calculation :
P(taxi at t-1)$*P(taxi|taxi) * P(yellow|taxi) = 0.225 * .25 * .75 = 0.042$
P(taxi at t-1)*$P(nottaxi|taxi) * P(yellow|nottaxi)$= 0.225*.75*.1 = 0.0168
max(0.042,.0168) = 0.042

having the probability for taxi higher so t2 in the sequence is taxi
so the most likely sequence of cars is not taxi, taxi , taxi

	t=0	t=1	t=2
taxi	0.4	0.225	0.042
not taxi	0.6	0.03	0.0168

2 Problem 2

1)
having t=1 and GPS=(8,6)

$$Zx \sim \mathcal{N}(8,\,3)$$

$$Zy \sim \mathcal{N}(6,\,3)$$

$$\mu 1 = \frac{(\sigma 0^2 + \sigma x^2) * Z1 + \sigma z^2 * (\mu 0 + v)}{\sigma 0^2 + \sigma x^2 + \sigma z^2}$$

$$\sigma 1^2 = \frac{(\sigma 0^2 + \sigma x^2) * \sigma z^2}{\sigma 0^2 + \sigma x^2 + \sigma z^2}$$

$$\mu 1x = \frac{(.1 + 1) * 8 + 3 * (0 + 10)}{.1 + 1 + 3} = 9.463$$

$$\sigma 1x^2 = \frac{(.1 + 1) * 3}{.1 + 1 + 3} = .8048$$

$$\mu 1y = \frac{(.1 + 1) * 6 + 3 * (0 + 5)}{.1 + 1 + 3} = 5.268$$

$$\sigma 1y^2 = \frac{(.1 + 1) * 3}{.1 + 1 + 3} = .8048$$

$$X1 \sim \mathcal{N}(9.463,\,.8048)$$

$$Y1 \sim \mathcal{N}(5.268,\,.8048)$$

2)
having t=2 and GPS=(19,9) ((.8048+1)*19)+(3*(9.463+10))/(.8048+1+3)

$$X2x \sim \mathcal{N}(X1 + 10,\,1)$$

$$X2x \sim \mathcal{N}(19.463,\,1)$$

$$Y2x \sim \mathcal{N}(Y1 + 5,\,1)$$

$$Y2x \sim \mathcal{N}(10.268,\,1)$$

$$Zx \sim \mathcal{N}(19.463,\,3)$$

$$Zy \sim \mathcal{N}(10.268,\,3)$$

$$\mu 2x = \frac{(.8048 + 1) * 19 + 3 * (9.463 + 10)}{.8048 + 1 + 3} = 19.289$$

$$\sigma 2x^2 = \frac{(.8048 + 1) * 3}{.8048 + 1 + 3} = 1.126$$

$$\mu 2y = \frac{(.8048 + 1) * 9 + 3 * (5.268 + 5)}{.8048 + 1 + 3} = 9.791$$

$$\sigma 2y^2 = \frac{(.8048 + 1) * 3}{.8048 + 1 + 3} = 1.126$$

$$X2 \sim \mathcal{N}(19.28,\,1.126)$$

$$Y2 \sim \mathcal{N}(9.791,\,1.126)$$

3 Problem 3

Part 1; K = 0

Values

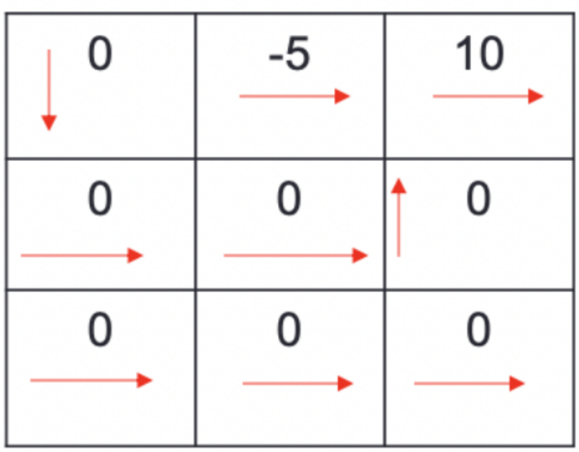
Position\Action	N	E	S	W	$V_{k+1}(s)$ [max]
0, 0	0	0 + .9 * 0 + .1 * 0 = 0	0 + .9 * 0 + .1 * 0 = 0	0	0
0, 1	-5	-5 + .9 * 0 + .05 * 0 + .05 * 0 = -5	-5 + .9 * 0 + .05 * 0 + .05 * 0 = -5	-5 + .9 * 0 + .05 * 0 + .05 * 0 = -5	-5
0, 2	10	10	10 + .9 * 0 + .1 * 0 = 10	10 + .9 * 0 + .1 * 0 = 10	10
1, 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0	0
1, 1	0 + .9 * 0 + .1/3 * 0 + .1/3 * 0 + .1/3 * 0 = 0	0 + .9 * 0 + .1/3 * 0 + .1/3 * 0 + .1/3 * 0 = 0	0 + .9 * 0 + .1/3 * 0 + .1/3 * 0 + .1/3 * 0 = 0	0 + .9 * 0 + .1/3 * 0 + .1/3 * 0 + .1/3 * 0 = 0	0
1, 2	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0
2, 0	0 + .9 * 0 + .1 * 0 = 0	0 + .9 * 0 + .1 * 0 = 0	0	0	0
2, 1	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0
2, 2	0 + .9 * 0 + .1 * 0 = 0	0	0	0 + .9 * 0 + .1 * 0 = 0	0

Policy

(when there is a tie, pick in order: east, south, north, west)

Position	π*
0, 0	south
0, 1	east
0, 2	do nothing
1, 0	east
1, 1	east
1, 2	north
2, 0	east
2, 1	east
2, 2	east

$\pi 1$



K = 1

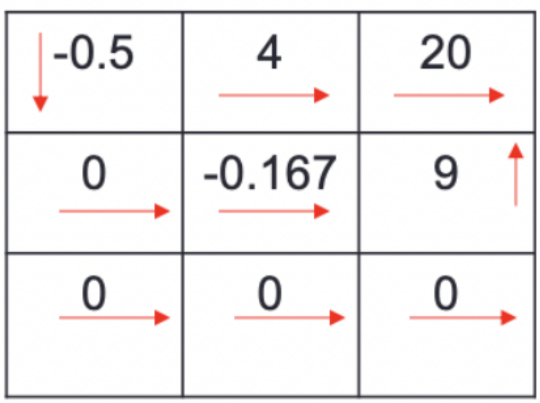
Position\Action	N	E	S	W	$V_{k+1}(s)$ [max]
0, 0	0, not allowed	0 + .9 * -5 + .1 * 0 = -4.5	0 + .9 * 0 + .1 * -5 = -0.5	0, not allowed	-0.5
0, 1	-5	-5 + .9 * 10 + .05 * 0 + .05 * 0 = 4	-5 + .9 * 0 + .05 * 10 + .05 * 0 = -4.5	-5 + .9 * 0 + .05 * 0 + .05 * 10 = -4.5	4
0, 2	10 + 1 * 10 = 20	10 + 1 * 10 = 20	10 + 1 * 10 = 20	10 + 1 * 10 = 20	20
1, 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0	0
1, 1	0 + .9 * -5 + .1/3 * 0 + .1/3 * 0 + .1/3 * 0 = -4.5	0 + .9 * 0 + .1/3 * -5 + .1/3 * 0 + .1/3 * 0 = -0.167	0 + .9 * 0 + .1/3 * -5 + .1/3 * 0 + .1/3 * 0 = -0.167	0 + .9 * 0 + .1/3 * -5 + .1/3 * 0 + .1/3 * 0 = -0.167	-0.167
1, 2	0 + .9 * 10 + .05 * 0 + .05 * 0 = 9	0	0 + .9 * 0 + .05 * 0 + .05 * 10 = 0.5	0 + .9 * 0 + .05 * 0 + .05 * 10 = 0.5	9
2, 0	0 + .9 * 0 + .1 * 0 = 0	0 + .9 * 0 + .1 * 0 = 0	0	0	0
2, 1	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0	0 + .9 * 0 + .05 * 0 + .05 * 0 = 0	0
2, 2	0 + .9 * 0 + .1 * 0 = 0	0	0	0 + .9 * 0 + .1 * 0 = 0	0

Policy

(when there is a tie, pick in order: east, south, north, west)

Position	π*
0, 0	South
0, 1	East
0, 2	do Nothing
1, 0	East
1, 1	East
1, 2	North
2, 0	East
2, 1	East
2, 2	East

π2



Part 2; t = 0, k = 0

Policy Evaluation

		$T(s, \pi_t(s), s') * V_k(s')$, where s' is __ of s:					
s	$R(s, \pi_t(s))$	N	E	S	W	s' = s	$V_{k+1}(s)$
0, 0	0	0	.9 * 0	.1 * 0	0	0	0
0, 1	-5	0	.9 * 0	.05 * 0	.05 * 0	0	-5
0, 2	10	0	0	0	0	1 * 0	10
1, 0	0	.05 * 0	.9 * 0	.05 * 0	0	0	0
1, 1	0	.1/3 * 0	.9 * 0	.1/3 * 0	.1/3 * 0	0	0
1, 2	0	0	0	0	0	1 * 0	0
2, 0	0	.1 * 0	.9 * 0	0	0	0	0
2, 1	0	.05 * 0	.9 * 0	0	.05 * 0	0	0
2, 2	0	0	0	0	0	0	10

k=1

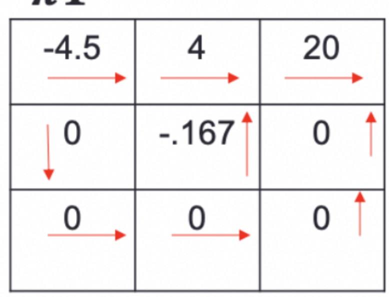
		$T(s, \pi_t(s), s') * V_k(s')$, where s' is __ of s:					
s	$R(s, \pi_t(s))$	N	E	S	W	s' = s	$V_{k+1}(s)$
0, 0	0	0	.9 * -5	.1 * 0	0	0	-4.5
0, 1	-5	0	.9 * 10	.05 * 0	.05 * 0	0	4
0, 2	10	0	0	0	0	1 * 10	20
1, 0	0	.05 * 0	.9 * 0	.05 * 0	0	0	0
1, 1	0	.1/3 * -5	.9 * 0	.1/3 * 0	.1/3 * 0	0	-0.167
1, 2	0	0	0	0	0	1 * 0	0
2, 0	0	.1 * 0	.9 * 0	0	0	0	0
2, 1	0	.05 * 0	.9 * 0	0	.05 * 0	0	0
2, 2	0	0	0	0	0	0	0

Policy improvement:

(when there is a tie, pick in order: east, south, north, west)

Position	π_{t+1}
0, 0	East
0, 1	East
0, 2	East, do nothing
1, 0	South
1, 1	North
1, 2	North
2, 0	East
2, 1	East
2, 2	North

$\pi 1$



t = 1, k = 2

Policy Evaluation (rounds to 3 decimal spaces)

		$T(s, \pi_t(s), s') * V_k(s')$, where s' is __ of s:					
s	$R(s, \pi_t(s))$	N	E	S	W	s' = s	$V_{k+1}(s)$
0, 0	0	0	.9 * 4	.1* 0	0	0	3.6
0, 1	-5	0	.9 * 20	.05 * -0.167	.05 * -4.5	0	12.76
0, 2	10	0	0	0	0	1 * 20	30
1, 0	0	.05 * -4.5	.05 * -0.167	.9 * 0	0	0	-0.233
1, 1	0	.9 * 4	.1/3 * 0	.1/3 * 0	.1/3 * 0	0	3.6
1, 2	0	.9 * 20	0	.05 * 0	.05 * -0.167	0	17.992
2, 0	0	.1 * 0	.9 * 0	0	0	0	0
2, 1	0	.05 * -0.167	.9 * 0	0	.05 * 0	0	-0.00835
2, 2	0	0	0	0	0	0	0

t = 1, k = 3

		$T(s, \pi_t(s), s') * V_k(s')$, where s' is __ of s:					
s	$R(s, \pi_t(s))$	N	E	S	W	s' = s	$V_{k+1}(s)$
0, 0	0	0	.9 * 12.76	.1 * -0.233	0	0	11.461
0, 1	-5	0	.9 * 30	.05 * 3.6	.05 * 3.6	0	22.36
0, 2	10	0	0	0	0	1 * 30	40
1, 0	0	.05 * 3.6	.05 * 3.6	.9 * 0	0	0	0.36
1, 1	0	.9 * 12.76	.1/3 * 17.992	.1/3* -0.00835	.1/3 * -0.233	0	12.075
1, 2	0	.9 * 30	0	.05 * 0	.05 * 3.6	0	27.18
2, 0	0	.1 * -0.233	.9 * -0.00835	0	0	0	-0.0308
2, 1	0	.05 * 3.6	.9 * 0	0	.05 * 0	0	0.18
2, 2	0	0.9 * 17.992	0	0	0.1* -0.00835	0	16.191

Policy improvement:

Position	π*
0, 0	East
0, 1	East
0, 2	East, do nothing
1, 0	East
1, 1	East
1, 2	North
2, 0	North
2, 1	East
2, 2	North

$\pi 2$