PCS5024 - Statistical Machine Learning

Escola Politecnica, USP

Anna H. Reali Costa, Fabio G. Cozman

Studant : Angel Felipe M. de Paula
USP : 11030561

Exercicio de Reinforcement Learning

May 6, 2019

PCS5024 - Statistical Machine Learning
Escola Politecnica, USP
Anna H. Reali Costa, Fabio G. Cozman

Estudante: Angel Felipe Magnossão de Paula
USP : 11030561

Exercicio 1: Indicar uma história com 12 passos e sua respectiva sequência de comandos (uma só) e respectiva probabilidade:

In [ ]:
```
Resp: Utilizando a mesma sequência de comandos contida nos slides
"[U ,U, R, R, R]" é impossível executar 12 passos. Tendo o mesmo
"Trasition Model" também será impossível obter a mesma probabilidade
obtidas no exemplo dado em classe. Porém, Tendo uma sequência de
no mínimo 12 paços, sim é possível chegar no "goal" (12 passos):

S: (1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2),
(3,3), (3,4), (4,1), (4,2), (4,3), (4,4)

A: [L, R, U, D]

h(x): [L, L, L, L, L, L, L, U, U, R, R, R]

p(x): (0,8)^12 =  0.068719476736

V(x): (-0.04*11) + 1= 0.56
```

Exercicio 2: Resolver usando o algorithm "valion interation":
a)Quantas interações ?
b)Quais os valores ?

In [ ]:
```
### Informarion:
Gama = 1/2
A: red, green
S: s1, s2
|A|= Duas ações and |S|= Dois estados
V0(s1)= 0 and V0(s2)= 0
```

```
### Calculation:
--------------------------------------------------------------------------
1.A)
V1(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V0(s2) + p(s1|s1,red)*V0(s1)),
                       (gama*(p(s2|s1,green)*V0(s2)))

V1(s1)= 3 + max(a) ( (0.5*(0.5*0 + 0.5*0)), (0.5*(1*0)) )
V1(s1)= 3 + max(a) ( (0.5*(0 + 0)), (0.5*(0)) )
V1(s1)= 3 + max(a) ( (0), (0)) )
V1(s1)= 3 + 0
V1(s1)= 3

1.B)
V1(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V0(s2)),
                       (gama*p(s1|s2,green)*V0(s1))))

V1(s2)= -1 + max(a) ( (0.5*(0.5*0)), (0.5*(1*0)) )
V1(s2)= -1 + max(a) ( (0.5*(0)), (0.5*(0)) )
V1(s2)= -1 + max(a) ( (0), (0) )
V1(s2)= -1 + 0
V1(s2)= -1
--------------------------------------------------------------------------
2.A)
V2(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V1(s2) + p(s1|s1,red)*V1(s1)),
                       (gama*(p(s2|s1,green)*V1(s2)))

V2(s1)= 3 + max(a) ( (0.5*(0.5*3 + 0.5*(-1))), (0.5*(1*(-1))) )
V2(s1)= 3 + max(a) ( (0.5*(1.5 + (-0.5)), (0.5*(-1)) )
V2(s1)= 3 + max(a) ( (0.5*(1), (-0.5) )
V2(s1)= 3 + max(a) ( (0.5), (-0.5) )
V2(s1)= 3 + 0.5
V2(s1)= 3.5

2.B)
V2(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V1(s2)),
                       (gama*p(s1|s2,green)*V1(s1))))

V2(s2)= -1 + max(a) ( (0.5*(0.5*(-1))), (0.5*(1*3)) )
V2(s2)= -1 + max(a) ( (0.5*(-0.5)), (0.5*3) )
V2(s2)= -1 + max(a) ( (0.25), (1.5) )
V2(s2)= -1 + 1.5
V2(s2)= 0.5
----------------------------------------------------------------------------
```

```
3.A)
V3(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V2(s2) + p(s1|s1,red)*V2(s1)),
                       (gama*(p(s2|s1,green)*V2(s2)))

V3(s1)= 3 + max(a) ( (0.5*(0.5*0.5 + 0.5*3.5), (0.5*(1*0.5) )
V3(s1)= 3 + max(a) ( (0.5*(0.25 + 1.75), (0.5*0.5) )
V3(s1)= 3 + max(a) ( (0.5*2), (0.25) )
V3(s1)= 3 + max(a) ( (1), (0.25) )
V3(s1)= 3 + 1
V3(s1)= 4

3.B)
V3(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V2(s2)),
                       (gama*p(s1|s2,green)*V2(s1)))

V3(s2)= -1 + max(a) ( (0.5*(0.5*0.5))), (0.5*(1*3.5)) )
V3(s2)= -1 + max(a) ( (0.5*0.25)), (0.5*3.5) )
V3(s2)= -1 + max(a) ( (0.125), (1.75) )
V3(s2)= -1 + 1.75
V3(s2)=  0.75
----------------------------------------------------------------------------
4.A)
V4(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V3(s2) + p(s1|s1,red)*V3(s1)),
                       (gama*(p(s2|s1,green)*V3(s2)))

V4(s1)= 3 + max(a) ( (0.5*(0.5*0.75 + 0.5*4), (0.5*(1*0.75) )

V4(s1)= 3 + max(a) ( (0.5*(0.375 + 2), (0.5*0.75) )
V4(s1)= 3 + max(a) ( (0.5*(2.375 ), (0.5*0.75) )
V4(s1)= 3 + max(a) ( (1.1875), (0.375) )
V4(s1)= 3 + 1.1875
V4(s1)= 4.1875 ~ 4.2

4.B)
V4(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V3(s2)),
                       (gama*p(s1|s2,green)*V3(s1)))

V4(s2)= -1 + max(a) ( (0.5*(0.5*0.75)), (0.5*(1*4)) )
V4(s2)= -1 + max(a) ( (0.5*0.375), (0.5*4) )
V4(s2)= -1 + max(a) ( (0.1875), (2) )
V4(s2)= -1 + 2
V4(s2)=  1
-----------------------------------------------------------------------------
```

```
5.A)
V5(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V4(s2) + p(s1|s1,red)*V4(s1)),
                       (gama*(p(s2|s1,green)*V4(s2)))

V5(s1) = 3 + max(a)((0.5*(0.5*1 + 0.5*4.2), (0.5*(1*1) )
V5(s1) = 3 + max(a)((0.5*(0.5 + 2.1), (0.5*1) )
V5(s1) = 3 + max(a)((0.5*2.6), (0.5) )
V5(s1) = 3 + max(a)((1.3), (0.5) )
V5(s1) = 3 + 1.3
V5(s1) = 4.3

5.B)
V5(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V4(s2)),
                       (gama*p(s1|s2,green)*V4(s1)))

V5(s2)= -1 + max(a) ( (0.5*(0.5*1)), (0.5*(1*4.2)) )
V5(s2)= -1 + max(a) ( (0.25), (2.1) )
V5(s2)= -1 + 2.1
V5(s2)=  1.1
-----------------------------------------------------------------------------
6.A)
V6(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V5(s2) + p(s1|s1,red)*V5(s1)),
                       (gama*(p(s2|s1,green)*V5(s2)))

V6(s1) = 3 + max(a)((0.5*(0.5*1.1 + 0.5*4.3)), (0.5*(1*1.1) )
V6(s1) = 3 + max(a)((0.5*(0.55 + 2.15)), (0.5*1.1) )
V6(s1) = 3 + max(a)((0.5*2.7), (0.55) )
V6(s1) = 3 + max(a)((1.35), (0.55) )
V6(s1) = 3 + 1.35
V6(s1) = 4.35 ~ 4.4

6.B)
V6(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V5(s2)),
                       (gama*p(s1|s2,green)*V5(s1)))

V6(s2)= -1 + max(a) ( (0.5*(0.5*1.1)), (0.5*(1*4.3)) )
V6(s2)= -1 + max(a) ( (0.5*0.55), (0.5*4.3) )
V6(s2)= -1 + max(a) ( (0.275), (2,15) )
V6(s2)= -1 + 2,15
V6(s2)=  1,15 ~ 1.2
-----------------------------------------------------------------------------
```

```
7.A)
V7(s1) = r(s1) + max(a)(gama*(p(s2|s1,red)*V6(s2) + p(s1|s1,red)*V6(s1)),
                       (gama*(p(s2|s1,green)*V6(s2)))

V7(s1) = 3 + max(a)((0.5*(0.5*1.2 + 0.5*4.4)), (0.5*(1*1.2) )
V7(s1) = 3 + max(a)((0.5*(0.6 + 2.2)), (0.5*1.2) )
V7(s1) = 3 + max(a)((0.5*2.8), (0.6) )
V7(s1) = 3 + max(a)((1.4), (0.6) )
V7(s1) = 3 + 1.4
V7(s1) = 4.4

7.B)
V7(s2) = r(s2) + max(a)(gama*(p(s2|s2,red)*V6(s2)),
                       (gama*p(s1|s2,green)*V6(s1)))

V7(s2)= -1 + max(a) ( (0.5*(0.5*1.2)), (0.5*(1*4.4)) )
V7(s2)= -1 + max(a) ( (0.5*0.6), (0.5*4.4) )
V7(s2)= -1 + max(a) ( (0.3), (2.2) )
V7(s2)= -1 + 2.2
V7(s2)=  1.2
-------------------------------------------------------------------------------
Criterio de parada: Vi(s)=Vi+1(s)
V6(s1)=V7(s1)
V6(s2)=V7(s2)

a)Número de intereacões: 14
b)Valores:
     v(s1) v(s2)
    |  0  |  0  |
    |  3  | -1  |
    | 3.5 | 0.5 |
    | 4.0 | 0.75|
    | 4.2 | 1.0 |
    | 4.3 | 1.1 |
    | 4.1 | 1.2 |
```

Exercício 3: Mostrar a melhor função de valor possível (V*(s)) ea melhor políticas possível (p*(s)) utilizando o algoritmo de Value Iteration:

Referência:

	1	2	3	4
3	(3,1)	(3,2)	(3,3)	(3,4)
2	(2,1)		(2,3)	(2,4)
1	(1,1)	(1,2)	(1,3)	(1,4)

Funcao de Valor:

0)

	1	2	3	4
3	0	0	0	100
2	0		0	-100
1	0	0	0	0

1)

	1	2	3	4
3	-1.00	-1.00	71.00	100
2	-1.00		-10.00	-100
1	-1.00	-1.00	-1.00	-1.00

2)

	1	2	3	4
3	-1.90	49.94	76.49	100
2	-1.90		40.22	-100
1	-1.90	-1.90	-1.90	-1.90

3)

	1	2	3	4
3	34.61	63.06	81.50	100
2	-2.71		48.69	-100
1	-2.71	-2.71	27.62	-2.71

4)

	1	2	3	4
3	47.28	69.03	82.72	100
2	23.43		53.07	-100
1	-3.44	18.40	33.57	9.64

5)

	1	2	3	4
3	55.07	70.98	83.22	100
2	37.26		54.33	-100
1	17.22	26.48	39.73	15.04

6)

	1	2	3	4
3	58.42	71.70	83.38	100
2	45.36		54.81	-100
1	29.76	32.37	41.86	19.96

7)

	1	2	3	4
3	59.96	71.94	83.44	100
2	49.22		54.97	-100
1	37.25	34.96	43.17	21.93

8)

	1	2	3	4
3	60.62	72.02	83.46	100
2	51.03		55.02	-100
1	40.94	36.38	43.70	23.06

9)

	1	2	3	4
3	60.91	72.05	83.46	100
2	51.83		55.04	-100
1	42.70	37.01	43.96	23.54

10)

	1	2	3	4
3	61.02	72.06	83.47	100
2	52.18		55.05	-100
1	43.49	37.32	44.08	23.77

11)

	1	2	3	4
3	61.07	72.07	83.47	100
2	52.33		55.05	-100
1	43.84	37.45	44.13	23.88

12)

	1	2	3	4
3	61.09	72.07	83.47	100
2	52.39		55.05	-100
1	43.99	37.52	44.16	23.92

13)

	1	2	3	4
3	61.10	72.07	83.47	100
2	52.42		55.05	-100
1	44.06	37.54	44.17	23.94

14)

	1	2	3	4
3	61.11	72.07	83.47	100
2	52.43		55.05	-100
1	44.09	37.56	44.17	23.95

15)

	1	2	3	4
3	61.11	72.07	83.47	100
2	52.43		55.05	-100
1	44.10	37.56	44.17	23.96

16)

	1	2	3	4
3	61.11	72.07	83.47	100
2	52.44		55.05	-100
1	44.10	37.57	44.17	23.96

17)

	1	2	3	4
3	61.11	72.07	83.47	100
2	52.44		55.05	-100
1	44.10	37.57	44.17	23.96

Política:

→	→	→	100
↑		↑	-100
↑	→	↑	←