BY: S. VISHAL
CS6700 - RL - ASSIGNMENT-1 CH18B020

1
a) Action space : { forage patch 1, forage patch 2, forage patch 3 }
(call them as a_1, a_2, a_3 respectively)

State space : { Animal is dead, 2 units of energy in reserve, 3 units of energy in reserve, 4 units of energy in reserve }

We want to maximise probability of survival at the end of 3 time periods.

$$g_3(x_3) = \begin{cases} 1 & \forall\, x_3 \in \{2, 3, 4\} \\ 0 & \text{if } x_3 = \text{dead}. \end{cases}$$

↓
Terminal cost

So we give a terminal cost of 1 if the animal has survived & all other transition costs to be 0. (& we want to maximise it)

i.e. $g_k(i, a, j) = 0 \quad \forall\, \{k, i, a, j\}$

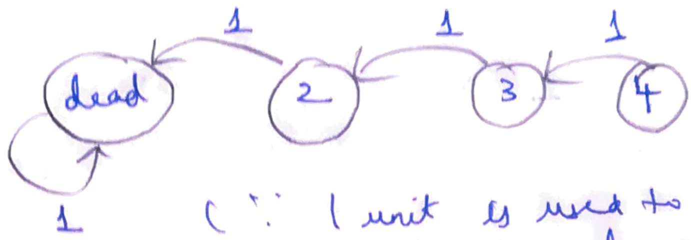
So by maximising J,
$$J = \max E\left(g_3(x_3) + \sum_{k=0}^{2} g_k(x_k, u_k(x_k), x_{k+1})\right)$$ (0)

$= \max E(g_3(x_3))$

we in fact maximise the probability of survival ($= E(g_3(x_3))$)

Let us note the transition probabilities for each action

a_1 : Forage patch 1



(∵ 1 unit is used to forage)

$P_{dd} = 1$
$P_{2d} = 1$
$P_{32} = 1$
$P_{43} = 1$
$P_{ij} = 0$ for all others

a_2 : Forage patch 2

Assume: risk of predation indep of finding food.

∴ risk of predation = 0.004, there is a 0.004 probability that the animal dies

In the 0.996 chance the animal survives, there is 0.3 chance to get food of 2 energy units

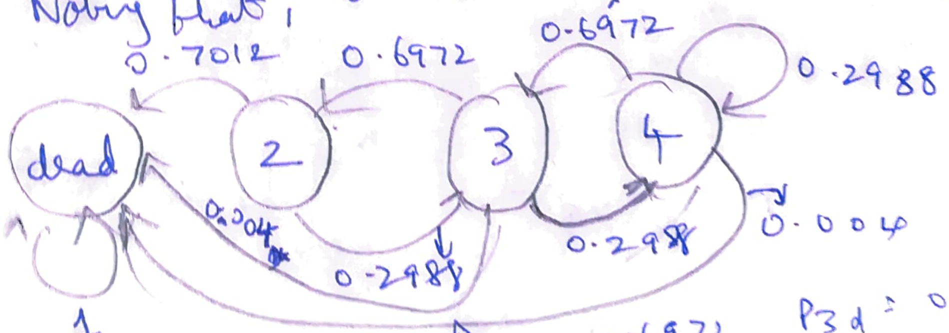
P(food | survival) = 0.3 ⇒ P(food ∩ survival) = 0.3 × P survival

(gaining 1 energy unit as a result - 1 unit will be used to forage)

& 0.7 chance it doesn't get food

(⇒ loses 1 unit as a result of foraging)

Noting that, 0.996 × 0.3 = 0.2988 & 0.996 × 0.7 = 0.6972



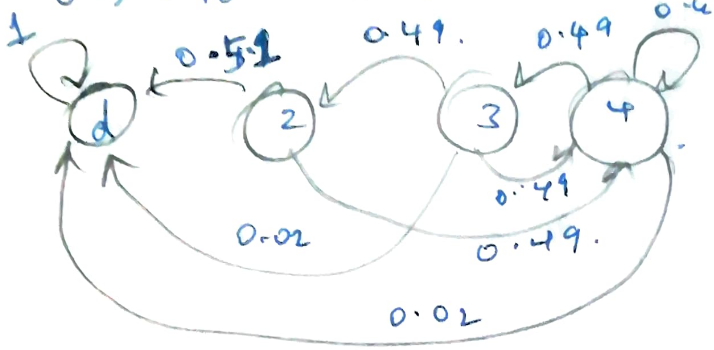
$P_{dd} = 1$
$P_{2d} = 0.7012$
$P_{23} = 0.2988$
$P_{32} = 0.6972$
$P_{34} = 0.2988$
$P_{43} = 0.6972$
$P_{3d} = 0.004$
$P_{4d} = 0.004$
$P_{ij} = 0$ for all other i,j

a3: Forage patch 3.

Here the energy gain 3 units ($\Rightarrow$ net gain of 2 units)

Proceeding similar to the case of a2

$0.5 \times 98 = 0.49.$



$P_{dd} = 1$

$P_{2d} = 0.51$

$P_{24} = 0.49$

$P_{32} = 0.49$

$P_{3d} = 0.02$

$P_{34} = 0.49$

$P_{43} = 0.49$ $\quad P_{44} = 0.49$

$P_{4d} = 0.02$ $\quad P_{ij} = 0 \ \forall$ other i & j

b) $$J_3^*(x_3) = g_3(x_3)$$
$$= \begin{cases} 1 & \forall\, x_3 \in \{2,3,4\} \\ 0 & \text{if } x_3 = d. \end{cases}$$

DP Algorithm:

$$J_2^*(x_2) = \max_a E_{x_3}\Big(g_2(x_2, a_i, x_3) + J_3^*(x_3) \Big)$$

$$= \max_a \sum_{\forall j \in S} P_{ij}(a) \Big(g_2(x_2, a, j) + J_3^*(j) \Big)$$

but $g(i, a, j) = 0$ always

$$\Rightarrow J_2^*(x_2) = \max_a \sum P_{ij}(a)\, J_3^*(j)$$

~~For~~ $x_2 = d \Rightarrow$ action doesn't matter the animal is already dead

action 1: forage patch 1

x_2	$\sum J^* p_{ij}$
d	0
2	$1+0=0$
3	$P_{32} J_3^*(2) = 1$ ($\because P_{3j} = 0 \ \forall\ P_{4j} \neq 2$)
4	$P_{43} J_3^*(3) = 1$

action 2: forage patch 2

x_2	$\sum J^* p_{ij}$
d	0
2	$0.2988\, J_3^*(3) = 0.2988$
3	$0.2988\, J_3^*(4) + 0.6972\, J_3^*(2) = 0.996$
4	$0.2988\, J_3^*(4) + 0.6972\, J_3^*(2) = 0.996$

action 3: forage patch 3

x_2	$\sum J^* p_{ij}$
d	0

x_2	$\sum P_{ij} J_3^*(j)$
2	$P_{24} J^*(4) = 0.49.$
3	$0.49\, J^*(2) + 0.49 J^*(4) = 0.98$
4	$0.49\, J^*(3) + 0.49 J^*(4) = 0.98$

choosing the actions that maximise the ~~probable~~ cost $J_2(x_2)$ we find

x_2	$J_2^*(x_2)$	a
1	0	— (doesn't matter)
2	0.49.	a_3 (forage patch 3)
3	1	a_1 (forage patch 1)
4	1	a_1 (forage patch 1)

Now, $J_1^*(x_1) = \max_a E\left(\cancel{g(x_1, a, j)} + J_2^*(j)\right)$

$= \max_a E\left(J_2^*(j_a) \right)$ ↳ over j

If we chose action a_1

$E(J_2^*(j)) = \sum P_{ij}(a)\, J_2^*(j)$

We make a table & compare $J_1(x_1)$ values for each action and then decide the optimal value

x_1	a_1	a_2	a_3
d	0	0	0
2	0	$0.2988 J_2^*(3) = 0.2988$	$0.49 J_2^*(4) = 0.49$
3	$1 \times J_2^*(2) = 0.49$	$0.2998 J_2^*(4) + 0.6972 J_2^*(2) = 0.6404$	$0.49 J_2^*(4) + 0.49 J_2^*(2) = 0.7301$
4	$1 \times J_2^*(3) = 1$	0.996	$0.49 J_2^*(4) + 0.49 J_2^*(3) = 0.98$

Choosing the action that maximises ~~probability of~~ turn the reward,

x_1	$J_1^*(x_1)$	Action
d	0	– (can't take any action – dead)
2	0.49	Forage patch 3 (a_3)
3	0.7301	Forage patch 3 (a_3)
4	1	Forage patch 1 (a_1)

Similarly, $J_0^*(x_0) = \max_a E(J_1^*(x_1))$

$$= \max_a \sum_{\forall j} P_{ij}(a)(J_1^*(x_1))$$

We once again make a table of action & reward

x_0	J_0^* a_1	a_2	a_3
d	0	0	0
2	0	$0.2988(0.73) = 0.218$	0.49
3	0.49	$0.2988(1) + 0.6972(0.49) = 0.6404$	$0.49 + (0.49)^2 = 0.7301$
4	0.7301	$(0.2998) + (0.6972)(0.73) = 0.8078$	$0.49 + 0.49 \times 0.7301 = 0.8477$

Choosing the action that optimises the reward,

x_0	$J_0^*(x_0)$	Action
d	0	~~For~~ – (dead)
2	0.49	Forage patch 3 (a_3)
3	0.7301	Forage patch 3 (a_3)
4	0.8477	Forage patch 3 (a_3)

Optimal policy for foraging obtained from the above analysis:

$$\pi^* = \{\mu_0(x), \mu_1(x), \mu_2(x)\}$$

$$\mu_0(x) = \begin{cases} \text{can't perform action} & x = d \\ a_3 & x = 2 \quad \text{(forage patch 3)} \\ a_3 & x = 3 \quad \text{(forage patch 3)} \\ a_3 & x = 4 \quad \text{(forage patch 3)} \end{cases}$$

$$M_1(x) = \begin{cases} \text{cant perform action} & x = d \\ a_3 & x = 2 \text{ (Forage patch 3)} \\ a_3 & x = 3 \text{ (Forage patch 3)} \\ a_1 & x = 4 \text{ (Forage patch 1)} \end{cases}$$

$$M_2(x) = \begin{cases} \text{cant perform action} & x = d \\ a_3 & x = 2 \text{ (forage patch 3)} \\ a_1 & x = 3 \text{ (forage patch 1)} \\ a_1 & x = 4 \text{ (forage patch 1)} \end{cases}$$