CS6700 : RL - ASSIGNMENT - 1

BY: S. VISHAL
CH18B020

(4) For convenience, I would like to label Kevin as player A, Karan as player B, Kannan as player C.

To model the problem as an MDP, where 'A' is the decision maker we need to know the decisions (in fact, optimal) of 'B' & 'C'.

2 - Game scenarios.

Consider the game has reduced to A & B and it is A's turn to fire. Let us analyse this from the point of view of B:

- B can't miss intentionally - has to shoot at A.
- Rule permits A to shoot into the air. But that will increase chance of B winning - because B will take more turns to fire & hence the probability of reaching the target (which will be a <u>geometric</u> distribution) increases

So, to obtain a <u>lower bound</u> of probability of B winning we can as well assume A always shoots & doesn't miss intentionally

$P_{B \text{ wins in } k \text{ turns}}$ = P(A misses k turns despite shooting)
× P(B miss k-1 turns consecutively before firing in last turn)

$$= (1-\alpha)^k (1-\beta)^{k-1} \beta.$$

$$\therefore P(B \text{ wins in } A \text{ vs } B) = \sum_{k=1}^{\infty} (1-\alpha)^k (1-\beta)^{k-1} \beta.$$

$$= \frac{(1-\alpha)}{1-(1-\alpha)(1-\beta)} \times \beta$$

Proceeding similarly

$$P(B \text{ wins in } B \text{ vs } C) = \frac{(1-\gamma)}{1-(1-\gamma)(1-\beta)} \times \beta$$

But $(1-\gamma) < (1-\alpha)$ $(\because \alpha > \gamma)$

$\left(\Rightarrow 1-(1-\alpha)(1-\beta) < 1-(1-\gamma)(1-\beta) \right)$

So we conclude

$$P(B \text{ wins } A \text{ vs } B) > P(B \text{ wins } B \text{ vs } C)$$

Similarly for C, arguing in a similar fashion,

$$P(C \text{ wins } C \text{ vs } A) > P(C \text{ wins } C \text{ vs } B)$$

From this analysis, we understand that B and C would seek to eliminate each other first because they prefer duelling A in a 2 game scenario, to maximise winning chance.

So, when all of ABC are alive,

* B shoots C in his turn
* C shoots B in his turn

~~States~~

Now that we have decided the possible actions of B & C we can build an MDP.

∴ Note that a single transition in the MDP is incorporates all changes in the position of the game between 2 turns of A (eg. if B & C are alive, one transition implies A, B, and C have all taken their turns)

State space : { A ~~lost~~ (1), A and B are alive (2), A and C are alive (3), A, B and C are alive (4), A only alive (A has won!) (5) }

Action space : {

state (1) : no action - A already out of the game (terminal)

state (2) : shoot at B, miss intentionally.

state (3) : shoot at C, miss intentionally

state (4) : shoot at B, shoot at C, miss intentionally

state (5) : A has won! - no actions (terminal)

~~costs~~ Rewards :

∵ We want to maximise probability of winning, ~~we will maximise~~ $g(i, a, j) = 0 \; \forall \{i, a, j\}$

$$g_N(x_N) = \begin{cases} 0 & \text{if } x \neq x_5 \\ 1 & \text{if } x = x_5 \end{cases}$$

Note that the A lost & A won are terminal states

& so objective $\max_a E\left(\alpha c + J_N^*(x_N) + \sum_{k=1}^{N} g(\cdot) \right)$ $\sum_{\forall k} g_k(x_k, u_k, x_{k+1})$

$$= \max_a E\left(J_N^*(x_N) \right)$$

$$= \max_a E\left(g_N'(x_N) \right)$$

which is eqvt to maximising probability of winy.

$\rightarrow$ ($\because$ probability = expectation of indicator variable for the event)

<u>Transition probabilities</u>

Let a_1: A shoots B a_2: A shoots C a_3 A shoots air.

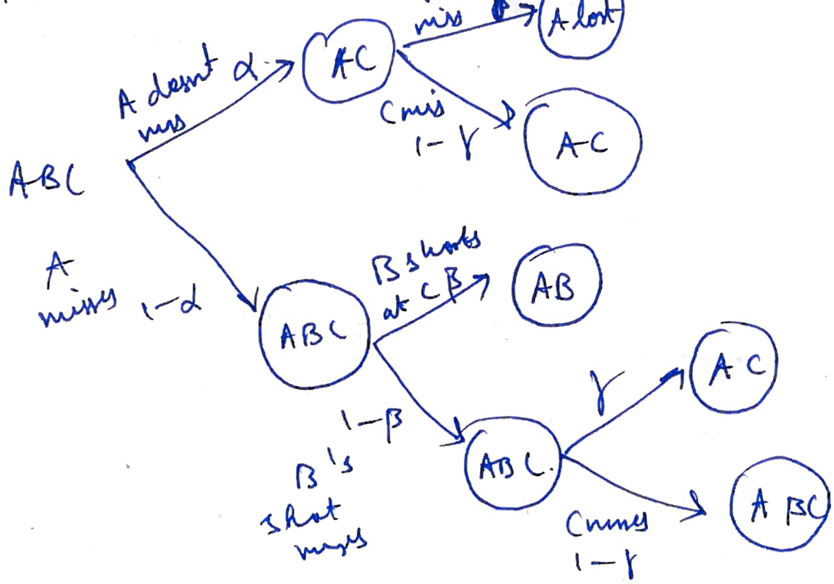
a) Action a_1:

From AB : A wins if A shoots properly : α

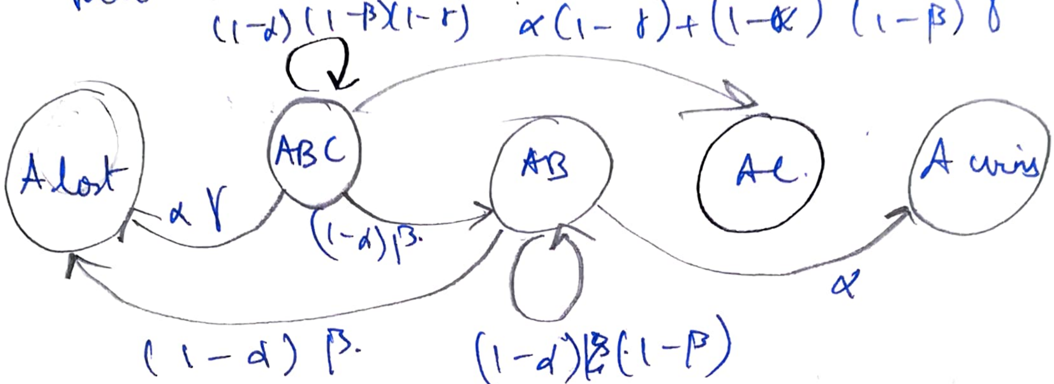
A ~~loses~~ loses if A miss & B shoots properly $(1-\alpha)\beta$

Back to same state if A & B both miss $(1-\alpha)(1-\beta)$

From ABC



Using these transition probabilities we get the markov chain as:



No transitions from AC because A can't shoot B in that ~~state~~ -

$P_{21} = \alpha\gamma$ $\quad P_{22} = (1-\alpha)(1-\beta)(1-\gamma)$ $\quad P_{24} = \alpha(1-\gamma) + (1-\alpha)(1-\beta)\gamma$

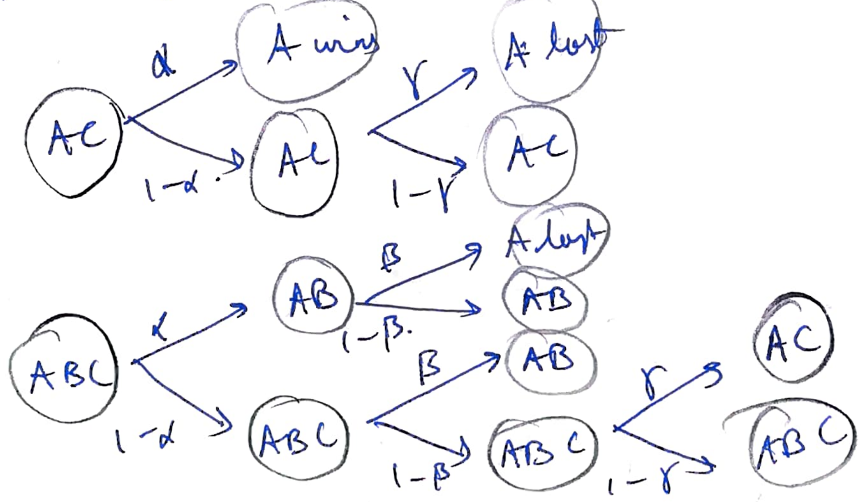
$P_{23} = (1-\alpha)\beta$

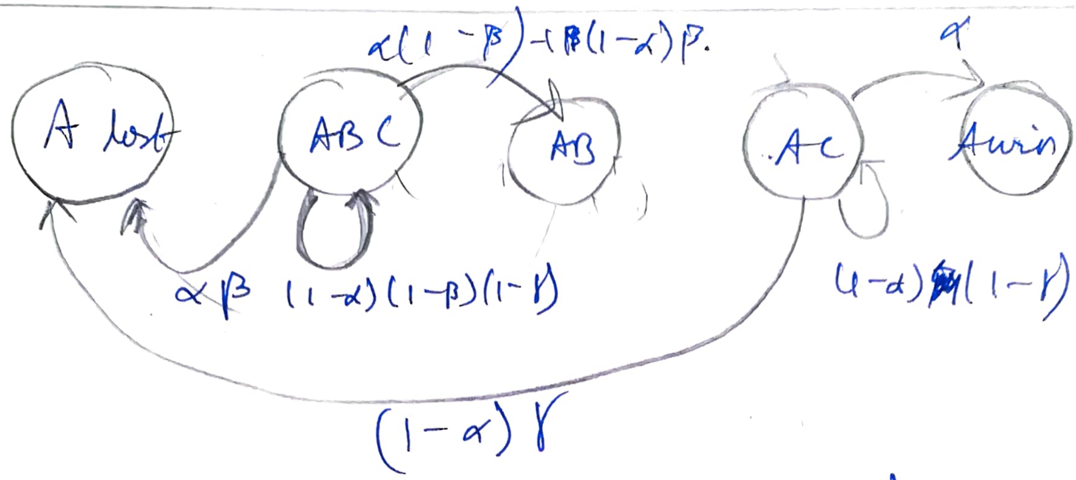
$P_{31} = (1-\alpha)\beta$ $\quad P_{33} = (1-\alpha)(1-\beta)$

$P_{35} = \alpha$; & other $P_{ij} = 0$

ii) Action a_2 : A shoots C

~~proceeding~~ Proceeding similarly,



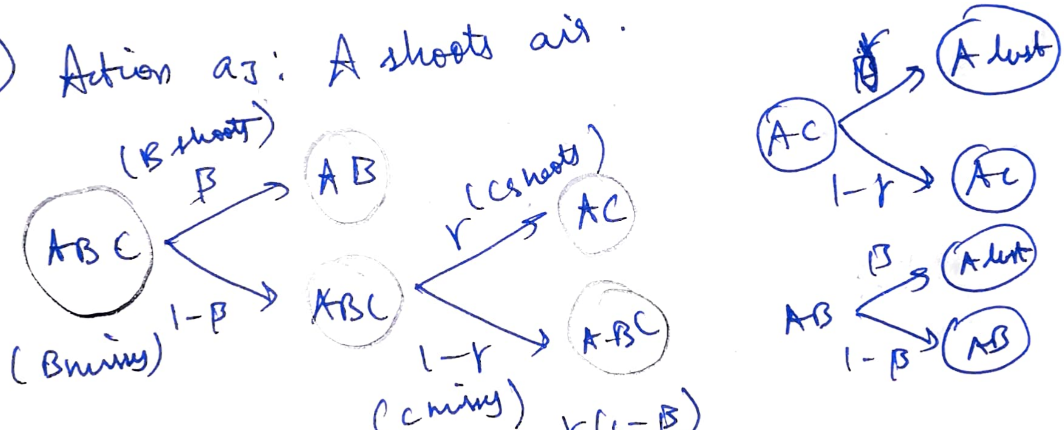


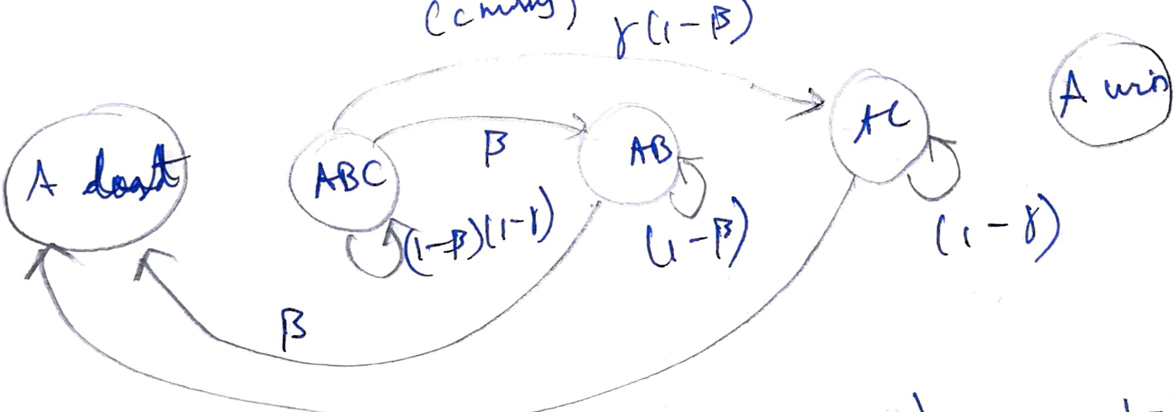
$P_{21} = \alpha\beta \qquad P_{22} = (1-\alpha)(1-\beta)(1-\gamma)$

$P_{23} = \alpha(1-\beta) + (1-\alpha)\beta \; ; \; P_{41} = (1-\alpha)\gamma$

$P_{44} = (1-\alpha)(1-\gamma) \; ; \; P_{45} = \alpha \; ;$ other $P_{ij} = 0$

iii) Action a_3: A shoots air.





$P_{22} = (1-\beta)(1-\gamma) \qquad P_{33} = (1-\beta) \qquad P_{44} = 1-\gamma$

$P_{23} = \beta \qquad P_{24} = (1-\beta)\gamma \qquad P_{31} = \beta \qquad P_{41} = \gamma$

other $P_{ij} = 0$

b) Bellman optimality equations

$$TJ^{+}_{N} = J^{+} \; ; \; J^{+}(A\,win) = 1$$

$$\Rightarrow J^{+}(i) = \max_{a \in A(i)} E\left(\cancel{g(i,a,j)} + J^{+}(j) \right)$$

($\because$ ~~maximising~~ rewards)

$$= \max_{a \in A(i)} E\left(J^{+}(j) \right)$$

~~J(AB) = max E~~

~~At AB state~~ AB, we have 2 actions allowed
- shoot B or shoot air

$$\Rightarrow J(AB) = \max \{ \underbrace{(1-\beta) J(AB)}_{\text{A shoots air}}, \; \underbrace{(1-\alpha)(1-\beta) J(AB) + \alpha \times J(A\,win)}_{\text{A shoots B}} \}$$

$$= \max \{ (1-\beta) J(AB), \; (1-\alpha)(1-\beta) J(AB) + \alpha \} \quad (1)$$

Similarly

$$J(AC) = \max \{ (1-\gamma) J(AC), \; (1-\alpha)(1-\gamma) J(AC) + \alpha \} \quad (2)$$

A shoots air

A shoots C

{2 actions : shoot air or shoot C }

At state ABC, all 3 actions are allowed, so we take max over the 3 action

$$J(ABC) = \max \Big\{ \Big[(1-\alpha)(1-\beta)(1-\gamma) J(ABC) + (\alpha(1-\gamma) + (1-\alpha)(1-\beta)\gamma) J(AC) + (1-\alpha)\beta J_{AB}\Big],$$

$$\Big[(1-\alpha)(1-\beta)(1-\gamma) J(ABC) + (\alpha(1-\beta) + (1-\alpha)\beta) J(AB) + (1-\alpha)(1-\beta)\gamma J(AC)\Big],$$

$$\Big[(1-\beta)(1-\gamma) J(ABC) + \beta J(AB) + \gamma(1-\beta) J(AC)\Big]\Big\} \quad \text{③}$$

c) Given $\alpha = 0.3$, $\beta = 0.5$, $\gamma = 0.6$. Substituting,

eqn ① : $J_{AB} = \max\left(0.5 J_{AB},\ \underset{= 0.35 J_{AB} + 0.3}{0.5 \times 0.7}\right)$

part i) giving $J_{AB} = 0$ - not correct - there must probability of winning

i.e. ii) > i when $J_{AB} = 0$

part ii) giving $J^*_{AB} = \dfrac{0.3}{0.65} = \boxed{0.4615}$ - Shoot B preferred

i.e. $0.5(0.4615) < 0.35(0.4615) + 0.3$

eqn ②: $J^*_{AC} = \max(0.4 J_{AC},\ 0.28 J(AC) + 0.3)$

using a similar argument,

$J^*_{AC} = \dfrac{0.3}{0.72} = \boxed{0.4167}$ - Shoot C preferred eqn (3)

(i) term :

$$J^{+}_{ABC} = (0.7)(0.5)(0.4)\,J_{ABC} + [(0.3)(0.4) + (0.7)(0.5)(0.6)]\,J_{AC} + (0.7)(0.5)\,J_{AB}$$

$$\Rightarrow J^{+}_{ABC} = \frac{0.299}{0.86} = 0.3477$$

ii) term :

$$J^{+}_{ABC} = 0.14\,J_{ABC} + [(0.3)(0.5) + (0.7)(0.5)]\,J_{AB} + (0.5)(0.7)(0.6)\,J_{AC}$$

$$\Rightarrow J^{+}_{ABC} = \frac{0.3183}{0.86} \simeq 0.37$$

iii) term :

$$J^{+}_{ABC} = 0.2\,J_{AB} + 0.5\,J_{AB} + 0.3\,J_{AC}$$

$$\Rightarrow J^{+}_{ABC} = \frac{0.356}{0.8} = \boxed{0.445}$$

Using $J^{+}_{ABC} = 0.445$,

(iii) term > (i) & (ii)

$\Rightarrow$ eqn is solved. — ~~shoot is cui~~ preferred.

State	$J^*(x)$	$\mu^*(x)$ (optimal action)
A lost	0	— (A can't do anything)
AB	0.4615	Shoot B
AC	0.4167	Shoot C
ABC	0.445	Shoot in air