Q 1.1

C) A search iteration may perform zero state expansions (that is, expand no states at all) yet guarantee that a previously found solution is at most -suboptimal.

Q 1.2

Q1 1.2

Update

$h(s_{start}) = \min_{s \in succ(s_{start})} c(s_{start}, s) + h(s)$

Initial heuristic : Consistent.

$\downarrow$

$h(s_{start}) \leq c(s_{start}, s) + h(s)$

where $s \in succ(s_{start})$

$\Rightarrow h(s_{start}) \leq \min_{s \in succ} c(s_{start}, s) + h(s)$

=> After update

$\Rightarrow h(s_{start}) = \min_{s \in succ} c(s_{start}, s) + h(s)$

$\therefore h(s_{start}) \leq c(s_{start}, s) + h(s)$

Hence Even After update heuristic is ~~com~~ consistent.

Q → 1.1 ⇒ C

Q2

a) Each state can 3 possible actions. Since there are 3 states, this implies total number of policies will be 3^3 = 27

Q2

b)

Let the initial policy be balanced for every state.

=> ~~$V(N) = R(N) + \sum T(\overset{N}{\cdot}, B, F$~~ Policy Evaluation

$$V^{\pi}(N) = R(N) + \sum T(N, aB, s_i)\left(\gamma V^{\pi}(s_i)\right)$$

where N : None F : For A : Against
B : Balanced O : offensive
D : Defensive

Similarly for $V^{\pi}(F)$ & $V^{\pi}(A)$

solving eqn we get

$V^{\pi}(N) = -20/169$ $\quad V^{\pi}(F) = 149/169 = 0.88$

$= -0.12$ $\quad V^{\pi}(A) = -189/169$

$= -1.12$

For Policy Improvement.

(1) — $$\pi_{k+1}(s_i) = \arg\max_a \left\{ r_i + \gamma \sum_j P_{ij}^a V^{\pi_k}(s_j) \right\}$$

where $P_{ij}^a = T(i, a, j)$

since r_i is constant for particular s, and γ is constant.

(2) — $\pi_{k+1}(s_i) = \arg\max_a \left\{ \sum_j T(i,a,j)(s_j) \right\}$

for $S =$ ~~F, A, N~~ F, A, N

action came out to be ~~'Offensive'~~ 'Balanced'

=> ~~New policy is O: 'Offensive' for all states.~~

-> ~~Policy Eval~~ :

Therefore, our first guess of policy was correct.

Optimal policy.

F -> Balanced

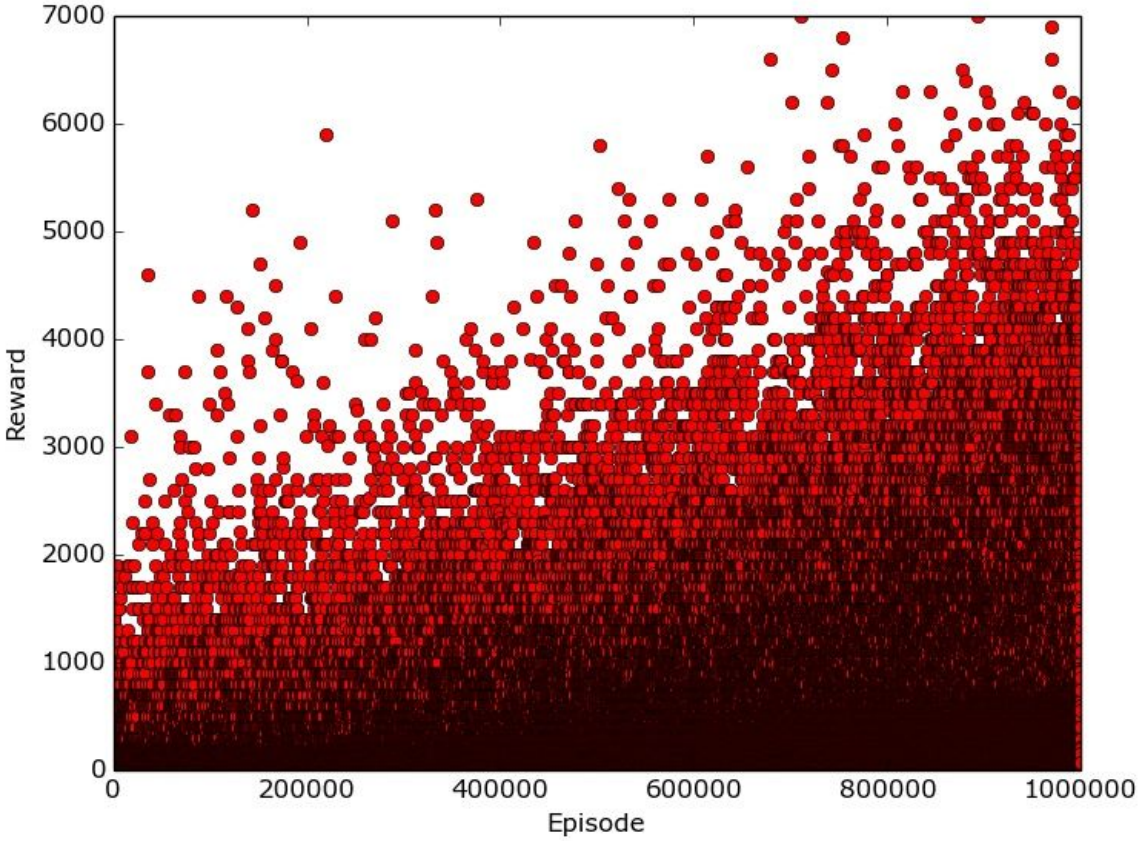
A -> Balanced

N -> Balanced

c) Since Discount factor is constant for all states, therefore ~~fendi eqn~~ equation (1) reduces to equation (2) which is independent of discount.

=> Hence changing the discount factor will not change the optimal policy.

Q3



Plot of the reward

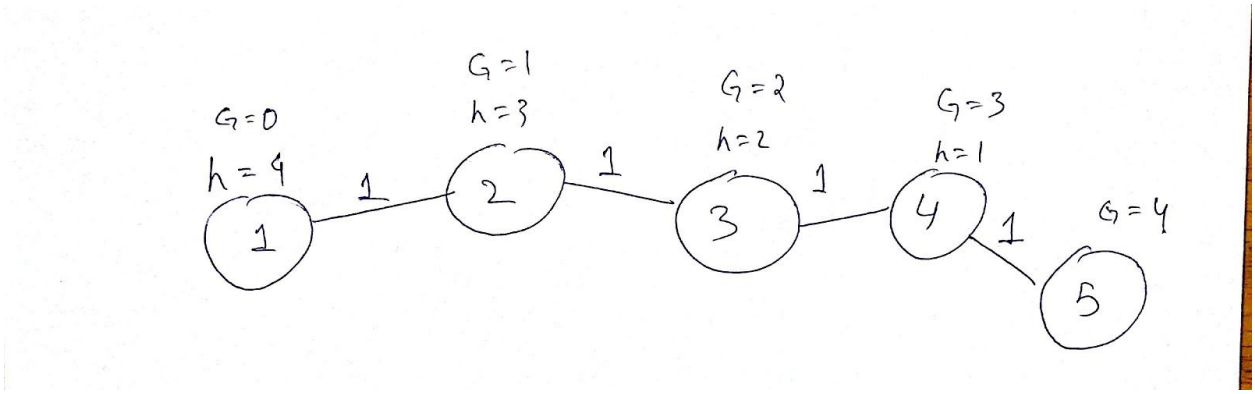
Policy is look like this
N,N,W,E,E,N,N,N,N,N,
N,N,N,N,W,W,E,N,N,N,
N,N,N,N,N,E,N,N,N,N,
E,N,E,N,W,W,N,N,N,E,
E,N,E,N,N,W,N,N,S,E,
N,N,N,N,N,N,N,N,N,S,
N,N,N,N,N,S,N,N,N,N,
N,S,N,N,E,N,N,N,N,N,
W,W,N,N,E,N,N,N,N,N,
N,E,N,E,N,N,N,N,N,N,

Where the top row has x = 0 and it increases downwards

Q4
Similarity metric could be a function of reward and domain. For example summation of number of same states, and same actions or same transition values. Further
Using similarity metric, and given the task, we can determine which is the policy of the past is the closest one with the current policy. Then this policy of the past is chosen to be reused, and using probabilistic bias in exploration strategy we can exploit this policy of the past.

Q5
d) None of These

Q6
Here is an example to incompleteness of weighted A*.
There are 5 states: 1,2,3,4,5: 1 being start and 5 being goal; each with edge cost determined by battery level as 1.
(show in fig below)



Now, let the initial battery level be 4.
h(s) = x of goal - x of s (where x is linear coordinate of the state).
Now if state 2 is expanded, then
h(3) = 5-3 = 2
So f(3) = 2 + 2 = 4;
Therefore final battery level of UAV would be 0 at goal state.
But if you inflate the heuristic 1 + delta then
h(3) = 2 + delta*2 > 2
There f(3)>4, hence it would be no more feasible, and the weighted A* will not be able to find solution even though it exist.