Q1C. Three drawbacks had been discussed in class: i) Does not handle confounders– can be rectified with Double ML (almost everyone wrote), ii) only handles linear relations– can be rectified with non-linear regression (a few students wrote) iii) cannot handle contemporaneous causal relations (hardly 1-2 students wrote). Each point carries 1 mark.

Q1D. The RCT conditions of statistical equivalence between Treatment and Control groups has not been followed. The AQI has improved in most cities in both groups anyway. The one city where it has drastically improved is T3, which had inadequate bus fleet per capita originally. The AQI improvement seems to have happened mostly in cities with better percentage of area under green cover. However, the specific metrics to be used (eg. number of cities where improvements happened, or average amount of improvement etc) can be chosen by own discretion. This was an open-ended question, any reasonable interpretation of the data has been awarded with some marks (4-6).

Q1E. fA is more accurate, so choose SA and ignore SB. The sum of the Shapley Values for each example equals the anomaly for it, so the expectation turns out to be 48. We find that HDI is most important factor in most cases, followed by PCI, but GDP is not important. For region 1, the HDI is quite low, which fails to compensate for the relatively high PCI and somewhat high Net GDP. Region 9 has very good HDI, but its people are dissatisfied due to low PCI.

Many students have considered average Shapley value for each factor, and concluded whether that factor has a positive or negative impact based on that. But this is wrong interpretation. We can consider absolute Shapley value of each feature to understand whether the feature's impact is strong or weak on the output. Whether the increase in the feature value impacts the outcome positively or negatively is a different question that needs to be seen separately.

Q2A. Need to consider GDP growth (diff. of GDP between successive years) for HMM!

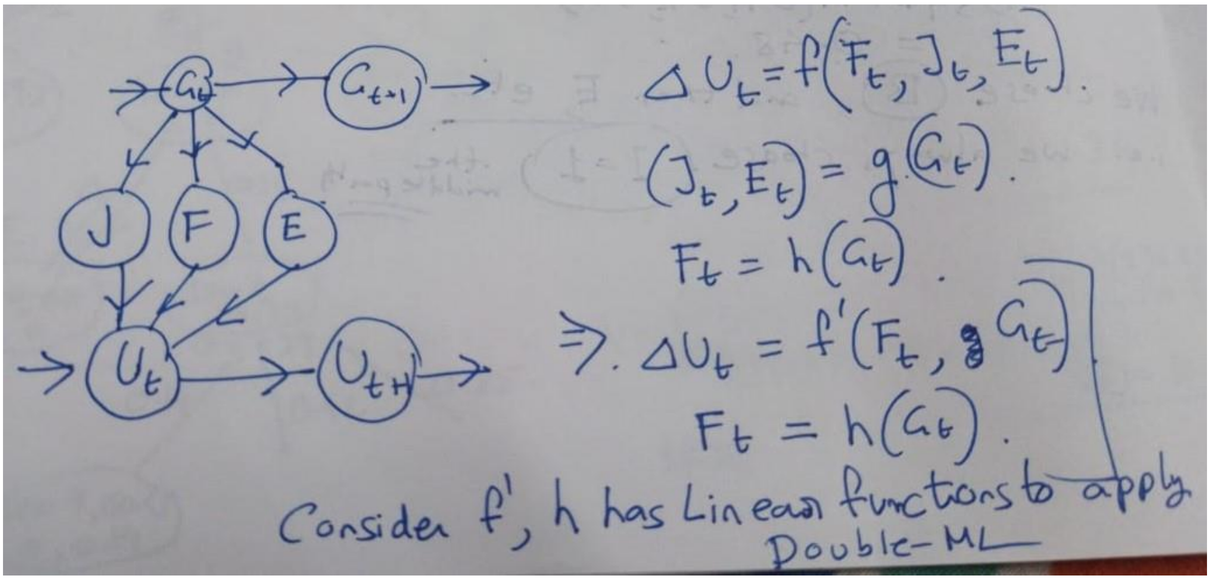
t	1	2	3	4	5	6	7	8	9	10
GDP	305	306	310	321	322	322	327	334	340	341
GDP growth	9	1	4	11	1	0	5	7	6	1

The HMM states can be decided by looking at the GDP growth values, eg. 0-4 (low), 5-7 (medium), >7 (high). 2 states will also do. Once each observation has been assigned a state variable value Z(t), it is easy to estimate transition distribution and emission distribution parameters. Gaussian distribution can be considered for emission.

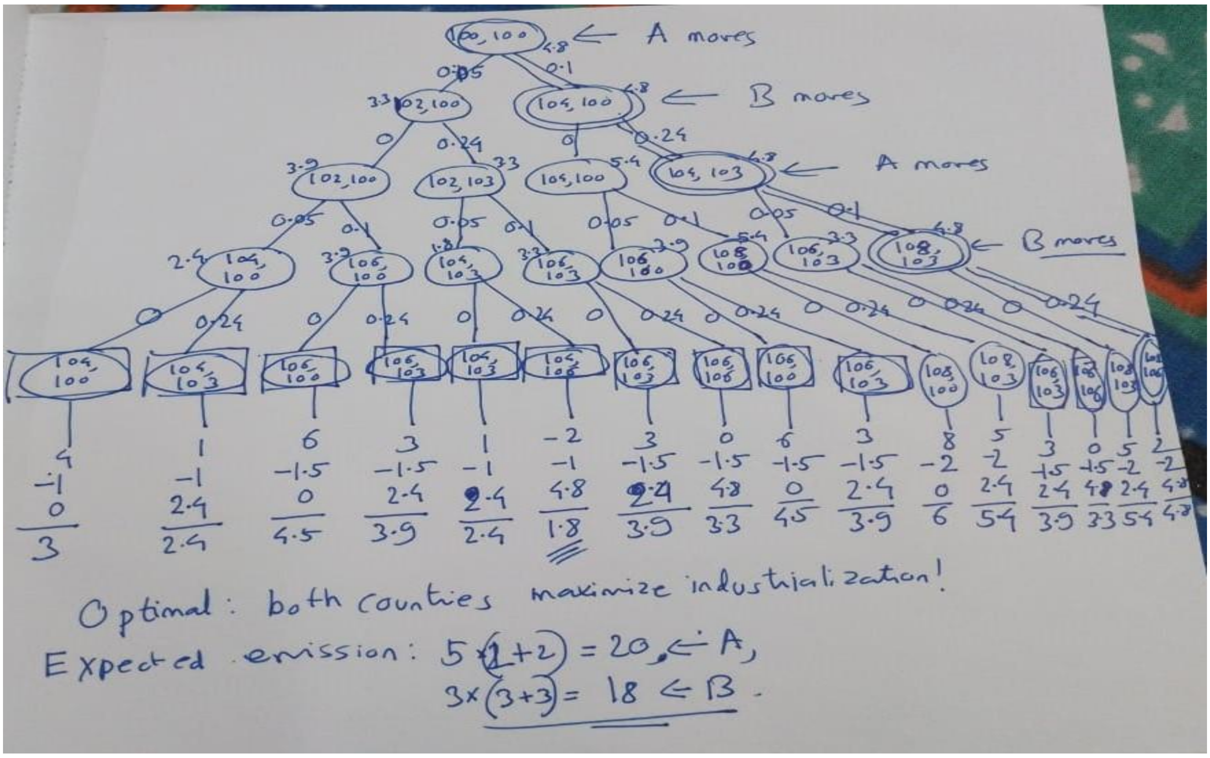
Q2B. Expected Number of employed people = Past number of employed people + expected no. of new jobs – expected no. of jobs abolished. For t=11, we need to consider the expected number of jobs created or abolished, based on possible value of the state variable. Since we know state variable at t=10, its possible values at t=11 can be found based on the emission distribution. The expected number of jobs created = (mean jobs created for Z=1)*prob(Z(11)=1) + (mean jobs created for Z=2)*prob(Z(11)=2) + (mean jobs created for Z=3)*prob(Z(11)=3). 2 marks for the calculating no. of employed and unemployed people till t=10, 2 more marks for t=11.

Q2D. RNN structure, hidden state update equation, emission equation should be clearly written, loss functions clearly defined, model parameters identified, and the parameter update process (co-ordinate descent) should be clearly explained for full marks.

Q2C.

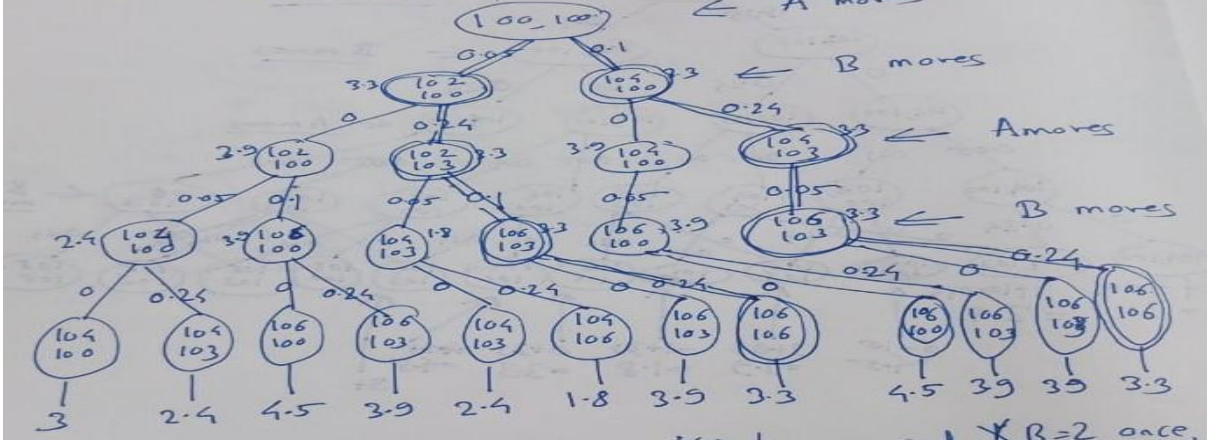


Q3B



Q3C

If A has cut off policy:



In this case A will chose $X_A=1$ once and $X_B=2$ once, while B will take $X_B=3$ both times.
Either way ~~both~~ the race will be tied.

Q4.

(i) $$u_2(b_2) = (b_2 - v_2) \cdot \Pr[\text{2nd player wins}]$$

$$= (b_2 - v_2) \cdot \prod_{\substack{i \in [5] \\ i \neq 2}} \Pr[\text{player } i \text{ bids more than } b_2]$$

each player is symmetric from the perspective of player 2 ←
$$= (b_2 - v_2) \prod_{\substack{i \in [5] \\ i \neq 2}} \Pr[\alpha v_i > b_2]$$

$$= (b_2 - v_2)\left(1 - \frac{b_2}{20\alpha}\right)^4$$

$u_2'(b_2^*) = 0$, b_2^* is a maximizer of $u_2(b_2)$.

$$\left(1 - \frac{b_2}{20\alpha}\right)^4 - \frac{4(b_2 - v_2)}{20\alpha}\left(1 - \frac{b_2}{20\alpha}\right)^3 = 0$$

$$\Rightarrow \quad 1 - \frac{b_2}{20\alpha} - \frac{1}{5\alpha}(b_2 - v_2) = 0$$

$$\Rightarrow \quad 20\alpha - b_2 - 4b_2 + 4v_2 = 0$$

$$\Rightarrow \quad 20\alpha - 5v_2\alpha + 4v_2 = 0$$

$$\Rightarrow \quad \alpha = \frac{4v_2}{5v_2 - 20} = \frac{6}{5}$$

$$\therefore \text{ bid of second player} = \frac{6}{5} \times 12 = \frac{72}{5}$$

(ii – v) since the payment of player i is a function of the bids of the other players only, these auctions are DSIC by Groves Theorem. Hence, bidding valuations is a dominant strategy for each player.

Q5A. The set of actions / pure strategies of player 1, $x_1 \in \{0,1,2 \ldots \ldots .100\}$. The set of actions / pure strategies of player 2, $x_2 \in \{0,1,2 \ldots \ldots .100\}$.

A (A+B)/3 (A+B)/2 B

Clearly $\frac{2}{3} . \frac{x_1+x_2}{2} = \frac{x_1+x_2}{3}$ is closer to min $\{x_1 , x_2\}$.

So the payoff matrix M is given by

M_{ij} = (1,0) if $i < j$

= (1/2 , ½) if i = j

where Player 1 is the column player and player 2 is the row – player.

Hence the only pure strategy Nash Equilibrium is (x_1 , x_2) = (0,0).

I have given 3 or more marks to those who got the payoffs right.

Q5B. Each of three men (A, P and D) either "like a guard" (L) or "dislike a guard" (D) There are 8 possible preference profiles for A, P and D. The set of preference profiles is given by S = {LLL, LLD, LDL, LDD, DLL, DLD, DDL, DDD}. Ofcourse LLD, LDL, DLL are exactly similar. Also DDL, DLD, LDD represent exactly similar scenarios. So we can shrink the set of possible preference profiles into S' = {LLL, LLD, LDD, DDD}. The selection rule is given by α ∈ {1,2,3}. Now the problem has 12 possible cases given by ($S' \times$ α). For each such case prove that if two men are revealing their preference truthfully, the third man does NOT have any incentive to lie.

Most students have NOT approached the problem correctly.

i. You first need to enumerate all possible preference profiles and then analyse for every α.

ii. Secondly most students have NOT used the definition of a dominated / dominating strategy.

Marks have been awarded as per the approach and arguments provided.

Q5C. Exactly as done in class. Two key parts:

i. The risk – free interval.

ii. The max. risk free payoff.

If any of the two above – mentioned points aren't present I have deducted accordingly.

Most of you have scored full (5) marks in this.