CS217 **Artificial Intelligence and Machine Learning** **Endsem**

21 Apr25 **1.30 PM to 4.30 PM** **50 marks**

(You have to solve 14 problems in 3 hours. No extra sheet will be given. Do rough work and then post the neat and clean answer onto the space provided in the question paper itself. Illegible, round-about, unnecessarily complicated answers will not be evaluated.)

(Q1 through Q10 are *multiple-choice* questions. There may be multiple correct answers for each question, so you must select all correct options. Below each question, an **ANSWER** field is provided. You must specify your chosen options only in this field. Marking options with ✓ (tick) or ✗ (cross) beside them will not be accepted. Specifying chosen options anywhere else in the answer sheet will not be considered for grading. Correct answer to each question will carry +1 marks and there is *no negative marking*.)

(Q11 through Q14 are *subjective* questions. You need to post the neat and clean answer onto the space provided in the question paper itself. No extra sheet will be provided for you to write solutions for the subjective questions. Maximum marks for each question is mentioned below the respective questions.)

Q1, Q2, and Q3 are based on the following context.

A recommender system logs the following data over a period of time: Red items were shown 80 times, of which they were clicked 24 times. Blue items were shown 120 times, of which they were clicked 36 times. Based on these logs, one can compute an estimate of each of these quantities:

1. Pr(click on red | shown red) which is the probability that the item will be clicked given that a red item is shown.
2. Pr(click on blue | shown blue) which is the probability that the item will be clicked given that a blue item is shown.
3. Pr(click on red | click) which is the probability of click on a red item given that a click happened.
4. Pr(click on blue | click) which is the probability of click on a blue item given that a click happened.

[Q1] What is the estimate for Pr(click on red | shown red), Pr(click on blue | shown blue), Pr(click on red | click), Pr(click on blue | click), in that order? - 1 Mark

(A) 0.20, 0.20, 0.40, 0.60
(B) 0.30, 0.30, 0.40, 0.60
(C) 0.30, 0.30, 0.25, 0.75
(D) 0.20, 0.20, 0.60, 0.40

Correct Answer: (B)

Solution:

Pr(click on red | shown red) = 24/80

Pr(click on blue | shown blue) = 36/120

Pr(click on red | click) = 24/(24+36)

Pr(click on blue | click) = 36/(24+36)

[Q2] Assuming that the probability of click given an item is a true measure of user preference, which of the following is correct based on the above example? - 1 Mark

(A) Red items are preferred more than blue
(B) Blue items are preferred more than red
(C) Both items are equally preferred
(D) Insufficient evidence

Correct Answer: (C)

Solution:

Pr(click on red | shown red) = Pr(click on blue | shown blue)

[Q3] In the example above, it was later revealed that red items were always shown at a higher position than any blue item in the logged recommender system. It is also known that higher positions have a better probability of getting clicked than lower positions. Which of the following is true about user preference in this case based on this additional evidence?

- 1 Mark

(A) Red items are preferred more than blue
(B) Blue items are preferred more than red
(C) Both items are equally preferred
(D) Insufficient evidence

Correct Answer: (B)

Solution:

With position bias we had that Pr(click on red | shown red) = Pr(click on blue | shown blue). Correcting for position bias we should have that Pr(click on red | shown red) < Pr(click on blue | shown blue) since red was advantaged due to position.

[Q4] What is the complexity of backward induction for *n* player game (each taking round robin moves, e.g., like 4 player ludo) with each player having at most *k* moves in each round and the maximum number of rounds is *d* (one round is when every player makes one move)?

- 1 Mark

(A) *O(log(knd))*

(B) *O(n^{kd})*

(C) *O(k^{nd})*

(D) *O(nkd)*

Correct Answer: (C)

Solution:

The vanilla backward induction visits every node of the game tree and makes a decision. There are O(k^{nd}) nodes in the game tree.

[Q5] Alpha-Beta Pruning is a variant of which search strategy? - 1 Mark

(A) Greedy search

(B) Depth-first search

(C) Breadth-first search

(D) Hill climbing search

Correct Answer: (B)

Solution:

Alpha-Beta Pruning is a variant of depth-first search. It is used in conjunction with the minimax algorithm, which typically employs a depth-first search strategy.

[Q6] Alpha-Beta Pruning is used to reduce the complexity of which problem in AI? - 1 Mark

(A) Combinatorial explosion in game trees

(B) Data encryption in AI systems

(C) Network optimization

(D) Image processing

Correct Answer: (A)

Solution:

Alpha-Beta Pruning is used to reduce the complexity caused by the combinatorial explosion in game trees. It eliminates the need to explore all possible moves and scenarios by pruning irrelevant branches.

[Q7] Which factor does not directly affect the performance of Alpha-Beta Pruning? - 1 Mark

(A) The number of players in the game

(B) The branching factor of the game tree

(C) The depth of the game tree

(D) The order of move evaluations

Correct Answer: (A)

Solution:

The number of players in the game does not directly affect the performance of Alpha-Beta Pruning. Factors like the branching factor, the depth of the tree, and the order of move evaluations have more direct impacts.

[Q 8] An AI algorithm applied to a robotic system should perform the following functions,

- 1 Mark

(A) Sense, modify, respond

(B) Sense, analyse, respond

(C) Analyse, modify, respond

(D) Sense, analyse, modify

Correct Answer: (B)

Solution:

A robotic system requires three basic functions for executing specified tasks: (i) sensing/measuring the environment using available hardware, (ii) on-board processing to estimate its current state and remaining sub-tasks, and (ii) respond using actuators to move towards the specified goals.

[Q 9] Bullwhip effects in supply chains (constant cycling between very high and very low inventory) can be a result of, - 1 Mark

(A) Delays

(B) Modelling errors

(C) Adversaries

(D) Lack of ethics

Correct Answer: (A)

Solution:

Bullwhip effects are observed when the decision-making step is out of sync with actual physical movement of components. If one does not account for previous decisions (which have yet to be implemented in the system), one may overcompensate for perceived shortages. As a result, an excess amount of orders are placed, which leads to over-inventory. This cycle repeats.

[Q 10] AI algorithms are required for Industry 4.0 because, - 1 Mark

(A) They are the latest available methods

(B) They help us handle adversarial situations

(C) They can compensate for errors

(D) They allow us to scale to large number of components

Correct Answer: (D)

Solution:

As discussed in the lecture, Industry 4.0 brings a large number n of components onto the internet, each of which is capable of talking with other O(n) devices in the system. To manage effective communication through so many links, one needs highly scalable algorithms which can handle heterogeneous data. This capability is only provided by AI, in contrast to traditional hand-coded algorithms.

[Q 11] Facts and Rules: In a country, every person always either speaks the truth or tells lies (but not both). A tourist T comes to a junction in the country and finds an inhabitant S of the country standing there. One of the roads at the junction leads to the capital of the country and the other does not. S can be asked only **yes/no** questions.

Question: What **single** yes/no question can T ask of S, so that the direction of the capital is revealed?

MUST: a logic based systematic formulation of the problem solution should be created. Think of correct variables and operators and truth tables. Common sense based arguments will not be accepted. Instruments of logic as done in courses will be needed. **The final question asked must be emerging from a logic based formula** which itself was derived step by step.

- 10 marks

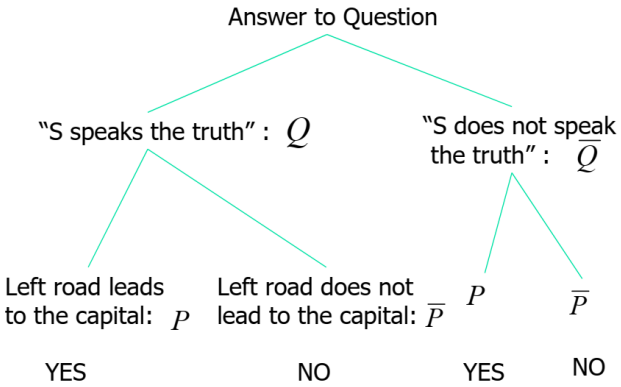
Solution:

Deciding the Propositions: a very difficult step- needs human intelligence

- P: Left road leads to capital
- Q: S always speaks the truth

The form of the question that the tourist should ask

- *Is R true?*
- *The answer is "yes" if and only if the left road leads to the capital*
- *The structure of R to be found as a function of P and Q*



P	Q	S's Answer	R
T	T	Yes	T
T	F	Yes	F
F	T	No	F
F	F	No	T

From the truth table

- *R is of the form (P x-nor Q) or (P ≡ Q)*

Hence the question T asks is**: *Is it true that the left road leads to the capital if and only if you speak the truth?***

[Q 12] Given N variables, there are $2^{2\wedge N}$ N Boolean Functions (BF). Your task is to design a Neural Universal Boolean Function Computer which can give the input output behaviour of ANY Boolean function of N variables. The architecture of the n/w is fixed, i.e., it does not change from one Boolean function to another. The n/w has a single output (0/1). There is only ONE hidden layer. There are N input neurons. The n/w is pure feedforward. You have to strictly adhere to this architecture. What changes from one BF to another is the set of connection weights and neuron biases. For example, given 2 input variables, the n/w computes the FALSE function, the AND function, OR, XOR and so on keeping the architecture fixed but changing the weights and biases.

Determine the size of hidden layer. Give the GENERAL SCHEME that applies across the BFs for fixing weights and biases of the neurons in the n/w all of WHICH ARE PERCEPTRONS.

-10 marks

Solution:

Universal Boolean Function Computer with a Feedforward N/W of perceptrons. There will be $2^{2\wedge N}$ output neurons and 2^N hidden layer neurons.

Suppose the function is XOR. The truth table is shown in table 1 below:

Row	X_1	X_0	OUT
0	0	0	0
1	0	1	1
2	1	0	1
3	1	1	0

Table 1: Truth table of XOR

Architecture: 2 neurons in the input layer. $2^2=4$ neurons in the hidden layer. Each hidden layer neuron *takes charge* of one row of input in the truth table, in the sense that given that input pattern this and only this one neuron fires. No other hidden layer neuron fires.

Only one neuron is kept as the output.

Weight assignment scheme:

Input to hidden layer:

For hidden layer neuron H_i,

(a) wt. from X_0 is $+2^0= 1$,if X_0 is 1 in the bit pattern for row i

else, it is $-2^0= -1$

(b) wt. from X_1 is $+2^1= 2$, if X_1 is 1 in the bit pattern for row i

else it is $-2^1= -2$

Hidden to output layer:

The OUT column of the truth table is replicated as weights from the hidden layer to the output layer.

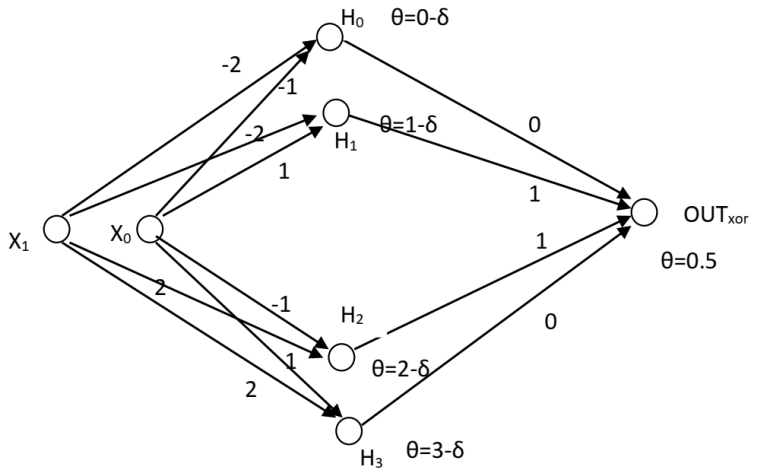


Figure 1: FF N/W for XOR

Threshold assignment scheme:

For output neuron; θ=*0.5* (only one neuron in the hidden layer fires and that needs to be passed on)

For hidden layer neurons: For H_i, θ= *(i-δ)*, where δ is a small positive quantity.

The corresponding FF N/W shown in figure 1.

The scheme is very general. For AND, OR or any other 2 variable function only the weights to the output neuron will change (equal to the OUT column for that function).

How does this work:

1. Suppose row *0* pattern *(0,0)* is presented. Only H_0 fires. Other hidden neurons do not. But the weight from H_0 to OUT is 0. Hence output of OUT remains 0.
2. Suppose row *1* pattern *(0,1)* is presented. Only H_1 fires. Other hidden neurons do not. The weight from H_1 to OUT is 1. Hence output of OUT becomes 1, as it should.
3. For row 2, only H_2 fires and so does OUT.
4. For row 3, only H_3 fires. But OUT does not, since the wt. value is 0.

General case: For N inputs: 2^N neurons in the hidden layer. Everything is same. Only that,

For hidden layer neuron H_i,

wt. from X_j is $+2^j$

if X_j is 1 in the bit pattern for row *i*

else, it is -2^j

Threshold of H_i = *(i-δ)*

Why does only one neuron fire in the hidden layer:

When the bit pattern for row *i* is presented, only at H_i does the dot product of W_{Hi} and *X* become equal to *i*. For all other neurons H_j the dot product is less than *j*. The weight

assignment is such that if the signs were all positive the weight values act like a binary to decimal converter.

[Q 13] Give a well formed formula involving predicates, variables, operators etc that is TRUE for ANY interpretation with infinite domain and FALSE for ANY interpretation with a finite domain. Prove your answer rigorously.

-3 +5=8 Marks

Solution:

One such formula is

$$\exists P[\forall x \exists y P(x,y) \wedge \forall x \neg P(x,x) \wedge \forall x \forall y \forall z[(P(x,y) \wedge P(y,z)) \Rightarrow P(x,z)]]$$

where $\neg$ means *negation* and $\Rightarrow$ means *implication*

The above formula yields the value TRUE for any interpretation with infinite domains and the value FALSE for any interpretation with finite domains.

Finite domain: say there are *N* elements. Let the predicate *P(x,y)* be expressed by a directed arc from *x* to *y*, where *x* and *y* are elements in the domain. As per the formula, each element in the domain is connected to another element in the domain. Now any time there is a cycle created through these connections, we will be forced to apply transitivity as expressed in the 3rd part of the formula and have a self loop *P(x,x)* which makes the above formula false. The cycle is inevitable, if the domain in finite. Since even if we always connect to a new element, the last element- the N^{th}- element will have to be connected to one of the previous elements. Thus for a finite domain, the formula will always be false.

Infinite domain: one can ALWAYS find a new element which to connect to. Hence the formula will always be true for an infinite domain.

[Q 14] (due to Judea Pearl) Suppose you live in an earthquake prone region. You have a new burglar alarm installed at home. The alarm is fairly reliable but is occasionally set off by minor earthquakes. You also have two neighbours John and Mary who have promised to call you when they hear an alarm. John nearly always calls when he hears the alarm, but sometimes confuses the telephone ringing with alarm and calls then too. Mary on the other had likes loud music and often misses the alarm altogether. Given the evidence of who called or not called we want to estimate the probability of burglary.

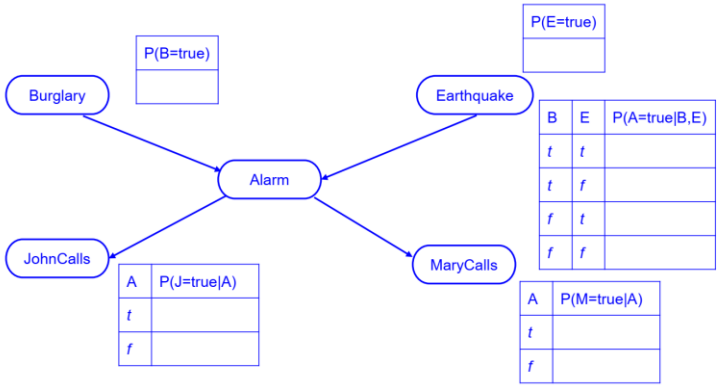
(a) Give the random variables of the situation. (2 marks)
(b) Which random variables are independent? Draw a dependency diagram. (4 marks)
(c) What probabilities need to be field-gathered from data/observation? (2 marks)
(d) What is the probability of John calling AND Mary calling AND Burglary happening? (4 marks)

- 12 Marks

Solution:

(a) Let B, E, A, J and M be respectively r.v.s for burglary, earthquake, alarm, john-calls and mary-calls. (2)

(b) The following diagram (Bayes Net) gives the dependencies.



The probabilities in the diagram are critical probabilities. They have to be filed-obtained, i.e., from data. Now we can answer the questions:

(c) Probability that John AND Mary call AND Burglary has happened:

P(B, J, M)

We have to introduce the other r.v.s by marginalization and get

$P(B,J,M)=\Sigma_E\Sigma_A P(B,J,M,E,A)=P(B,J,M,E,A)+P(B,J,M,\sim E,A)+P(B,J,M,E,\sim A)+P(B,J,M,\sim E,\sim A)$

Apply chain rule and see that P(B,J,M,E,A)

$=P(B,E).P(A|B,E).P(J|A,B,E).P(M|J,A,B,E)$

Apply independence as in the diagram and see that the above

$=P(B).P(E).P(A|B,E).P(J|A).P(M|A)$

All of which are known. For example, if P(B)=0.001, P(E)=0.002, P*A|B,E)=0.95, P(J|A)=0.9 and P(A|M)=0.7, then the above value is 0.000001 which is the probability of all B,E,A,J,M happening together. The value is small because the prior probability of B and E are themselves very low.

The other terms can similarly be obtained.