CS217 **Artificial Intelligence and Machine Learning** **Midsem**

22Feb25 **2.30 PM to 4.30 AM** **30 marks**

(All questions are multiple choice questions, except Q10 and Q11. For Q1 to Q10, more than one answer may be possible, in which case all of them have to be given. If the answer is wrong which may include not choosing all the correct options, you get -1; else 1.5 marks for each. Q10 is of 7 marks and Q11 of 4+4 marks. Use the supplied A4 sheet for answering Q10 and Q11. No extra sheet will be given

VIMP: do rough work and then post the neat and clean answer to the two sides of the A4 size paper. Illegible, round-about, unnecessarily complicated answers will not be evaluated.)

Q1. Searle's Chinese Room experiment __________ the Turing Test. Fill in the gap with the correct option from below:

(a) reinforces
(b) challenges
(c) is irrelevant to
(d) None of the above

Ans: (b)

Justification: as per definition of TT and SCR

Q2. For the sentence "$foxes_1$ $foxes_2$ fox_3 fox_4 $foxes_5$", the main verb is:

(a) $foxes_2$
(b) fox_3
(c) fox_4
(d) $foxes_5$

Ans: (c)

Justification: the meaning of the sentence is "Those foxes which other foxes deceive (fox), in turn deceive foxes"; hence fox_4.

Q3. In 8-puzzle, the goal state *<1,2,3,4,5,6,7,8,b>* (row major order, 'b' means blank) is reachable from:

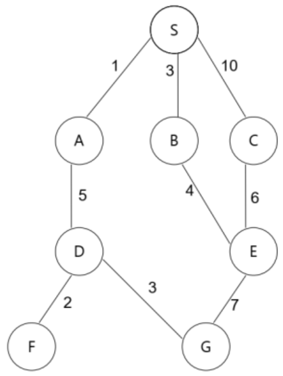
(a) <1,2,4,3,5,6,7,8,b>
(b) <1,2,3,4,6,5,8,7,b>
(c) <3,2,1,4,5,6,7,8,b>
(d) *None of the above*

Ans: (b)

Justification: the goal state has 0 (which is even parity) inversion, i.e., larger number before a smaller number. (a) has 1 (odd parity) inversion <4, 3>. (b) has 2 inversions (even parity)

<6,5>, <8,7>, and (c) has 3 <3,2>, <2,1>, <3,1>. Only a pair of states with the same parity of inversion can reach each other.

Q4. Consider the search graph



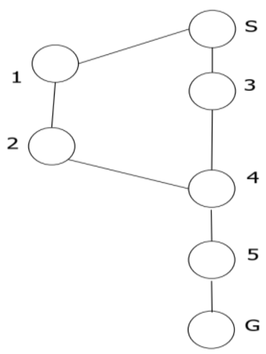
The heuristics values are, h(A)=13, h(B)=16, h(C)=18, h(D)=8, h(E)=12, h(F)=1000, h(G)=0. The cost of the path found by Algorithm-A is

(a) At most 14
(b) At least 14
(c) Greater than 18
(d) Depends on the order in which the nodes are expanded

Ans: (a)

Justification: the heuristic overestimates *h*()* by 5, hence the found path need not have a cost more than *optimal+5*. (proved in class slides)

Q5. Consider the search graph (each arc cost is 1):



It is seen in an A* algorithm that parent pointer redirection happens. The heuristic values of all nodes are 0, except that of (choose the correct option/s which is/are consistent with this observation of parent pointer redirection):

(a) 3, which is 3
(b) 4, which is 2
(c) 5, which is 1
(d) None of the above

Ans: (a)

Justification:

In case of (b), the expansion sequence is (OL, CL)

OL:S; CL: Empty

OL: 1 3 (F(1)=1, F(3)=4); CL: S

OL: 3 2 (F(2)=2, F(3)=4); CL: 1, S

OL: 3, 4 (F(3)=4, F(4)=3); CL: 2, 1, S

OL: 3, 5 (F(3)=4, F(5)=4); CL: 4, 2, 1, S

Now if 3 is expanded, 4 will need parent pointer redirection. This is happening because 3 violates monotone restriction.

Only (a) DEFINITELY ensures parent pointer redirection; other conditions may, but may not too.

Q6. Consider the proof of the theorem $\Phi \rightarrow P$ in Hilbert's propositional calculus, where Φ is the syntactic 'false' symbol.

$\Phi,\ P \rightarrow \Phi \models \Phi$ (using deduction theorem)

⇨ $\Phi \models XYX$
⇨ $\Phi \models$ XYZ, PQR

$\models$ P

⇨ $\models \Phi \rightarrow P$ (proving the proposition

XYZ and PQR are intermediate steps (what are they?). The prefinal step of 'P' comes through (choose the correct option):

(a) Material transfer to the right of |=, application of Axiom 1 and application of Modus Ponens
(b) Material transfer to the right of |=, application of Axiom 2 and application of Modus Tolens
(c) Material transfer to the right of |=, application of Axiom 3 and application of Modus Ponens
(d) None of the above

Ans: (c)

Justification:

XYZ is $(P \rightarrow \Phi) \rightarrow \Phi$ and PQR is $[(P \rightarrow \Phi) \rightarrow \Phi] \rightarrow P$ *(A3);* now apply MP.

Q7. A probabilistic BP (PBP) algorithm goes as follows. A weight in a network (which has many weights) is changed by gradient descent rule with probability P, and by an amount $c.|w|$ with probability $(1-P)$, where c is a fraction and $|w|$ is magnitude of w. For a particular instance of PBP, a weight w is such that the loss L vs. w curve (a projection in the L-w plane) has a local minimum at $<w=0,\ L=0>$ and $L=0.5w^2$ in a small neighbourhood of this minimum. For a weight value of w=-0.01 (minus 0.01) and learning rate of 0.1, which of the conditions on *c* can be expected to make the n/w bypass the local minimum in a single step at this value of *w*?

(a) c<0.1
(b) c>0.1
(c) c=0.2
(d) None of the above

Ans: (d)

Justification: The gradient descent component of weight change is –(0.1 X -0.01)=0.001. The other component is 0.01c

Now, for bypassing the local minimum we need a step size of at least 0.01 (w is at a distance of 0.01 from <0,0>).

P X 0.001 + (1-P) X 0.01c > 0.01

➔ P+10c(1-P)>10

⇨ P(1-10c)>10(1-c)

Since, 0<=P<=1, and c is a fraction, this is impossible.

Q8. In an instance of BP training, a momentum factor *m* is used, such that the component of weight change by momentum factor is far greater than the component by gradient descent. The learning rate is 0.001 and the momentum factor *m*=0.95. As the number of iterations increase to a large value, the weight change value for a particular weight which was initially changed by an amount of 0.02 tends to

(a) 0.2
(b) 0.5
(c) 0.4
(d) None of given options is correct

Ans: (c)

Justification:

Since the gradient descent component of the weight change rule is insignificant compared to the component due to momentum factor, we have

$$\Delta w_\infty = \Delta w_0 \text{ X } (1+m+m^2+m^3+\ldots) = \Delta w_0/(1-m) = 0.02/(1-0.95) = 0.02/0.05 = 0.4$$

Q9. A pure and fully connected feedforward neural net has $2P$ input neurons $(X_i, i=1,2P)$, $2Q$ hidden neurons $(H_j, j=1,2Q)$ and $2R$ output neurons $(O_k, k=1,2R)$. The weight naming convention is destination-source, e.g., O_2-H_1 means the connection FROM H_1 TO O_2. For O_k-H_j and H_j-X_i, the symmetry breaking condition/s is/are (choose all that is correct):

(a) For $k=1,2R$ and $j=1,Q$, break $O_k\text{-}H_j=O_k\text{-}H_{2Q-j+1}$
(b) For $j=1,Q$ and $k=1,R$, break $O_k\text{-}H_j=O_{k+R}\text{-}H_{j+Q}$,
(c) For $i=1,2P$ and $j=1,Q$, break $H_j\text{-}X_i= H_{2Q-j+1}\text{–}X_i$
(d) None of the above help break symmetry

Ans: (a),(c)

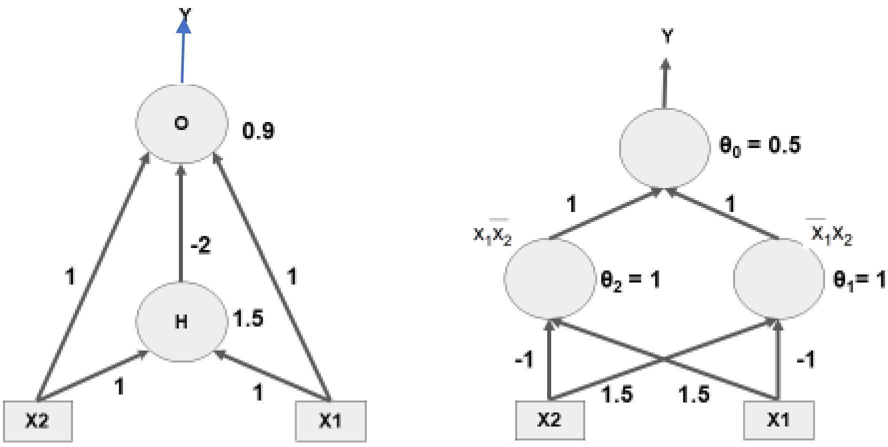
Justification: keep the hidden layer horizontal and the i/p and o/p layers vertical. Now break symmetry of pairs of weights around the vertical line of symmetry.

Q10. XOR is being computed by a feedforward n/w. There is no restriction on the architecture except that it is feedforward. Let MAX be the number of hidden layer neurons that is <u>sufficient</u> to compute XOR and let MIN be the number that is <u>necessary</u>. Then *MAX-MIN* is

(a) At most 1
(b) Exactly 1
(c) At least 1
(d) Exactly 0

<u>Ans</u>: (b)

Justification**:**



- MAX 2 hidden layer neuron are <u>sufficient</u> for feedforward XOR (right n/w)
- MIN 1 hidden layer neuron is <u>necessary</u> for feedforward XOR (left n/w)
- Hence MAX-MIN = 1 (exactly 1)

Q11 (answer on one side of the page). Prove rigorously that the number of threshold functions of m variables is $O(2^{\mu})$, where $\mu = m^2$. You can make use of the solution of the recurrence relation proved in the class.

-7 marks

Ans:

Proved in class, n hyperplanes in a space of dimension d passing through origin produce regions

$$R_{n,d} = 2\sum_{i=0}^{d-1} C_i^{n-1}$$

For the given situation, $n=2^m$ and $d=m+1$ (bias absorbed as additional weight), hence the required quantity

$$
\begin{aligned}
&= 2\sum_{i=0}^{m+1-1} C_i^{2^m-1} \\
&= 2\sum_{i=0}^{m} C_i^{2^m-1} \\
&\approx 2\sum_{i=0}^{m} C_i^{2^m} \\
&\le 2\sum_{i=0}^{m} (2^m)^i \quad (use\ \ {}^nC_r \le n^r) \\
&= 2.\frac{(2^m)^m - 1}{2^m - 1} \quad (G.P.\ series) \\
&\approx 2^{m^2-m+1} \\
&= O(2^{m^2})
\end{aligned}
$$

Q12.

(a) (answer on half page) In perceptron training, if it ever happens that a weight vector reappears, then the function attempted to be trained on the perceptron must be linearly NON-separable. You have to prove, necessarily making use of the *G(W_n)* function defined in the class.

-4 marks

Ans:

$G(W_n)=(W_n.W^*)/|Wn|$

Now $W_n.W^*=(W_{n-1}+X_{fail}).W^*=W_{n-1.}W^*+X_{fail}.W^*= W_{n-1.}W^*+positive\_no$

So $W_n.X^*>W_{n-1}.W^*$

This means W_ns are strictly increasing. So, if all assumptions are correct then no weight can repeat. Since, however, a weight has repeated, the assumption of W* existing is wrong, i.e., the function is linearly non-separable.

(b) (answer on half page) Prove in the Hilbert system of propositional calculus the commutativity of the AND operation (one direction is enough). You can use only axioms, modus ponens and any metatherem. If you need a theorem, you have to prove it.

- 4marks

Ans:

Apply Deduction Theorem

To show P.AND.Q → Q.AND.P

i.e., [(P→(Q→F))→F] →[(Q→(P→F))→F]

i.e., [(P→(Q→F))→F], (Q→(P→F)) |= F --(A)

Now (Q→(P→F)) |= (P→(Q→F)) --(B)

since (Q→(P→F)), P, Q|=F

From (B), (A) follows.