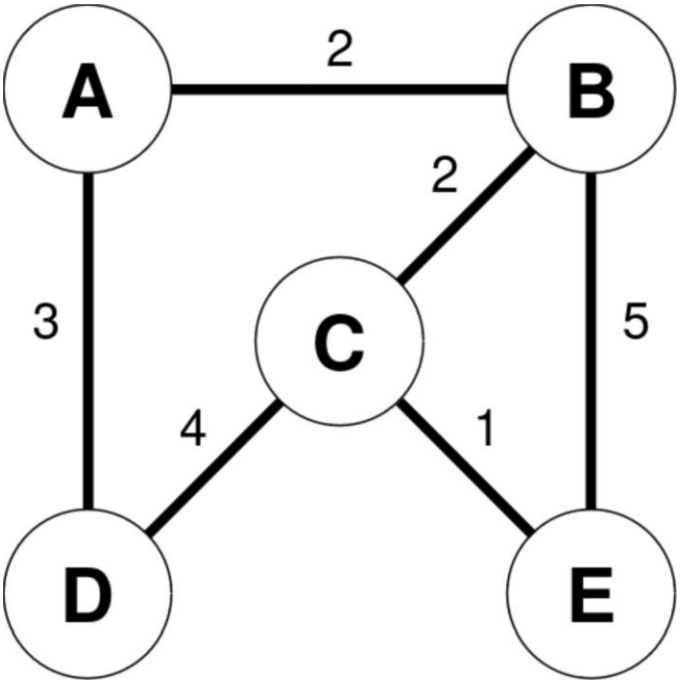
AI MSc Sample entrance exam paper

1. We run the Kruskal algorithm on the graph given in the figure below. Which edge is selected fifth by the algorithm into the spanning tree?



a. AD

b. AB

c. BE

d. CD

e. CE

f. BC

g. the spanning tree does not have five edges

---

2. Considering P ≠NP, which of the following algorithms is in P?

a. finding a Hamiltonian cycle

b. the bin packing problem

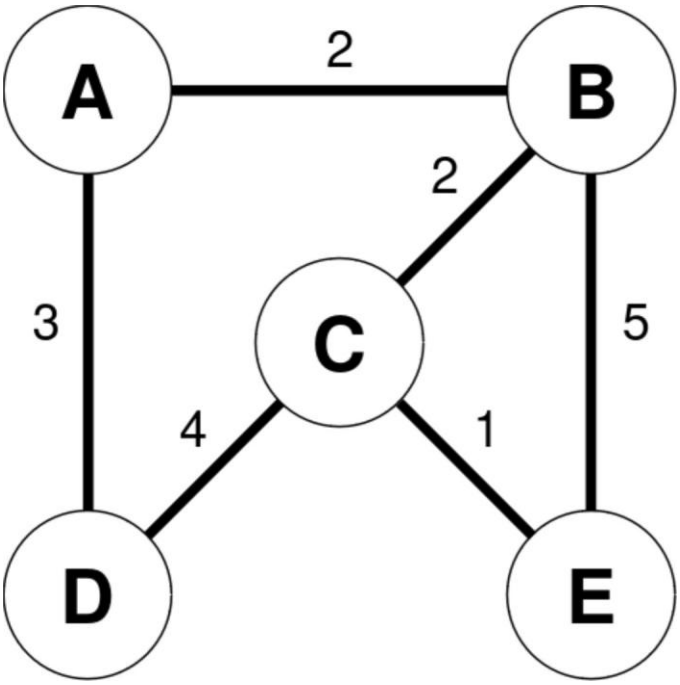
c. executing a simple neural network (i.e. not a large language model) in the inference phase

d. job scheduling

e. the 3SAT problem

+Note that this sample entrance exam contains more tasks for practice purposes than the official exam.

3. We run the Kruskal algorithm on the graph given in the figure below. At which step can edge AB be selected?



a. fifth

b. third

c. first

d. fourth

e. it is not selected

---

4. Which of the following properties is false for the output of the Kruskal algorithm?

a. we obtain the minimum-weight spanning tree

b. it is not connected

c. it cannot contain a cycle

d. there may be an edge in the input graph that is not part of the output graph

5. You and your friends went hiking in the mountains. Upon arriving at your accommodation, you noticed there is no mobile signal, and no one planned the route in advance. You must now plan the route using a paper map. Since the trails follow ridgelines and no cycles are possible in the terrain, **which algorithm would you use to recommend routes as quickly as possible?**

a. Bellman–Ford algorithm

b. Floyd–Warshall algorithm

c. Kruskal algorithm

d. Dijkstra algorithm

---

6. Which of the following statements is not true for the Bellman–Ford shortest path algorithm?

a. the input graph may not contain cycles with negative total weight

b. the algorithm outputs the distances from a single source to all vertices

c. its running time is of order if the input is a graph

d. the input graph may contain negative edge weights

---

7. Which algorithm would you choose to find the shortest paths from one vertex to all others in a weighted graph as efficiently as possible?

a. Dijkstra algorithm

b. Bellman–Ford algorithm

c. Floyd–Warshall algorithm

d. Ford–Fulkerson algorithm

---

8. Given a graph with vertex set and weighted edge set . Starting from vertex , we search for the shortest path to vertex using Dijkstra's algorithm. During the execution of the algorithm, how many times is the length of the path updated (assuming that initially it is set to infinity)?

a. once (1 time)

b. twice (2 times)

c. 3 times

d. 4 times

e. 5 times

---

9. Which statement is true about greedy search?

a. Greedy search is not complete, but it is optimal.

b. Greedy search is a best-first search in which the node that is closest to the goal according to the heuristic is selected.

c. Greedy search is complete, but not optimal.

d. Greedy search is a best-first search in which the node with the lowest cost from the start node is selected.

e. Greedy search is a best-first search in which the node selected is the one that is closest to the goal based on the backward cost.

---

10. Which statement is true about heuristics in the A* algorithm?

a. It only uses the heuristic after finding the optimal path.

b. The heuristic only modifies the edge cost.

c. No heuristic is needed.

d. The heuristic is random.

e. If the heuristic is admissible, A is optimal.

f. The heuristic can be negative.

---

11. Which of the following statements is true regarding search algorithms?

a. Breadth-first search is complete.

b. *A** search is never optimal.

c. Iterative deepening search is never complete.

d. Depth-first search is complete in all cases.

e. Uniform-cost search is not complete.

---

12. Which of the following is a local search strategy?

a. Depth-First Search (DFS)

b. Breadth-First Search (BFS)

c. Hill-climbing search

d. Alpha–Beta Pruning

e. Uniform-Cost Search (UCS)

f. A* star search

---

13. AI agents are commonly described using the sense–think–act cycle. Which phase does training a neural network belong to?

a. sense

b. think

c. act

d. none of the above, since this is a design-time step rather than run-time phase

---

14. AI agents are typically described using the sense–think–act cycle. Consider an application that receives input from a keyboard, performs inference using a neural network, and displays the result on a screen. Which phase is associated with receiving the keyboard input?

a. think

b. sense

c. act

---

15. Compared to a simple reflex agent, which type of environment can a model-based reflex agent handle?

a. partially observable environment

b. single-agent environment

c. episodic environment

d. multi-agent environment

---

16. AI agents are typically described using the sense–think–act cycle. Consider an application that receives input from a keyboard, performs inference using a neural network, and displays the result on a screen. Which phase is associated with executing the neural network?

a. act

b. think

c. sense

---

17. Consider a self-driving car operating autonomously on a highway. The vehicle is a physical system that may fail with a small probability (e.g., a tire blowout). Which of the following environment models is not appropriate for representing this scenario?

a. continuous environment

b. episodic environment

c. sequential environment

d. nondeterministic environment

e. multi-agent environment

---

18. In which type of environment is it not necessary for an agent to maintain internal state variables?

a. sequential environment

b. nondeterministic environment

c. discrete environment

d. fully observable environment

---

19. In a CSP, a unary constraint means that...

a. the constraints involve relationships among multiple variables, e.g. $X \neq Y \neq Z$

b. each variable can only take one type of value, e.g. X=0, Y=0, Z=0

c. each variable can only take true/false values.

d. the constraints involve multiple variables, e.g. X=1, Y=0, Z=1.

e. the constraint involves a single variable, e.g. X=4.

---

20. Which heuristic can be used for selecting variable values in a CSP?

a. Degree heuristic

b. Minimum Remaining Values (MRV) heuristic

c. Least Constraining Value (LCV) heuristic

d. Least Constraining Variable (LCV) heuristic

e. Minimal remaining variable (MRV) heuristic

---

21. What effect does alpha–beta pruning have on minimax search with an optimal move ordering?

a. It eliminates necessary evaluation.

b. It reduces the number of unexamined nodes to the minimum possible.

c. It increases the likelihood of heuristic errors.

d. It reduces the number of evaluated nodes to the minimum possible.

e. It increases the number of evaluated nodes.

f. It does not change the search time.

---

22. Which statement is characteristic of first-order logic?

a. It does not use quantifiers.

b. It does not distinguish individual objects.

c. It cannot express relations.

d. It contains predicates, variables, and quantifiers.

e. It only evaluates the truth values of statements.

f. It only deals with true or false statements.

---

23. Application of Bayes-rule. During a routine medical checkup, a diagnostic device may display an alert indicating that a patient has a certain condition. This may mean either that the patient truly has the condition, whose probability is $P(Condition) = 0.081$, or that the device produced an incorrect result. The requirement for the device is that, if the condition is actually present, it should give an alert with high probability $P(Alert|\ Condition) = 0.837$, and if the condition is not present, false alerts should be rare, with probability $P(Alert|\neg\ Condition) = 0.145$. Based on all this, what is the probability that the device does **not** display an alert, yet actually the patient has the condition, $P(Condition|\neg Alert\ )$?

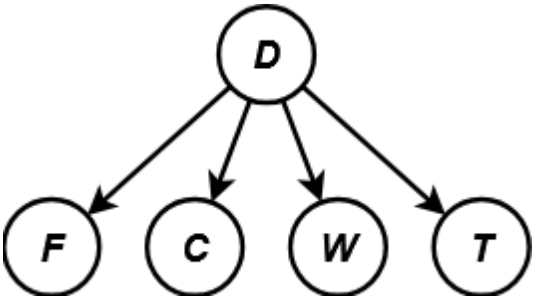
a. 0.0165

b. 0.785

c. 0.0132

---

24. A rare disease ***D*** can cause four conditionally independent symptoms which can be represented using a **naive Bayesian network**: fever ***F***, cough ***C*** , elevated white blood cell count ***W*** , positive rapid test result ***T,*** with the following structure:



We know, that:

P(D)=0.02 ***P(F|D)=0.8*** ***P(F|¬D)=0.1*** ***P(T|D)=0.9*** ***P(T|¬D)=0.05***

The patient has a fever and a positive rapid test result, while the variables ***C*** and ***W***are not yet observed.

Given the known evidence, what is the probability that the patient is diseased ***P(D | F, T)***?

a. 0.746

b. 0.424

c. 0.0144

d. 0.993

---

25. What is recall (sensitivity) in supervised learning?

a. None of the answers is correct.

b. The proportion of correctly classified positive examples (TP) out of all examples.

c. The proportion of correctly classified negative examples (TN) out of all actual negative examples.

d. The proportion of correctly classified positive examples (TP) out of all examples.

e. The proportion of all correct classifications.

f. The proportion of correctly classified positive examples (TP) out of all actual positive examples.

---

26. Which statement is true about the role of datasets in supervised learning?

a. Training data is always negligible.

b. With validation data, we perform the final evaluation independent of training.

c. All three datasets can be randomly combined.

d. Test data is used for training.

e. Test data is used for hyperparameter optimization.

f. Training data is used for learning, and validation data is used for hyperparameter tuning.

---

27. Which is characteristic of logistic regression?

a. Applicable to linear regression problems.

b. Prediction of a continuous target variable using probabilistic estimation.

c. Can be used for linearly separable regression.

d. Its typical loss function is the mean squared error (MSE).

e. Prediction of a continuous target variable.

f. Binary classification by estimating class membership probabilities.

---

28. Which statement is true regarding linear and logistic regression?

a. Both linear and logistic regression models are used to predict continuous values, i.e., to solve regression problems.

b. Linear regression predicts continuous values, while logistic regression predicts class probabilities.

c. Both can only be used for binary classification.

d. Both linear and logistic models are used to predict class labels, i.e., to solve classification problems.

e. Logistic regression predicts continuous values, while linear regression predicts class probabilities.

f. Both can only be used for continuous variables.

---

29. Which statement is true about the branches of decision trees?

a. The branching factor is predefined.

b. The branches are random.

c. They depend only on the smallest attribute value.

d. The branches are determined based on the values of the attributes.

e. They always depend on the attribute value with the highest entropy.

f. They do not depend on entropy.

---

30. Which statement is true about the depth of a decision tree?

a. The greater the depth, the lower the risk of overfitting.

b. The depth is always optimal.

c. The greater the depth, the higher the risk of overfitting.

d. It has no meaning in deterministic trees.

e. The depth of the tree is always at least 2.

f. The depth does not affect the performance of the tree.

---

31. Which characteristic is true for a decision tree?

a. It cannot be used in supervised learning.

b. It only operates with stochastic values.

c. It only searches for a local optimum.

d. A hierarchical structure with nodes and branches.

e. It uses only linear functions to make decisions.

f. A table of states and actions.

---

32. An emerging winery has decided to monitor the winemaking process through continuous quality control in order to ensure product quality and its distinctive flavor profile. To this end, they have commissioned a team of experts to taste wines produced using different methods and their variants, in order to determine whether they provide a pleasant taste experience or not.

Based on the obtained results (and the chemical composition of the wines), they began **constructing a decision tree**. At a node that still needs further expansion, there are 80 different wines, of which 60 are classified as positive (acceptable) and 20 as negative (not acceptable).

Another possible attribute is the wine's acidity. Taking this into account, the following [positive, negative] distributions are obtained:

Acidity	Mild [10,0]	Medium [50,15]	Acidic [0,5]

What is the information gain if the Acidity attribute is selected for evaluation, given that the initial entropy is 0.811?

a. 0 b. 0.811

c. 1 d. 0.7793

e. 0.6332 f. 0.1778

---

33. According to the Bellman equation, the utility of a state is

a. the sum of the expected discounted utilities of all possible next states, assuming the agent chooses the optimal action

b. the maximum of the expected discounted utilities of all possible next states, assuming the agent chooses the optimal action

c. the difference between the immediate reward for the given state and the discounted expected utility of the next state, assuming the agent chooses the optimal action

d. the immediate reward for the given state, assuming the agent chooses the optimal action

e. the sum of the immediate reward for the given state and the discounted expected utility of the next state, assuming the agent chooses the optimal action

34. What is the purpose of an exploration function?

a. To map a continuous state space into a discrete one

b. To assign higher priority to less frequently visited state-action pairs

c. To reduce the learning rate automatically over time

d. To terminate an episode if the agent receives a negative reward

e. To strictly enforce exploitation after a certain number of steps

f. To explicitly calculate the transition probability matrix T(s,a,s')

---

35. Which statement is false regarding Markov Decision Processes (MDP)?

a. The MDP has actions: $a \in A$

b. The MDP has a set of states: $s \in S$

c. The MDP has a reward function R(s,a,s')

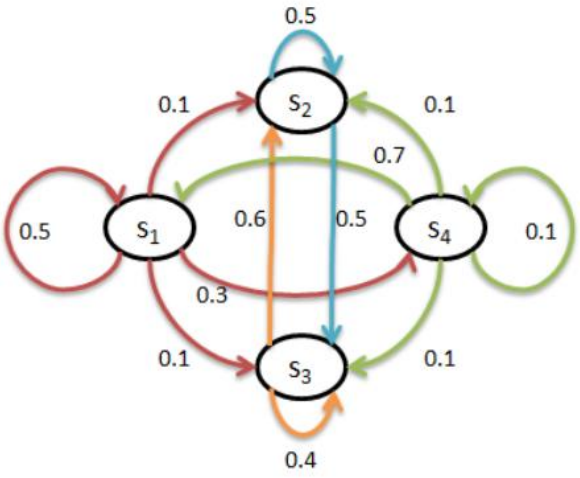
d. The MDP has a state transition function T(s,a,s')

e. The MDP has a courage factor: α

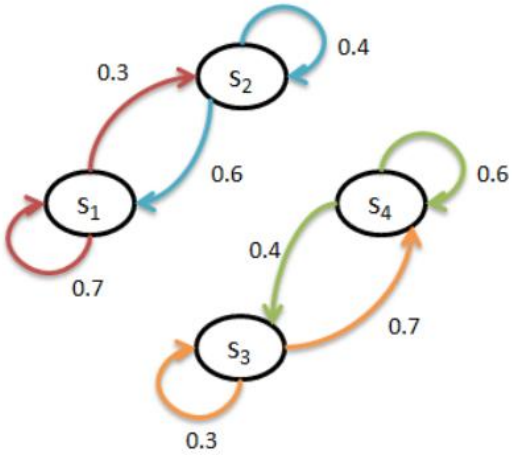
---

36. To solve an object relocation problem with an agent, we perform active reinforcement learning. A total of 4 states are modeled: **s1**, **s2**, **s3**, **s4**. We assume that in the initial state (**s1**), the object is fixed by the agent's actuator (grasped by the robot arm), but the agent is not in the correct location. In the target state (**s4**), it is in the right place and has the object. (**s2**: does not have the object and is not in the right place, **s3**: does not have the object but is in the right place.) The available actions are **a1** and **a2**, motion and object recapture. During the move action (**a1**), the object may 'fall out' of the agent's actuator (robot arm), and during the capture action (**a2** object recapture), the object may fall at first (i.e. capture unsuccessful). The state transition graph for each action is as follows:

T(s,a1,s')



T(s,a2,s')



The utility value of each state is as follows:

	s1	***s2***	***s3***	***s4***
U(s)	-0.9	-6	-2.2	8

Determine the new utility of state **s1** under the optimal strategy, assuming that the **immediate reward** is **R(s1) = -2.5** and the **discount factor is 0.7**.

a. -1.709

b. 0.8

c. -1.37

d. 1.13

e. 0

f. 1

Which action will be (part of) the optimal policy TT* for s1 ?

a. a2

b. Both are equally appropriate.

c. None of them.

d. a1

37. What is the role of α (learning rate) in reinforcement learning?

a. It defines the total number of actions available to the agent

b. It controls the exploration probability in the ε-greedy strategy

c. It scales the initialization value of the Q-table

d. It determines the duration of an episode

e. It scales the value of future rewards

f. It determines to what extent new information overwrites old information during updates

38. For the action-value function Q(s,a) (state-action pairs), it is true that

a. the environment model, i.e., the state transition probability matrix T(s,a,s') and rewards R(s,a,s'), is always required to calculate Q(s,a) values

b. it assigns an expected utility to choosing a specific action in a given state

c. the Bellman equation cannot be applied to Q(s,a) values

d. it cannot be learned directly based on reward feedback

e. temporal difference learning of Q(s,a) is based on an exponentially moving average, weighted by the discount factor and the state transition probability matrix T(s,a,s')

39. What is the role of the ε (epsilon) parameter in the ε-greedy policy?

a. It determines the number of layers in the neural network

b. It determines the learning rate of Q-values

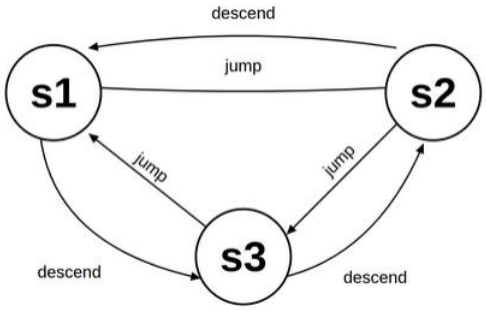
c. It defines the total number of states in the environment

d. It specifies the value of the maximum immediate reward

e. It determines the weight of future rewards

f. It controls the probability of choosing a random action for exploration

40. The aim of Flappy Birds Game is to score as many points as possible with the bird you control.



Time difference (IK) based Q-learning can be used to determine the optimal procedure. The bird's environment consists of three states: **s1, s2, s3**, in each of which two actions can be applied **: descent** and **jump**.

The **learning rate** is **α**: 0.2 and the **discount factor γ:** 0.5

The table of Q values:

	descent	**jump**
s1	4.3	4.9
s2	1	8.9
s3	0	3.4

The bird performs the following action **from the initial state s1:**
Action= descent, Reward= -5.

Update the Q-value table accordingly for the episode.
Enter the modified **Q(s,a)** following **the event**.

Multiple choice 1 Question 40

a. 0 d.-12.98

b. -1.72 e. 2.78

c. 4.1 f. 1

41. Backpropagation is fundamentally based on:

a. Eigenvalue decomposition

b. The central limit theorem

c. The Fourier transform

d. The chain rule of calculus

---

42. Early stopping is a regularization technique that:

a. Stops training when the learning rate becomes small.

b. Stops training when training loss reaches zero.

c. Stops training when validation performance no longer improves.

d. Stops training after a fixed number of epochs.

---

43. What primarily distinguishes a deep neural network from a shallow one?

a. It only uses linear units

b. It uses non-linear activations

c. There are more neurons in the input layer

d. It has more than one hidden layer

---

44. Dropout primarily helps to:

a. Ensure that all neurons have the same activation

b. Increase training speed

c. Completely prevent vanishing gradients

d. Reduce overfitting by randomly deactivating neurons during training

---

45. How does batch normalization help in training?

a. It increases the problem of vanishing gradients

b. It accelerates convergence and stabilizes network training

c. It makes the training process sensitive to the initial weight initialization

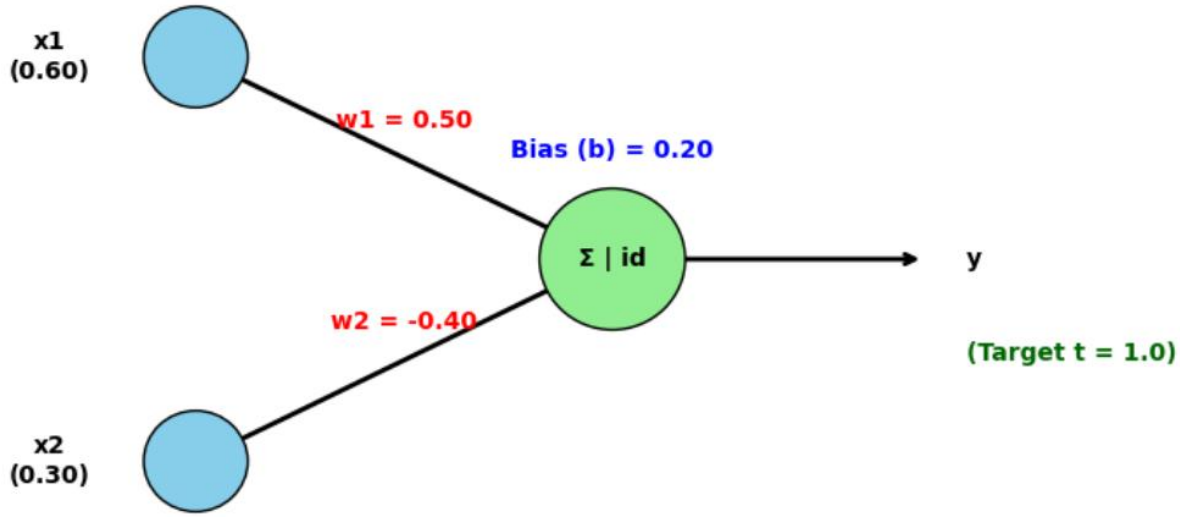
d. It randomizes the weights

e. It randomly deactivates layers

f. It increases the chance of overfitting

---

46. A modified perceptron shown in the figure below is utilized for a regression task.



A.) The perceptron uses a linear activation function. Compute the activated output of the perceptron:

a. 0.38 b. -0.85

c. -0.62 d. 0.12

e. 0.78 f. -0.26

B.) Compute the gradient of the error with respect to w1, given that we use mean squared error (MSE): $E = \frac{1}{2} \cdot (t - y)^2$ and the learning rate is: $\alpha = 0.2$

a. 0.1452 b. -0.0333

c. -0.2906 d. 0.3438

e. -0.3720 f. 0.6233

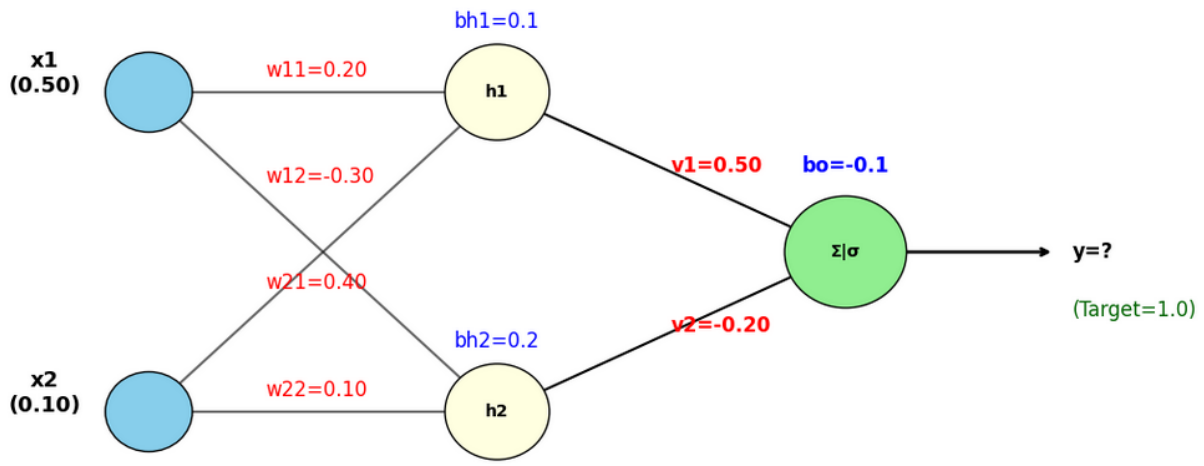
C.) Update the weight w1 and provide its new value:

a. 0.3545 b. 0.6464

c. -0.2355 d. 0.5372

e. -0.4879 f. 0.5744

47. A classification neural network is shown in the figure below.



A.) Both the hidden layer and the output layer use the sigmoid activation function. Calculate the activated output of the network.

a. 0.5192 b. 0.0864 c. 0.6251

d. 0.2846 e.-0.0042 f. 0.7462

B.) Calculate the gradient of the error with respect to (v1), given that:

- we use binary cross-entropy loss, whose derivative is:

$$\frac{\partial E}{\partial y} = \frac{y - t}{y(1 - y)}$$

- the derivative of the sigmoid function is:

$$\frac{\partial y}{\partial z} = y(1 - y)$$

- the learning rate is: $\alpha = 0.5$

a. -0.7464 b. -0.2691 c. 0.3921

d. -0.1147 e. 0.3374 f. 0.5333

C.) Update the weight v1 and provide its new value:

a. -0.66 b. 0.3173 c. -0.4303

d. 0.6346 e. 0.3313 f. 0.3655