Misurata University - Faculty of Engineering
Artificial intelligence course, department of E&E Engineering
Final Exam

Lec: Ali Abdulshahed

Date: 20/06/2026

Time: 150 minutes

Answer the Following Questions

Q1 (20 marks):

A feedforward artificial neural network, shown in Figure 1, is being trained using the Backpropagation algorithm. The available training sample consists of the following inputs and desired output:

x1 = 0.18, x2 = 0.54, yd = 0.90

For the neural network architecture shown in Figure 1, the current weights and biases are:

Hidden layer:

$$W = \begin{bmatrix} 0.9 & 1.35 \\ 0.45 & 1.8 \end{bmatrix} \qquad b = \begin{bmatrix} 0.9 \\ -0.45 \end{bmatrix}$$

Output layer:

$$W = [2.7 \quad 1.8] \qquad b = [-3.6]$$

The topology of the neural network is then modified as shown in Figure 2. For the modified architecture, the current weights and biases are:

Hidden layer:

$$W = \begin{bmatrix} 0.9 & 1.35 \\ 0.45 & 1.8 \\ 2.25 & 2.7 \end{bmatrix} \qquad b = \begin{bmatrix} 0.9 \\ -0.45 \\ 1.35 \end{bmatrix}$$

Output layer:

$$W = [2.7 \quad 1.8 \quad 0.9] \qquad b = [-3.6]$$

Assume that the Sigmoid activation function is used in all neurons.

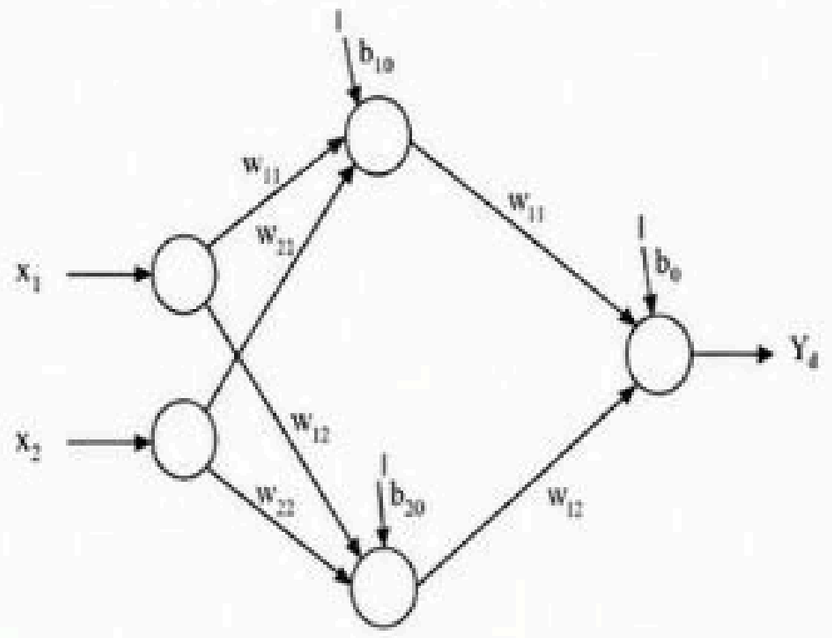

Figure 1

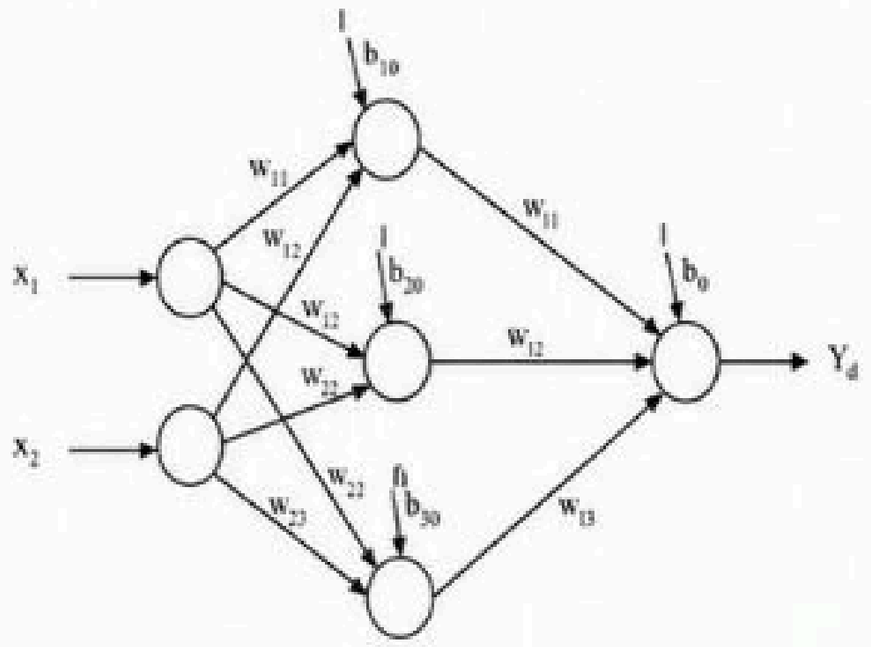

Figure 2

(a) Compute the outputs of the two neural networks shown in Figure 1 and Figure 2. Compare the results in terms of how close each output is to the desired output yd = 0.90.

(b) For the network shown in Figure 1, perform one training iteration using the Backpropagation algorithm. Determine the updated values of all weights and biases, assuming a learning rate of 0.10.

(c) Based on your result in part (b), determine whether the training iteration improved the accuracy of the network. Support your answer by comparing the error before and after training.

Q2 (15 marks):

Consider a 3 × 4 grid-world problem with the following assumptions:

- Let $s_{i,j}$ denote the position in row i and column j.
- s_{11} is the initial state.
- There is a wall at s_{23}.
- s_{24} and s_{34} are goal states.
- The robot escapes the world upon reaching either goal state.
- There are four actions: up, down, left, and right.
- Each action is possible in every state.

The transition model $P(s' \mid s, a)$:

- An action achieves its intended effect with probability 0.7.
- An action leads to a 90-degree left turn with probability 0.1.
- An action leads to a 90-degree right turn with probability 0.2.
- If the robot bumps into a wall, it stays in the same square.

The reward function $R(s)$ is the reward of entering state s:

- $R(s_{24}) = -1$
- $R(s_{34}) = 1$
- For all other non-terminal states, $R(s) = -0.04$

Table 1: State-value estimates for the grid world

	1	2	3	4
1	0.705	0.655	0.611	0.388
2	0.762	X	0.660	-1
3	0.812	0.868	0.918	+1

(a) The robot is in state s_{21} and chooses the action Right. What is the probability that the robot remains in state s_{21}?

(b) Using the values shown in Table 1, identify the optimal action for state s_{33}. Justify your answer.

(c) Write the Bellman optimality equation and provide a clear explanation of each term used in the equation.

(d) Explain the role of the discount factor (gamma) in the Bellman equation.

(e) What is the difference between immediate reward and future value?

(f) Explain why this grid world can be modeled as an MDP.

Q3 (10 marks):

Write the full English names of the following Artificial Intelligence terms, then provide their Arabic meanings:
ML, DL, ANN, RAG, (RL), MDP, (NLP), BP, (LLM), MLP.

Q4 (15 marks):

Question: Self-Attention Using 2-Dimensional Embeddings

Consider the sentence: (Child Reads Book). Each token is represented by a two-dimensional embedding vector:

Token	Symbol	2-D Embedding Vector
Child	x1	[1.0, 0.5]
Reads	x2	[0.2, 1.0]
Book	x3	[0.8, -0.2]

The model uses the following matrices to generate Query (Q), Key (K), and Value (V) vectors:

W^Q	W^K	W^V
[0.6 0.2] [0.1 0.7]	[0.5 −0.1] [0.3 0.8]	[0.4 0.6] [0.2 0.5]

Focus on the second word, "Reads", represented by x_2. Answer the following:

(a) What is meant by an Embedding in Natural Language Processing?

(b) What is meant by Attention in language models?

(c) Briefly explain the roles of Query (Q), Key (K), and Value (V).

(d) Compute Q_2 for "Reads" using: $Q_2 = x_2\, W^Q$.

(e) Compute K_1, K_2, and K_3 using: $K_i = x_i\, W^K$.

(f) Compute V_1, V_2, and V_3 using: $V_i = x_i\, W^V$.

(g) Compute the scaled attention scores.

(h) Apply Softmax to convert the scores into attention weights.

(i) Compute the final attention output vector: Output2 = $y_1\, V_1 + y_2\, V_2 + y_3\, V_3$.

(j) Identify which word receives the highest attention from "Reads", and justify your answer numerically.

Note to Students: Show all calculation steps clearly.

Good Luck