Set A

CSE440: Natural Language Processing II - QUIZ 2

Date: 3 November 2025

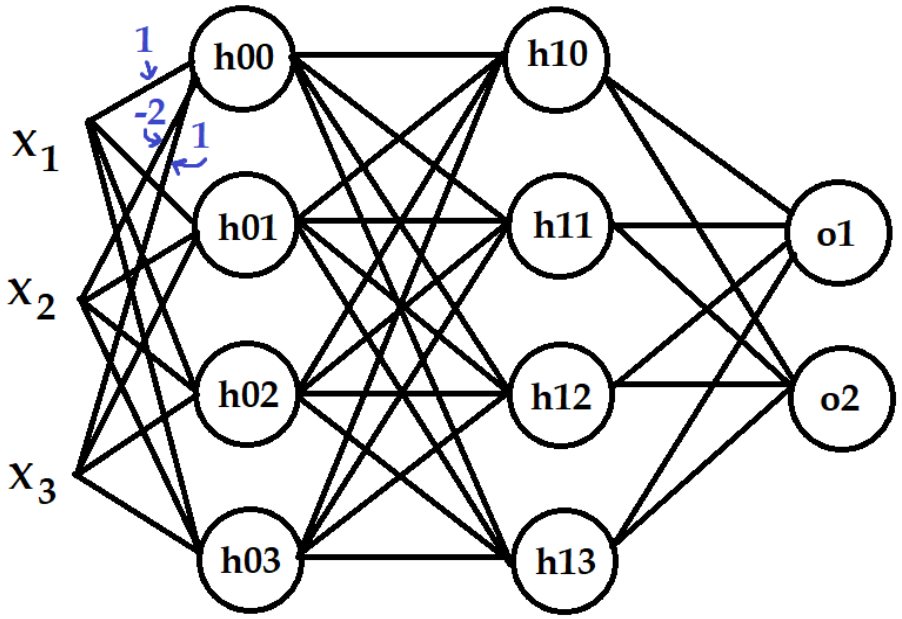
This is an MCQ quiz. Each question is worth 1. Write your answers serially on the side.

Name: **Student ID:** **Section:**

Answers:

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Consider the following neural network and answer question 1 to 3.
Here, the first hidden layer has RELU and second hidden layer and the final layer has sigmoid as activation function.



1. Which is the correct shape of weight matrices for the above network if input matrix is X = $[X_1,X_2,X_3]$, shape [1x3]:
   A. $W^{(1)}$ = [4x3], $W^{(2)}$ = [4x4], $W^{(3)}$ = [2x4]
   B. $W^{(1)}$ = [3x3], $W^{(2)}$ = [4x4], $W^{(3)}$ = [2x2]
   C. $W^{(1)}$ = [3x4], $W^{(2)}$ = [4x4], $W^{(3)}$ = [4x2]
   D. $W^{(1)}$ = [4x4], $W^{(2)}$ = [16x16], $W^{(3)}$ = [4x4]

2. Which of the following is the correct equation for h1* (h1* means output of all neurons of h1)?
   A. h1* = max(0, sigmoid($XW^{(1)}$) $W^{(2)}$)
   B. h1* = sigmoid(max(0.1 $XW^{(1)}$, $XW^{(1)}$) $W^{(2)}$)
   C. h1* = sigmoid(max(0 , $XW^{(1)}$) $W^{(2)}$)
   D. h1* = max(sigmoid(0 , $XW^{(1)}$) $XW^{(2)}$)

3. If X = [2, 1, -1] and $W^{(1,0)T}$ = [1,-2,1], what will be the output of the h00 neuron?
   A. -1
   B. 0
   C. 1
   D. None of the above

4. If a classifier performs well on training but poorly on unseen data, it is:
   A. Underfitted
   B. Overfitted
   C. Regularized correctly

D. Converged perfectly

5. In Naive Bayes, the term “naive” refers to:
   A. Its simplicity
   B. Conditional independence assumption among features
   C. The absence of training
   D. Its poor performance

6. Your Naive Bayes model predicts zero probability for a test sentence because a word never appeared for that specific class in training data. How to fix it?
   A. Drop the word
   B. Use smoothing
   C. Increase learning rate
   D. Apply dropout

7. A Logistic Regression has an output y_pred = 0.78. The actual label for the input data was 1. What is the cross Entropy loss for this prediction? Use log base 10 for your calculations.
   A. 0.107
   B. 0.248
   C. - 0.248
   D. - 0.107

8. Bag of Words ignores:
   A. Word order
   B. Word frequency
   C. Word presence
   D. Vocabulary size

9. Recall is defined as:
   A. TP / (TP + FP)
   B. TP / (TP + TN + FP + FN)
   C. TP / (TP + FN)
   D. (TP + TN) / (TP + TN + FP + FN)

10. Which metric is most useful when false positives are costly?
   A. Accuracy
   B. Precision
   C. Recall
   D. F1 Score

Set B

CSE440: Natural Language Processing II - QUIZ 2

Date: 3 November 2025

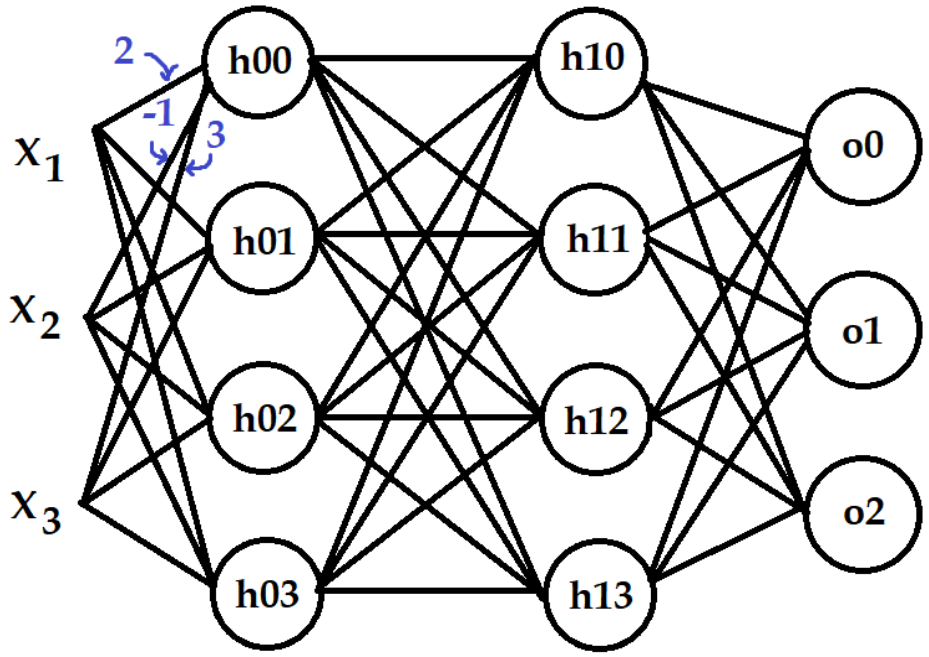
This is an MCQ quiz. Each question is worth 1. Write your answers serially on the side.

Name: **Student ID:** **Section:**

Answers:

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Consider the following neural network and answer question 1 to 3.
Here, the first hidden layer has sigmoid and second hidden layer and the final layer has RELU as activation function.



1. Which is the correct shape of weight matrices for the above network if input matrix is X = $[X_1,X_2,X_3]$, shape [1x3]:
   A. $W^{(1)}$ = [3x4], $W^{(2)}$ = [4x4], $W^{(3)}$ = [4x3]
   B. $W^{(1)}$ = [4x3], $W^{(2)}$ = [4x4], $W^{(3)}$ = [3x4]
   C. $W^{(1)}$ = [4x4], $W^{(2)}$ = [16x16], $W^{(3)}$ = [9x9]
   D. $W^{(1)}$ = [3x3], $W^{(2)}$ = [4x4], $W^{(3)}$ = [3x3]

2. If X = [2, 3, 1] and $W^{(1,0)T}$ = [2,-1,3], what will be the output of the h00 neuron?
   A. 0.98
   B. 0.017
   C. 4
   D. None of the above

3. Which of the following is the correct equation for h1*? (h1* means output of all neurons of h1)
   A. h1* = max(0, sigmoid($XW^{(1)}$) $W^{(2)}$)
   B. h1* = sigmoid(max(0.1 $XW^{(1)}$, $XW^{(1)}$) $W^{(2)}$)
   C. h1* = sigmoid(max(0 , $XW^{(1)}$) $W^{(2)}$)
   D. h1* = max(sigmoid(0 , $XW^{(1)}$) $XW^{(2)}$)

4. Precision is defined as:
   A. TP / (TP + FP)

B. TP / (TP + TN + FP + FN)
C. TP / (TP + FN)
D. (TP + TN) / (TP + TN + FP + FN)

5. Which metric is most useful when false negatives are costly?
A. Accuracy
B. Precision
C. Recall
D. F1 Score

6. Underfitting is manifested by:
A. Low training performance and low validation performance
B. Low validation performance but high training performance
C. Low testing performance but high training performance
D. None of the above

7. Your Naive Bayes model predicts zero probability for a test sentence because a word never appeared for that specific class in training data. How to fix it?
A. Increase learning rate
B. Apply dropout
C. Drop the word
D. Use smoothing

8. A Logistic Regression has an output y_pred = 0.34. The actual label for the input data was 1. What is the cross Entropy loss for this prediction? Use log base 10 for your calculations.
A. 0.468
B. - 0.468
C. - 1.07
D. 1.07

9. In Naive Bayes, the term “naive” refers to:
A. The absence of training
B. Its poor performance
C. Its simplicity
D. Conditional independence assumption among features

10. Bag of Words ignores:
A. Word presence
B. Vocabulary size
C. Word order
D. Word frequency

Answers

Set A

1 C
2 C
3 B
4 B
5 B
6 B
7 A
8 A
9 C
10 B

Set B

1 A
2 C
3 A
4 A
5 C
6 A
7 D
8 A
9 D
10 C