SUMMARY OF MARKING GUIDELINES:

- If you want you can use this as a convenient place to keep your overall instructions to your markers. This way, you can find these instructions years later when you are looking back at the exam.
  If you are seeing this and not a cover page, you have 'solutions' and 'markingspace' both set to true in your main `.tex` file.

Question 1. Binary Linear Classification [4 MARKS]

You are tasked with classifying a data set perfectly with a binary linear classification model.
We have provided a **partial** data set and the binary linear classification model below.

A partial data set

x_1	x_2	t
1	1	0
0	0	1
⋮	⋮	⋮

Binary Linear Classification Model

$$z = \mathbf{w}^T \mathbf{x}$$

$$y = \begin{cases} 1, & \text{if } z \geq 2 \\ 0, & \text{if } z < 2 \end{cases}$$

Your friend Halia plotted the data and came up with the decision boundary below.

$$x_2 = -0.5\, x_1 + 0.5$$

Derive the weights w_0, w_1, w_2 for the decision boundary.
Hint: Start by rewriting the decision boundary in the form: $a\, x_1 + b\, x_2 + c = 0$.

There is no need to show any work. However, if your answers are incorrect,
you can show work for the following to receive partial marks.

1. The model with your weights classifies the first data point ($x_1 = 1, x_2 = 1, t = 0$) correctly.
2. The model with your weights classifies the second data point ($x_1 = 0, x_2 = 0, t = 1$) correctly.

SOLUTION

The correct weights are

$$w_0 = 2.5, w_1 = -0.5, w_2 = -1$$

Since our model uses 2 as the threshold, let's derive weights for an equivalent model, where 5 is absorbed into w_0 and the threshold is 0. In other words,

$$\begin{aligned} & w_1 x_1 + w_2 x_2 + w_0 \geqslant 2 \\ & \Rightarrow w_1 x_1 + w_2 x_2 + (w_0 - 2) \geqslant 0 \\ & \Rightarrow w_1 x_1 + w_2 x_2 + w_0' \geqslant 0 \end{aligned}$$

where $w_0 = w_0' + 2$

We start by rewriting the decision boundary.

$$0.5x_1 + x_2 - 0.5 = 0$$

Next, we need to convert the equation to an inequality. To figure out the direction of the inequality, we will plug in the second data point to the left-hand side of the equation above. We get -0.5, which is less than 0. Since the second data point has $t = 1$, we need to flip all the signs in the equation above.

$$-0.5x_1 - x_2 + 0.5 \geqslant 0$$

Therefore, $w_1 = -0.5$, $w_2 = -1$, and $w_0' = 0.5$, which means $w_0 = 0.5 + 2 = 2.5$.

MARKING SCHEME:

- If all three weights are correct, give full marks.
- Otherwise, 1 mark if the model classifies the first data point correctly, and 1 mark if the model classifies the second data point correctly.

Question 2. Design a Small Neural Network by Hand [7 MARKS]

Consider a function represented by the truth table below.
The inputs are binary: $x_i \in \{0, 1\}$, where $i = 1, 2, 3$. The target is also binary: $t \in \{0, 1\}$.

x_1	x_2	x_3	t
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0

Table 1: Data Set for Neural Network

Part (a) Number of Hidden Units [1 MARK]

In the next part, you will design a 2-layer neural network to model this function.
Explain why **we need at most 2 units in the hidden layer** in no more than 2 sentences.

SOLUTION

There are two data points with $t = 1$. We can design 2 hidden units so that each hidden unit responds to exactly one of these two data points.

Question 2. (CONTINUED)

Part (b) Design a Small Neural Network [6 MARKS]

x_1	x_2	x_3	t
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	0

Table 2: A Copy of Table 1

Design a **two-layer** neural network to represent the function in Table 2 (a copy of Table 1). Every unit in the neural network should use the threshold activation function below.

$$\phi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The computations of the neural network are given below.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \mathbf{h} = \phi(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) \qquad y = \phi(\mathbf{W}^{(2)}\mathbf{h} + \mathbf{b}^{(2)})$$

SOLUTION

$$\mathbf{W}^{(1)} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \qquad \mathbf{b}^{(1)} = \begin{pmatrix} -1.5 \\ -0.5 \end{pmatrix}$$

$$\mathbf{W}^{(2)} = \begin{pmatrix} 1 & 1 \end{pmatrix} \qquad \mathbf{b}^{(2)} = \begin{pmatrix} -1.5 \end{pmatrix}$$

Explain what each hidden/output unit represents.

- The first hidden unit represents: $\neg x_1 \wedge x_2 \wedge x_3$
- The second hidden unit represents: $x_1 \wedge \neg x_2 \wedge \neg x_3$
- The output unit represents: $h_1 \vee h_2$

MARKING SCHEME:

- 2 marks for $\mathbf{W}^{(1)}$ and $\mathbf{b}^{(1)}$.
- 1 mark for $\mathbf{W}^{(2)}$ and $\mathbf{b}^{(2)}$.
- 3 marks for the meanings of the hidden/output units. 1 mark each.

Question 3. Backpropagation [7 MARKS]

Consider the 3-layer neural network in Figure 1 below.

The input layer has $D+1$ units. The first hidden layer has J units and uses a univariate activation function called α. That is, $h_j^{(1)} = \alpha(z_j^{(1)})$ where $z_j^{(1)}$ denotes the input to α. The second hidden layer has K units and uses a multivariate activation function called β. That is, $h_j^{(2)} = \beta(z^{(2)})$ where $z^{(2)}$ denotes the input vector to β. Finally, the output y is a scalar. We will use the cross-entropy loss function.

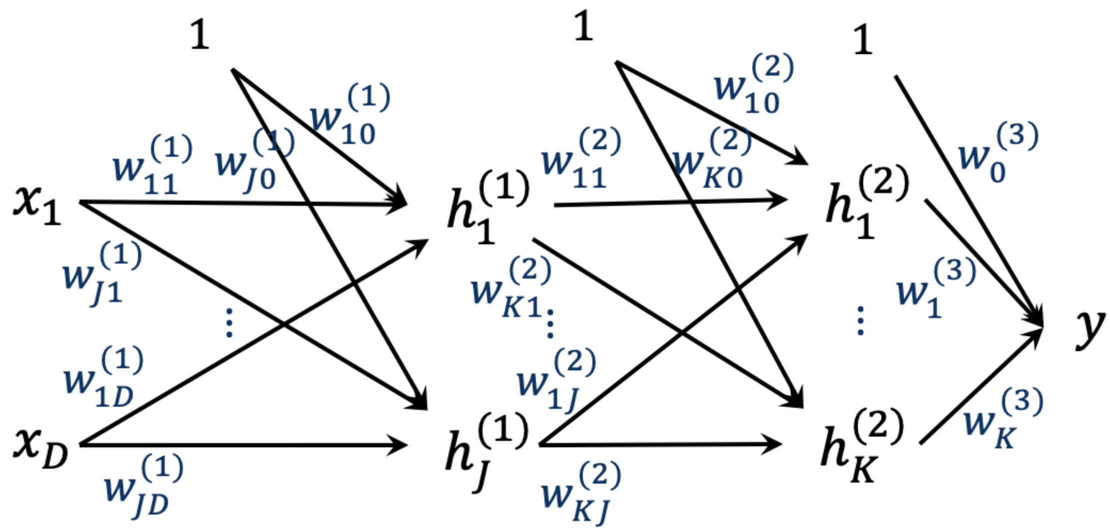

Figure 1: Neural Network for Backpropagation

Part (a) Computation Graph [4 MARKS]

Draw the vectorized computation graph below. **Label any matrix/vector with its dimensions.**
For example, the input vector has dimensions $(D+1) \times 1$.

SOLUTION

The vectorized computation graph is in Figure 2.

MARKING SCHEME:

- 2 marks for the nodes and connections in the computation graph.
- A common mistake: deduct 1 mark if the graph is missing one or both z.
- 2 marks for the dimensions. We will accept both J and $J+1$ (and K and $K+1$) as the correct dimensions.

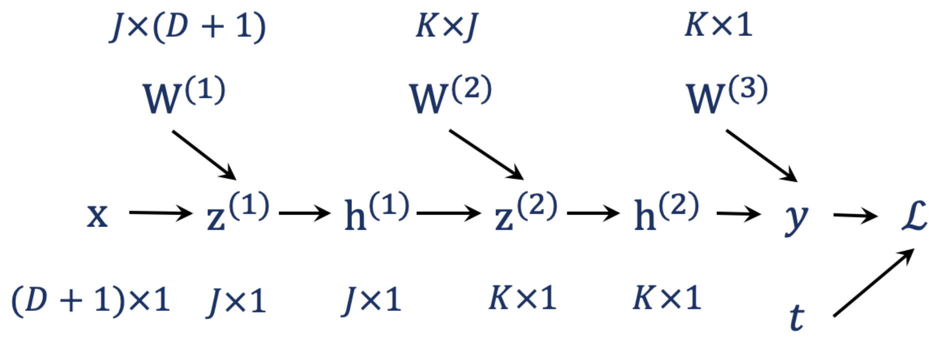

Figure 2: Solution: Vectorized Computation Graph

Question 3. (CONTINUED)

Part (b) [1 MARK]

Write the **vectorized** expression for computing $\mathbf{z}^{(2)}$ in terms of previously computed quantities, during the **forward** pass of the backpropagation algorithm. **Label** each quantity with its **dimensions**.

SOLUTION

$$\mathbf{z}^{(2)} = \mathbf{W}^{(2)}\mathbf{h}^{(1)}$$

$\mathbf{z}^{(2)}$ is $K \times 1$, $\mathbf{W}^{(2)}$ is $K \times J$, and $\mathbf{h}^{(1)}$ is $J \times 1$.

MARKING SCHEME:

- All or nothing
- We will accept both J and $J + 1$ (and K and $K + 1$) as the correct dimensions.

Part (c) [1 MARK]

Write the **vectorized** expression for the error signal $\overline{\mathbf{z}^{(1)}}$ in terms of previously computed quantities, during the **backward** pass of the backpropagation algorithm. **Label** each quantity with its **dimensions**.

You can use $\alpha'(\mathbf{v})$ to represent a vector of the derivatives of α applied to each component of $\mathbf{v}$ separately.

SOLUTION

$$\overline{\mathbf{z}^{(1)}} = \overline{\mathbf{h}^{(1)}} \odot \alpha'(\mathbf{z}^{(1)})$$

$\mathbf{z}^{(1)}$ is $J \times 1$, $\overline{\mathbf{h}^{(1)}}$ is $J \times 1$, and $\alpha'(\mathbf{z}^{(1)})$ is $J \times 1$.

MARKING SCHEME:

- All or nothing
- We will accept both J and $J + 1$ (and K and $K + 1$) as the correct dimensions.

Part (d) [1 MARK]

Write the **vectorized** expression for the error signal $\overline{\mathbf{h}^{(1)}}$ in terms of previously computed quantities, during the **backward** pass of the backpropagation algorithm. **Label** each quantity with its **dimensions**.

SOLUTION

$$\overline{\mathbf{h^{(1)}}} = \mathbf{W^{(2)}}^{\mathbf{T}}\overline{\mathbf{z^{(2)}}}$$

$\overline{\mathbf{h^{(1)}}}$ is $J \times 1$, $\mathbf{W^{(2)}}^{\mathbf{T}}$ is $J \times K$, and $\overline{\mathbf{z^{(2)}}}$ is $K \times 1$.

MARKING SCHEME:

- All or nothing
- We will accept both J and $J + 1$ (and K and $K + 1$) as the correct dimensions.

Question 4. Naive Bayes [8 MARKS]

Part (a) Learn The Model [5 MARKS]

Data Point	A	B	Class
1	L	10	Y
2	M	10	Y
3	M	10	Y
4	S	10	Y
5	S	20	Y
6	L	20	N
7	L	20	N
8	L	20	N
9	M	10	N
10	M	20	N
11	S	20	N

Table 3: Data Set for Naive Bayes

Learn a Naive Bayes Model for the data set in Table 3 using **Maximum Likelihood Estimation**. **There is no need to show any work.**

SOLUTION

The prior probabilities of the class label:

$$P(Class = Y) = 5/11$$

The conditional probabilities involving A:

$$P(A = L|Class = Y) = 1/5, P(A = M|Class = Y) = 2/5$$
$$P(A = L|Class = N) = 1/2, P(A = M|Class = N) = 1/3$$

The conditional probabilities involving B:

$$P(B = 10|Class = Y) = 4/5$$
$$P(B = 10|Class = N) = 1/6$$

MARKING SCHEME:

- 1 mark for the prior probability.
- 2 marks for the conditional probabilities involving A. Deduct 1 mark for each mistake.
- 2 marks for the conditional probabilities involving B. Deduct 1 mark for each mistake.

Question 4. (CONTINUED)

Part (b) Predict the Class [3 MARKS]

Data Point	A	B	Class
New	M	20	?

Table 4: New Data Point

Suppose we are given a Naive Bayes model for the dataset in Table 3.
Describe how we can predict the class for the new data point in Table 4.

In your answer, include the two items below.

1. Write the formulas for the probabilities we need to calculate.
   Make sure that every probability in the formula is available in the Naive Bayes model.

2. Describe how we will determine which class label to predict.
   Assume that there will be no ties.

You do not need to perform the calculations.

SOLUTION

We need to calculate the two probabilities below.

$$P(class = Y)P(A = M|class = Y)P(B = 20|class = Y)$$
$$P(class = N)P(A = M|class = N)P(B = 20|class = N)$$

We will predict the class corresponding to the larger of the two probabilities.

MARKING SCHEME:

- 2 marks for writing the two formulas.
- 1 mark for explaining how to pick the class.