SC4001 Exam 2017-2018 Sem 1

1. (a) Consider a layer of K neurons learning using the Hebbian learning rule. The Hebbian rule states that a weight of a neuron is changed proportional to the product of the input and the output of the neuron. Let $\boldsymbol{w_k}$ denote the weight of the kth neuron and $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots \boldsymbol{w}_K)$ be the weight matrix connected to the layer.

   (i) If the output of the kth neuron for an input x is y_k, write Hebbian learning equations for weight vector $\boldsymbol{w_k}$ and weight matrix $\boldsymbol{W}$.

   (6 marks)

   (ii) Given a set of input patterns $\{\boldsymbol{x}_p\}_{p=1}^{P}$, write the Hebbian batch learning equation for weight matrix $\boldsymbol{W}$.

   (4 marks)

1. (b) A three-layer discrete perceptron network receives 2-dimensional inputs $(x_1, x_2)^2 \in \boldsymbol{R}^2$ and has four hidden neurons and one output neuron. The shaded region of Figure Q1 shows the input space for which the output of the neuron, $y = 1$. Draw the perceptron network, clearly indicating the values of weights and biases.

   (15 marks)

Answer

f

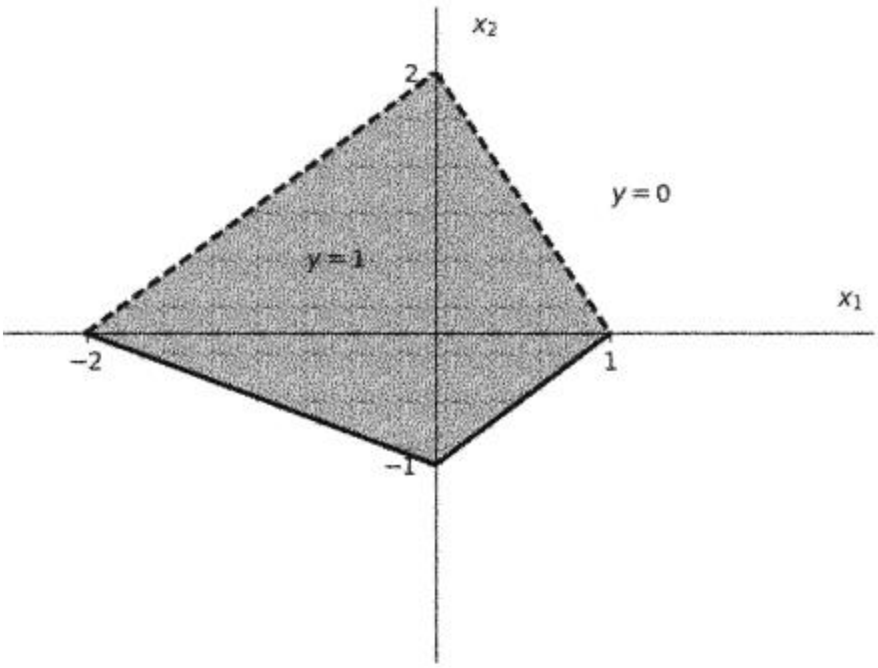


Figure Q1

1. (b) cont

Answer

Line 1 (passing through (-2, 0) and (0, 2))
Line 2 (passing through (0, 2) and (1, 0))
Line 3 (passing through (1, 0) and (0, -1))
Line 4 (passing through (0, -1) and (-2, 0))

shaded above the line is +1 > 0
shaded below the line is -1 ≤ 0

Line 1 (passing through (-2, 0) and (0, 2))

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{2 - 0}{0 - (-2)} = 1$$

$$y = mx + c$$
$$0 = 1(-2) + c$$
$$c = 2$$

$$x_2 = (1)x_1 + 2$$
$$x_1 - x_2 + 2 = 0$$

since the shaded-region below the line,
$u_1 = x_1 - x_2 + 2$

Line 2 (passing through (0, 2) and (1, 0))

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0 - 2}{1 - 0} = \frac{-2}{1} = -2$$

$$y = mx + c$$
$$2 = -2(0) + c$$
$$c = 2$$

$$x_2 = -2x_1 + 2$$
$$2x_1 + x_2 - 2 = 0$$

since the shaded-region below the line,
$u_2 = 2x_1 + x_2 - 2$

Line 3 (passing through (1, 0) and (0, -1))

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(-1) - 0}{0 - 1} = \frac{-1}{-1} = 1$$

$$y = mx + c$$
$$0 = 1(1) + c$$
$$c = -1$$

$$x_2 = (1)x_1 - 1$$
$$x_1 - x_2 - 1 = 0$$

since the shaded-region above the line,
$u_3 = -x_1 + x_2 + 1$

Line 4 (passing through (0, -1) and (-2, 0))

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0 - (-1)}{(-2) - 0} = \frac{1}{-2} = -\frac{1}{2}$$

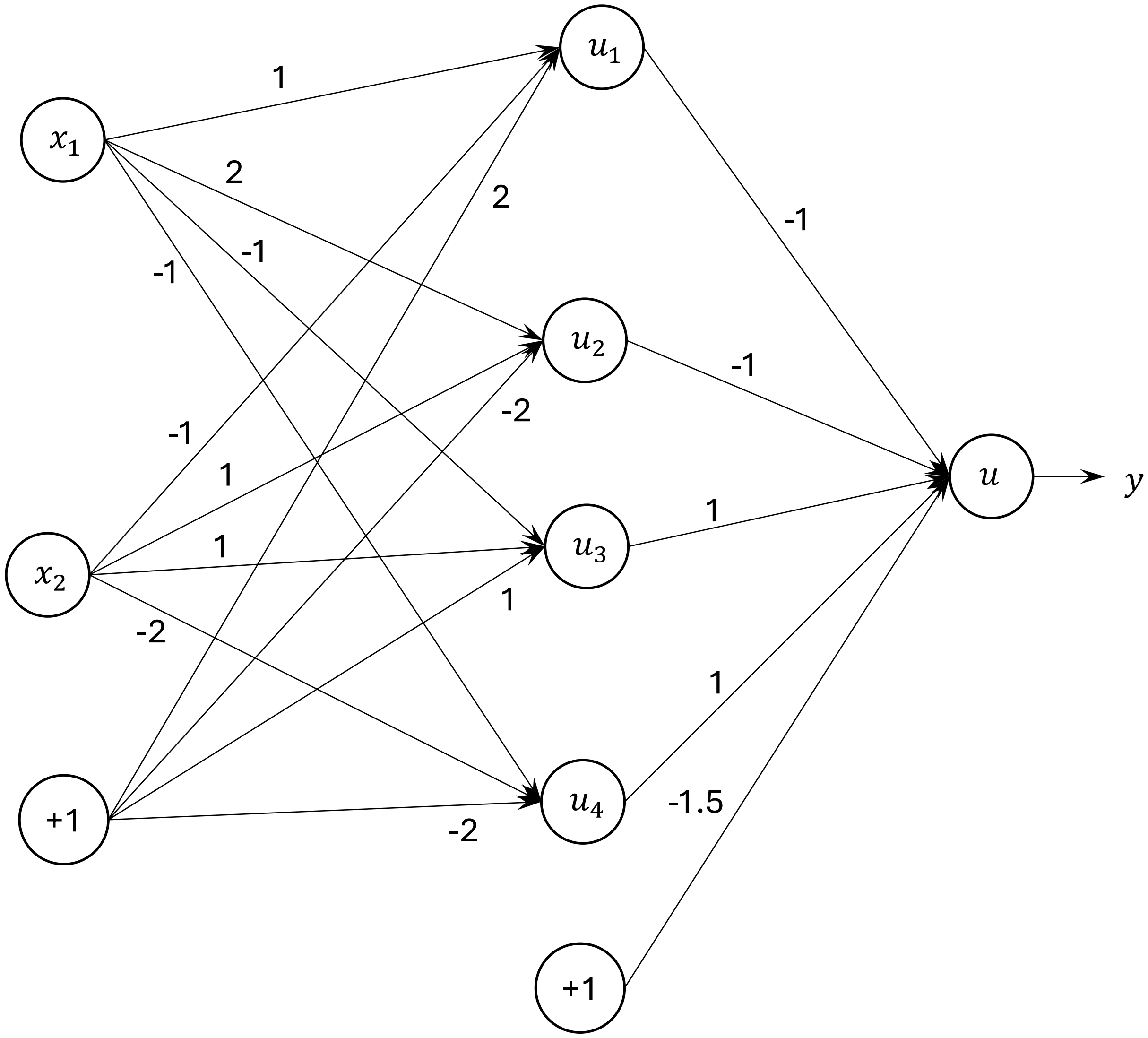
$$y = mx + c$$
$$-1 = -\frac{1}{2}(0) + c$$
$$c = -1$$

$$x_2 = \left(-\frac{1}{2}\right)x_1 - 1$$
$$2x_2 = -x_1 - 2$$
$$x_1 + 2x_2 + 2 = 0$$

since the shaded-region above the line,
$u_4 = -x_1 - 2x_2 - 2$

1. (b) cont

Answer



$$u = y_1 - y_2 + y_3 - y_4 - 1.5$$

2. The three-layer feedforward network shown in Figure Q2 receives 2-dimensional inputs $(x_1, x_2)^T \in \boldsymbol{R}^2$ and produces an output label $y \in \{0, 1\}$. The hidden-layer is a perceptron layer and the output-layer is a softmax layer, and each layer has two neurons. The weights and biases of the network are initialized as indicated in the figure.

The network is to train to produce outputs $y_1 = 0$ and $y_2 = 1$ for input patterns $\boldsymbol{x}_1 = \begin{pmatrix} 1.5 \\ 2.0 \end{pmatrix}$ and $\boldsymbol{x}_2 = \begin{pmatrix} -2.0 \\ -0.5 \end{pmatrix}$, respectively, using gradient descent learning. The learning factor $\alpha = 0.4$.

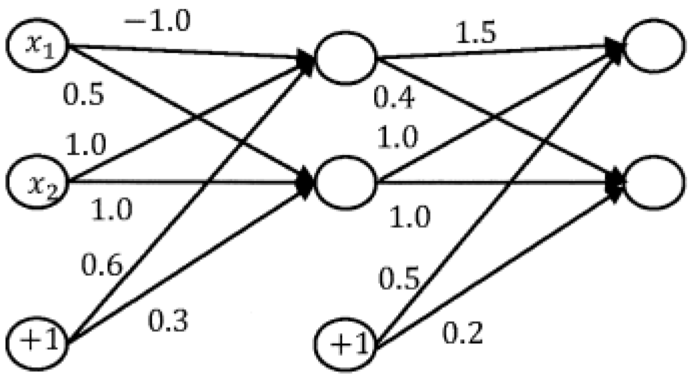


Figure Q2

For one iteration of batch gradient descent:

2. (a) Write the initial weight matrices and bias vectors connected to the hidden and output layers.

(4 marks)

Answer

$$W = \begin{pmatrix} -1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix} \qquad b = \begin{pmatrix} 0.6 \\ 0.3 \end{pmatrix} \quad V = \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 1.0 \end{pmatrix} \qquad c = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix}$$

2. (b) Write the input data matrix and the output target vector.

(2 marks)

Answer

$$\boldsymbol{x}_1 = \begin{pmatrix} 1.5 \\ 2.0 \end{pmatrix} \quad \boldsymbol{x}_2 = \begin{pmatrix} -2.0 \\ -0.5 \end{pmatrix}, \quad d_1 = 0 \quad d_2 = 1$$

$$X = x^T = \begin{pmatrix} 1.5 & 2.0 \\ -2.0 & 0.5 \end{pmatrix}, \qquad D = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

2. (c) Find the synaptic inputs and activations of the hidden and output layers.

(6 marks)

Answer

$$W = \begin{pmatrix} -1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix}, \quad b = \begin{pmatrix} 0.6 \\ 0.3 \end{pmatrix}, \quad V = \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 0.5 \end{pmatrix}, \quad c = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} 1.5 & 2.0 \\ -2.0 & -0.5 \end{pmatrix} \qquad B = b^T = (0.6 \quad 0.3), \quad C = c^T = (0.5 \quad 0.2)$$

Synaptic input to hidden-layer,

$$\boldsymbol{Z} = \boldsymbol{XW} + \boldsymbol{B} = \begin{pmatrix} 1.5 & 2.0 \\ -2.0 & -0.5 \end{pmatrix} \begin{pmatrix} -1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix} + \begin{pmatrix} 0.6 & 0.3 \\ 0.6 & 0.3 \end{pmatrix}$$

$$= \begin{pmatrix} (\ 1.5)(-1) + (\ 2.0)(1) & (\ 1.5)(0.5) + (\ 2.0)(1) \\ (-2.0)(-1) + (-0.5)(1) & (-2.0)(0.5) + (-0.5)(1) \end{pmatrix} + \begin{pmatrix} 0.6 & 0.3 \\ 0.6 & 0.3 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5 & 2.75 \\ 1.5 & -1.50 \end{pmatrix} + \begin{pmatrix} 0.6 & 0.3 \\ 0.6 & 0.3 \end{pmatrix} = \begin{pmatrix} 0.5 + 0.6 & 2.75 + 0.3 \\ 1.5 + 0.6 & -1.50 + 0.3 \end{pmatrix}$$

$$= \begin{pmatrix} 1.1 & 3.05 \\ 2.1 & -1.20 \end{pmatrix}$$

Output of the hidden layer,

$$\boldsymbol{H} = g(\boldsymbol{Z}) = \frac{1}{1 + e^{-Z}} = \begin{pmatrix} 0.75 & 0.95 \\ 0.89 & 0.23 \end{pmatrix}$$

Synaptic input to output-layer,

$$\boldsymbol{U} = \boldsymbol{HV} + \boldsymbol{C} = \begin{pmatrix} 0.75 & 0.95 \\ 0.89 & 0.23 \end{pmatrix} \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 0.5 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.2 \end{pmatrix}$$

$$= \begin{pmatrix} (0.75)(1.5) + (0.95)(1) & (0.75)(0.4) + (0.95)(0.5) \\ (0.89)(1.5) + (0.23)(1) & (0.89)(0.4) + (0.23)(0.5) \end{pmatrix} + \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.2 \end{pmatrix}$$

$$= \begin{pmatrix} 2.08 & 0.78 \\ 1.57 & 0.47 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.2 \end{pmatrix} = \begin{pmatrix} 2.08 + 0.5 & 0.78 + 0.2 \\ 1.57 + 0.5 & 0.47 + 0.2 \end{pmatrix}$$

$$= \begin{pmatrix} 2.58 & 0.98 \\ 2.07 & 0.67 \end{pmatrix}$$

Output layer activation

$$f(\boldsymbol{U}) = \frac{e^{\boldsymbol{U}}}{\sum_{k=1}^{K} e^{\boldsymbol{U}_k}} = \begin{pmatrix} \dfrac{e^{2.58}}{e^{2.58} + e^{0.98}} & \dfrac{e^{0.98}}{e^{2.58} + e^{0.98}} \\ \dfrac{e^{2.07}}{e^{2.07} + e^{0.67}} & \dfrac{e^{2.07}}{e^{2.07} + e^{0.67}} \end{pmatrix} = \begin{pmatrix} 0.83 & 0.17 \\ 0.80 & 0.20 \end{pmatrix}$$

Summary

- Synaptic input to hidden-layer: $\boldsymbol{Z} = \begin{pmatrix} 1.1 & 3.05 \\ 2.1 & -1.20 \end{pmatrix}$
- Output of the hidden layer: $\boldsymbol{H} = \begin{pmatrix} 0.75 & 0.95 \\ 0.89 & 0.23 \end{pmatrix}$
- Synaptic input to output-layer: $\boldsymbol{U} = \begin{pmatrix} 2.58 & 0.98 \\ 2.07 & 0.67 \end{pmatrix}$
- Output layer activation $f(\boldsymbol{U}) = \begin{pmatrix} 0.83 & 0.17 \\ 0.80 & 0.20 \end{pmatrix}$

2. (d) Find the error matrices, Δ_s, at the hidden and output layers.

(8 marks)

Answer

$$\boldsymbol{X} = \begin{pmatrix} 1.5 & 2.0 \\ -2.0 & -0.5 \end{pmatrix} \quad \boldsymbol{D} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad f(\boldsymbol{U}) = \begin{pmatrix} 0.83 & 0.17 \\ 0.80 & 0.20 \end{pmatrix} \quad \boldsymbol{H} = \begin{pmatrix} 0.75 & 0.95 \\ 0.89 & 0.23 \end{pmatrix}$$

Targets as a one hot matrix:

$$V = \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 0.5 \end{pmatrix}$$

$$\boldsymbol{K} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Gradient $\nabla_{\boldsymbol{U}} J$,

$$\nabla_U J = -\left(\boldsymbol{K} - f(\boldsymbol{U})\right) = -\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.83 & 0.17 \\ 0.80 & 0.20 \end{pmatrix}\right) = -\begin{pmatrix} 1 - 0.83 & 0 - 0.17 \\ 0 - 0.80 & 1 - 0.20 \end{pmatrix}$$

$$= -\begin{pmatrix} 0.17 & -0.17 \\ -0.80 & 0.80 \end{pmatrix} = \begin{pmatrix} -0.17 & 0.17 \\ 0.80 & -0.80 \end{pmatrix}$$

Gradient $\nabla_{\boldsymbol{Z}} J$,

$$g'(\boldsymbol{Z}) = \boldsymbol{H} \cdot (\boldsymbol{1} - \boldsymbol{H}) = \begin{pmatrix} 0.75 & 0.95 \\ 0.89 & 0.23 \end{pmatrix} \cdot \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0.75 & 0.95 \\ 0.89 & 0.23 \end{pmatrix}\right)$$

$$= \begin{pmatrix} 0.75(1 - 0.75) & 0.95(1 - 0.95) \\ 0.89(1 - 0.89) & 0.23(1 - 0.23) \end{pmatrix} = \begin{pmatrix} 0.19 & 0.05 \\ 0.10 & 0.18 \end{pmatrix}$$

$$\nabla_Z J = \nabla_u J V^T \cdot g'(\boldsymbol{Z}) = \begin{pmatrix} -0.17 & 0.17 \\ 0.80 & -0.80 \end{pmatrix} \begin{pmatrix} 1.5 & 1.0 \\ 0.4 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} 0.19 & 0.05 \\ 0.10 & 0.18 \end{pmatrix}$$

$$= \begin{pmatrix} (-0.17)(1.5) + (\ \ 0.17)(0.4) & (-0.17)(1.0) + (\ \ 0.17)(0.5) \\ (\ \ 0.80)(1.5) + (-0.80)(0.4) & (\ \ 0.80)(1.0) + (-0.80)(0.5) \end{pmatrix} \cdot \begin{pmatrix} 0.19 & 0.05 \\ 0.10 & 0.18 \end{pmatrix}$$

$$= \begin{pmatrix} -0.19 & -0.09 \\ 0.88 & 0.40 \end{pmatrix} \cdot \begin{pmatrix} 0.19 & 0.05 \\ 0.10 & 0.18 \end{pmatrix} = \begin{pmatrix} -0.04 & 0.004 \\ 0.09 & 0.070 \end{pmatrix}$$

Output layer:

$$\nabla_V J = H^T \nabla_U J = \begin{pmatrix} 0.75 & 0.89 \\ 0.95 & 0.23 \end{pmatrix} \begin{pmatrix} -0.17 & 0.17 \\ 0.80 & -0.80 \end{pmatrix}$$

$$= \begin{pmatrix} (0.75)(-0.17) + (0.89)(0.80) & (0.75)(0.17) + (0.89)(-0.80) \\ (0.95)(-0.17) + (0.23)(0.80) & (0.95)(0.17) + (0.23)(-0.80) \end{pmatrix}$$

$$= \begin{pmatrix} 0.58 & -0.58 \\ 0.02 & -0.02 \end{pmatrix}$$

$$\nabla_c J = (\nabla_U J)^{\boldsymbol{T}} 1_p = \begin{pmatrix} -0.17 & 0.17 \\ 0.80 & -0.80 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} (-0.17)(1) + (\ \ 0.17)(1) \\ (\ \ 0.80)(1) + (-0.80)(1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

2. (d) cont

Answer

Hidden layer:

$$\boldsymbol{X} = \begin{pmatrix} 1.5 & 2.0 \\ -2.0 & -0.5 \end{pmatrix} \qquad \nabla_Z J = \begin{pmatrix} -0.04 & 0.004 \\ 0.09 & 0.070 \end{pmatrix}$$

$$\nabla_W J = X^T(\nabla_Z J) = \begin{pmatrix} 1.5 & -2.0 \\ 2.0 & -0.5 \end{pmatrix}\begin{pmatrix} -0.04 & 0.004 \\ 0.09 & 0.070 \end{pmatrix}$$

$$= \begin{pmatrix} (1.5)(-0.04) + (-2.0)(0.09) & (1.5)(0.004) + (-2.0)(0.07) \\ (2.0)(-0.04) + (-0.5)(0.09) & (2.0)(0.004) + (-0.5)(0.07) \end{pmatrix}$$

$$= \begin{pmatrix} -0.24 & -0.13 \\ -0.13 & -0.03 \end{pmatrix}$$

$$\nabla_b J = (\nabla_{\boldsymbol{Z}} J)^{\boldsymbol{T}} 1_p = \begin{pmatrix} -0.04 & 0.09 \\ 0.004 & 0.07 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} (-0.040)(1) + (0.09)(1) \\ (\ 0.004)(1) + (0.07)(1) \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.07 \end{pmatrix}$$

<u>Summary</u>

Output layer

$$\nabla_V J = \begin{pmatrix} 0.58 & -0.58 \\ 0.02 & -0.02 \end{pmatrix}$$

$$\nabla_c J = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Hidden layer

$$\nabla_W J = \begin{pmatrix} -0.24 & -0.13 \\ -0.13 & -0.03 \end{pmatrix}$$

$$\nabla_b J = \begin{pmatrix} 0.05 \\ 0.07 \end{pmatrix}$$

2. (d) Find new weight matrices and bias vectors.

(5 marks)

Answer

$$W = \begin{pmatrix} -1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix}, \quad b = \begin{pmatrix} 0.6 \\ 0.3 \end{pmatrix}, \quad V = \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 0.5 \end{pmatrix}, \quad c = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} \qquad \alpha = 0.4$$

$$\nabla_V J = \begin{pmatrix} 0.58 & -0.58 \\ 0.02 & -0.02 \end{pmatrix} \quad \nabla_c J = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \nabla_W J = \begin{pmatrix} -0.24 & -0.13 \\ -0.13 & -0.03 \end{pmatrix} \; \nabla_b J = \begin{pmatrix} 0.05 \\ 0.07 \end{pmatrix}$$

$$\boldsymbol{V} = \boldsymbol{V} - \alpha \nabla_V J = \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 0.5 \end{pmatrix} - 0.4 \begin{pmatrix} 0.58 & -0.58 \\ 0.02 & -0.02 \end{pmatrix} = \begin{pmatrix} 1.5 & 0.4 \\ 1.0 & 0.5 \end{pmatrix} - \begin{pmatrix} 0.23 & -0.23 \\ 0.02 & -0.01 \end{pmatrix}$$

$$= \begin{pmatrix} 1.27 & 0.63 \\ 0.98 & 0.51 \end{pmatrix}$$

$$\boldsymbol{c} = \boldsymbol{c} - \alpha \nabla_c J = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} - 0.4 \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix}$$

$$\boldsymbol{W} = W - \alpha \nabla_W J = \begin{pmatrix} -1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix} - 0.4 \begin{pmatrix} -0.24 & -0.13 \\ -0.13 & -0.03 \end{pmatrix}$$

$$= \begin{pmatrix} -1.0 & 0.5 \\ 1.0 & 1.0 \end{pmatrix} - \begin{pmatrix} -0.10 & -0.05 \\ -0.05 & -0.01 \end{pmatrix} = \begin{pmatrix} -0.90 & 0.55 \\ 1.05 & 1.01 \end{pmatrix}$$

$$\boldsymbol{b} = \boldsymbol{b} - \alpha \nabla_b J = \begin{pmatrix} 0.6 \\ 0.3 \end{pmatrix} - 0.4 \begin{pmatrix} 0.05 \\ 0.07 \end{pmatrix} = \begin{pmatrix} 0.6 \\ 0.3 \end{pmatrix} - \begin{pmatrix} 0.02 \\ 0.03 \end{pmatrix} = \begin{pmatrix} 0.58 \\ 0.27 \end{pmatrix}$$

<u>Summary</u>

Updated $V = \begin{pmatrix} 1.27 & 0.63 \\ 0.98 & 0.51 \end{pmatrix}$

Updated $\boldsymbol{c} = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix}$

Updated $W = \begin{pmatrix} -0.90 & 0.55 \\ 1.05 & 1.01 \end{pmatrix}$

Updated $\boldsymbol{b} = \begin{pmatrix} 0.58 \\ 0.27 \end{pmatrix}$

3. (a) The first layer of a convolutional neural network (CNN) consists of a convolution layer of neurons having weights $w = \begin{pmatrix} 0.2 & 1.0 \\ 0.1 & -0.2 \end{pmatrix}$ and rectified linear (ReLU) activation functions, and a max pooling layer having a pooling window of 2x2 size. The bias connected to the convolution layer is 0.3.

An input image I is applied to the input layer of the CNN:

$$\boldsymbol{I} = \begin{pmatrix} 0.4 & 0.1 & -0.2 & 0.3 & -0.5 \\ -0.8 & -0.2 & 0.5 & 0.4 & 0.1 \\ 0.0 & 0.2 & 1.0 & -0.3 & 0.2 \end{pmatrix}$$

Find the feature maps at the first convolution layer and pooling layer.

(10 marks)

Answer

$$\boldsymbol{I} = \begin{pmatrix} 0.4 & 0.1 & -0.2 & 0.3 & -0.5 \\ -0.8 & -0.2 & 0.5 & 0.4 & 0.1 \\ 0.0 & 0.2 & 1.0 & -0.3 & 0.2 \end{pmatrix} \qquad w = \begin{pmatrix} 0.2 & 1.0 \\ 0.1 & -0.2 \end{pmatrix}$$

- P = 0
- S = 1
- bias = 0.3
- filter size = 2 x 2

convolution size:

$$\text{row} = \frac{R - F + 2P}{S} + 1 = \frac{3 - 2 + 2(0)}{1} + 1 = 2$$

$$\text{col} = \frac{C - F + 2P}{S} + 1 = \frac{5 - 2 + 2(0)}{1} + 1 = 4$$

Synaptic inputs to the feature map with filter w: $u = Conv(\boldsymbol{X}, w) + b$

$$u(1,1) = (0.4)(0.2) + (0.1)(0.1) + (-0.8)(0.1) + (-0.2)(-0.2) + 0.3$$
$$= 0.08 + 0.01 - 0.08 + 0.04 + 0.3 = 0.35$$

$$u(1,2) = (0.1)(0.2) + (-0.2)(0.1) + (-0.2)(0.1) + (0.5)(-0.2) + 0.3$$
$$= 0.02 - 0.02 - 0.02 - 0.1 + 0.3 = 0.18$$

$$u(1,3) = (-0.2)(0.2) + (0.3)(0.1) + (0.5)(0.1) + (0.4)(-0.2) + 0.3$$
$$= -0.04 + 0.03 + 0.05 - 0.08 + 0.3 = 0.26$$

$$u(1,4) = (0.3)(0.2) + (-0.5)(0.1) + (0.4)(0.1) + (0.1)(-0.2) + 0.3$$
$$= 0.06 - 0.05 + 0.04 - 0.02 + 0.3 = 0.33$$

$$u(2,1) = (-0.8)(0.2) + (-0.2)(0.1) + (0.0)(0.1) + (0.2)(-0.2) + 0.3$$
$$= -0.16 - 0.02 + 0 - 0.04 + 0.3 = 0.08$$

$$u(2,2) = (-0.2)(0.2) + (0.5)(0.1) + (0.2)(0.1) + (1.0)(-0.2) + 0.3$$
$$= -0.04 + 0.05 + 0.02 - 0.2 + 0.3 = 0.13$$

$$u(2,3) = (0.5)(0.2) + (0.4)(0.1) + (1.0)(0.1) + (-0.3)(-0.2) + 0.3$$
$$= 0.1 + 0.04 + 0.1 + 0.06 + 0.3 = 0.6$$

$$u(2,4) = (0.4)(0.2) + (0.1)(0.1) + (-0.3)(0.1) + (0.2)(-0.2) + 0.3$$
$$= 0.08 + 0.01 - 0.03 - 0.04 + 0.3 = 0.32$$

3. (a) cont

Answer

$$u = \begin{pmatrix} 0.35 & 0.18 & 0.26 & 0.33 \\ 0.08 & 0.13 & 0.60 & 0.32 \end{pmatrix}$$

Feature maps at the convolution layer:

$$y = relu(u) = \max\{0, u\} = \begin{pmatrix} 0.35 & 0.18 & 0.26 & 0.33 \\ 0.08 & 0.13 & 0.60 & 0.32 \end{pmatrix}$$

pooling size:

$$\text{row} = \frac{R-F}{S} + 1 = \frac{2-2}{1} + 1 = 1$$

$$\text{col} = \frac{C-F}{S} + 1 = \frac{4-2}{1} + 1 = 3$$

Feature maps at the pooling layer:

$$P_{max} = \begin{pmatrix} 0.35 & 0.60 & 0.60 \end{pmatrix}$$

3. (b) State how an autoencoder is trained in order to

3. (b) (i) remove noise in data.

(4 marks)

Answer

- Create a noisy version of the original data by adding random noise.
- The noisy data is fed into the autoencoder as input, while the original, clean data is used as the target output.
- The autoencoder consists of an encoder that compresses the input data into a latent representation and a decoder that reconstructs the data from this representation.
- The loss function measures the difference between the reconstructed output and the original clean data.
- The autoencoder is trained to minimize this loss, effectively learning to remove the noise during reconstruction.

3. (b) (ii) extract sparse features of data.

(4 marks)

Answer

- Apply a regularization term to the loss function that penalizes non-sparse representations.

- Similar to a standard autoencoder, but with modifications to enforce sparsity in the latent space.

- The autoencoder is trained on the original data, with the regularization term ensuring that the latent representation remains sparse.

- The loss function includes both the reconstruction error and the sparsity penalty.

- The autoencoder is trained to minimize this combined loss, resulting in sparse feature extraction.

3. (c) An autoencoder with a single hidden-layer attempts to reconstruct 3-dimensional input patterns $x \in \boldsymbol{R}^3$. The hidden-layer has two neurons with activation functions $g(u) = \frac{1}{1+e^{-u}}$ and the output layer has three neurons with activation functions $f(u) = \frac{1-e^{u}}{1+e^{u}}$. The weight matrix $\boldsymbol{W}$ connected to the hidden-layer, the bias vector $\boldsymbol{b}$ of the hidden-layer, and bias vector $\boldsymbol{b}'$ of the output layer are given by

$$\boldsymbol{W} = \begin{pmatrix} 1.8 & -2.3 \\ -1.2 & 2.9 \\ -3.2 & -2.2 \end{pmatrix}, \ \boldsymbol{b} = \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}, \text{ and } \boldsymbol{b}' = \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix}$$

For input patterns $\boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix}$ and $\boldsymbol{x}_2 = \begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix}$, find

3. (c) (i) the hidden layer activations.

(3 marks)

Answer

$$\boldsymbol{W} = \begin{pmatrix} 1.8 & -2.3 \\ -1.2 & 2.9 \\ -3.2 & -2.2 \end{pmatrix},\ \boldsymbol{b} = \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}, \text{ and } \boldsymbol{b}' = \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix}\ \boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix}\qquad \boldsymbol{x}_2 = \begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix}$$

$$h = g(u) = \frac{1}{1 + e^{-u}}$$

For input $\boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix}$,

$$u_1 = W^T x_1 + b = \begin{pmatrix} 1.8 & -1.2 & -3.2 \\ -2.3 & 2.9 & -2.2 \end{pmatrix}\begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}$$

$$= \begin{pmatrix} (\ \ 1.8)(0.5) + (-1.2)(0.2) + (-3.2)(-0.8) \\ (-2.3)(0.5) + (\ \ 2.9)(0.2) + (-2.2)(-0.8) \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}$$

$$= \begin{pmatrix} 3.22 \\ 1.19 \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix} = \begin{pmatrix} 2.52 \\ 1.19 \end{pmatrix}$$

$$h_1 = g(u_1) = \frac{1}{1 + e^{-u_1}} = \begin{pmatrix} 0.297 \\ 0.767 \end{pmatrix}$$

For input $\boldsymbol{x}_2 = \begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix}$,

$$u_2 = W^T x_2 + b = \begin{pmatrix} 1.8 & -1.2 & -3.2 \\ -2.3 & 2.9 & -2.2 \end{pmatrix}\begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}$$

$$= \begin{pmatrix} (\ \ 1.8)(0.6) + (-1.2)(-0.4) + (-3.2)(0.7) \\ (-2.3)(0.6) + (\ \ 2.9)(-0.4) + (-2.2)(0.7) \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}$$

$$= \begin{pmatrix} -0.68 \\ -4.08 \end{pmatrix} + \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix} = \begin{pmatrix} -1.38 \\ -4.08 \end{pmatrix}$$

$$h_2 = g(u_2) = \frac{1}{1 + e^{-u_2}} = \begin{pmatrix} 0.201 \\ 0.017 \end{pmatrix}$$

<u>Summary</u>

- hidden activation for $\boldsymbol{x}_1$: $h_1 = \begin{pmatrix} 0.297 \\ 0.767 \end{pmatrix}$
- hidden activation for $\boldsymbol{x}_2$: $h_2 = \begin{pmatrix} 0.201 \\ 0.017 \end{pmatrix}$

3. (c) (ii) the reconstruction errors.

(4 marks)

Answer

$$\boldsymbol{W} = \begin{pmatrix} 1.8 & -2.3 \\ -1.2 & 2.9 \\ -3.2 & -2.2 \end{pmatrix}, \ \boldsymbol{b} = \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}, \text{ and } \boldsymbol{b}' = \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix} \ \boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix} \qquad \boldsymbol{x}_2 = \begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix}$$

$$h_1 = \begin{pmatrix} 0.297 \\ 0.767 \end{pmatrix} \qquad h_2 = \begin{pmatrix} 0.201 \\ 0.017 \end{pmatrix}$$

$$y = f(u) = \frac{1 - e^u}{1 + e^u}$$

For $\boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix}$ P = 3

$$u_1 = Wh_1 + b' = \begin{pmatrix} 1.8 & -2.3 \\ -1.2 & 2.9 \\ -3.2 & -2.2 \end{pmatrix} \begin{pmatrix} 0.297 \\ 0.767 \end{pmatrix} + \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix}$$

$$= \begin{pmatrix} (\ \ 1.8)(0.297) + (-2.3)(0.767) \\ (-1.2)(0.297) + (\ \ 2.9)(0.767) \\ (-3.2)(0.297) + (-2.2)(0.767) \end{pmatrix} + \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix}$$

$$= \begin{pmatrix} -1.23 \\ 1.868 \\ -2.638 \end{pmatrix} + \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix} = \begin{pmatrix} -0.230 \\ 1.068 \\ -0.380 \end{pmatrix}$$

$$y_1 = f(u) = \frac{1 - e^u}{1 + e^u} = \begin{pmatrix} 0.114 \\ -0.488 \\ 0.188 \end{pmatrix}$$

Reconstruction error (MSE):

$$J_{mse1} = \frac{1}{P} \sum_{p=1}^{P} \left\| \boldsymbol{y}_p - \boldsymbol{x}_p \right\|^2$$
$$= \frac{1}{3} [\|0.114 - 0.5\|^2 + \|-0.488 - 0.2\|^2 + \|0.188 - (-0.8)\|^2]$$
$$= \frac{1}{3} [(0.386)^2 + (0.688)^2 + (0.988)^2] = \frac{1}{3}(1.598) = 0.533$$

3. (d) (ii) cont

Answer

$$\boldsymbol{W} = \begin{pmatrix} 1.8 & -2.3 \\ -1.2 & 2.9 \\ -3.2 & -2.2 \end{pmatrix}, \ \boldsymbol{b} = \begin{pmatrix} -0.7 \\ 0.0 \end{pmatrix}, \text{ and } \boldsymbol{b}' = \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix} \ \boldsymbol{x}_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ -0.8 \end{pmatrix} \qquad \boldsymbol{x}_2 = \begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix}$$

$$h_1 = \begin{pmatrix} 0.297 \\ 0.767 \end{pmatrix} \qquad h_2 = \begin{pmatrix} 0.201 \\ 0.017 \end{pmatrix}$$

$$y = f(u) = \frac{1 - e^u}{1 + e^u}$$

For $\boldsymbol{x}_2 = \begin{pmatrix} 0.6 \\ -0.4 \\ 0.7 \end{pmatrix}$ P = 3

$$u_2 = Wh_2 + b' = \begin{pmatrix} 1.8 & -2.3 \\ -1.2 & 2.9 \\ -3.2 & -2.2 \end{pmatrix} \begin{pmatrix} 0.201 \\ 0.017 \end{pmatrix} + \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix}$$

$$= \begin{pmatrix} (\ 1.8)(0.201) + (-2.3)(0.017) \\ (-1.2)(0.201) + (\ 2.9)(0.017) \\ (-3.2)(0.201) + (-2.2)(0.017) \end{pmatrix} + \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix}$$

$$= \begin{pmatrix} 0.323 \\ -0.192 \\ -0.681 \end{pmatrix} + \begin{pmatrix} 1.0 \\ -0.8 \\ 2.6 \end{pmatrix} = \begin{pmatrix} 1.323 \\ -0.992 \\ 1.919 \end{pmatrix}$$

$$y_1 = f(u) = \frac{1 - e^u}{1 + e^u} = \begin{pmatrix} -0.579 \\ 0.459 \\ -0.744 \end{pmatrix}$$

Reconstruction error (MSE):

$$J_{mse2} = \frac{1}{P} \sum_{p=1}^{P} \left\| \boldsymbol{y}_p - \boldsymbol{x}_p \right\|^2$$
$$= \frac{1}{3} [\|-0.579 - 1.323\|^2 + \|0.459 - (-0.992)\|^2 + \|-0.744 - 1.919\|^2]$$
$$= \frac{1}{3} [(1.902)^2 + (1.451)^2 + (2.663)^2] = \frac{1}{3}(12.815) = 4.272$$

Summary

Reconstruction error (MSE):

$J_{mse1} = 0.533$

$J_{mse2} = 4.272$

4. (a) A recurrent neural network (RNN) receives 2-dimensional input patterns $x \in \boldsymbol{R}^2$ and has one hidden layer with recurrent connections. The RNN has three neurons in the hidden layer and one neuron in the output layer. All neurons have logistic activation functions.

The weight matrices U connecting the input to the hidden layer, $\boldsymbol{W}$ connecting the previous hidden state to the next hidden state, and $\boldsymbol{V}$ connecting the hidden output to the output layer are given by

$$\boldsymbol{U} = \begin{pmatrix} -1.0 & 0.5 & 0.2 \\ 0.5 & 0.1 & -2.0 \end{pmatrix}, \boldsymbol{W} = \begin{pmatrix} 2.0 & 1.3 & -1.0 \\ 1.5 & 0.0 & -0.5 \\ -0.2 & 1.5 & 0.4 \end{pmatrix} \text{ and } \boldsymbol{V} = \begin{pmatrix} 2.0 \\ -1.5 \\ 0.2 \end{pmatrix}$$

All bias connections to neurons are set to 0.1 and the hidden-layer activations are initialized to zeros.

Find the output of the network for a sequence $(\boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3))$ of input patterns:

$$\boldsymbol{x}(1) = \begin{pmatrix} 1.0 \\ 2.0 \end{pmatrix}, x(2) = \begin{pmatrix} -1.0 \\ 1.0 \end{pmatrix} \text{ and } \boldsymbol{x}(3) = \begin{pmatrix} 0.0 \\ 3.0 \end{pmatrix}$$

(15 marks)

Answer

$$\boldsymbol{U} = \begin{pmatrix} -1.0 & 0.5 & 0.2 \\ 0.5 & 0.1 & -2.0 \end{pmatrix}, \boldsymbol{W} = \begin{pmatrix} 2.0 & 1.3 & -1.0 \\ 1.5 & 0.0 & -0.5 \\ -0.2 & 1.5 & 0.4 \end{pmatrix} \text{ and } \boldsymbol{V} = \begin{pmatrix} 2.0 \\ -1.5 \\ 0.2 \end{pmatrix}$$

$$\boldsymbol{x}(1) = \begin{pmatrix} 1.0 \\ 2.0 \end{pmatrix}, x(2) = \begin{pmatrix} -1.0 \\ 1.0 \end{pmatrix} \text{ and } \boldsymbol{x}(3) = \begin{pmatrix} 0.0 \\ 3.0 \end{pmatrix}$$

$$\boldsymbol{b} = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \text{ and } c = 0.1 \qquad h = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\phi(u) = \text{sigmoid}(u) = \frac{1}{1 + e^{-u}}$$

$$\sigma(u) = \text{sigmoid}(u) = \frac{1}{1 + e^{-u}}$$

4. (a) cont

Answer

$$\phi(u) = \text{sigmoid}(u) = \frac{1}{1+e^{-u}}$$

$$\sigma(u) = \text{sigmoid}(u) = \frac{1}{1+e^{-u}}$$

At $t = 1$, $\boldsymbol{x}(1) = \begin{pmatrix}1.0\\2.0\end{pmatrix}$,

$$\boldsymbol{h}(t) = \quad \phi(\boldsymbol{U}^T\boldsymbol{x}(t) + \boldsymbol{W}^T\boldsymbol{h}(t-1) + \boldsymbol{b})$$

$$\boldsymbol{h}(1) = sigmoid(\boldsymbol{U}^T\boldsymbol{x}(1) + \boldsymbol{W}^T\boldsymbol{h}(0) + \boldsymbol{b})$$

$$= \text{sigmoid}\left(\begin{pmatrix}-1.0 & 0.5\\ 0.5 & 0.1\\ 0.2 & -2.0\end{pmatrix}\begin{pmatrix}1.0\\2.0\end{pmatrix} + \begin{pmatrix}2.0 & 1.5 & -0.2\\ 1.3 & 0.0 & 1.5\\ -1.0 & -0.5 & 0.4\end{pmatrix}\begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}\right)$$

$$= \text{sigmoid}\left(\begin{pmatrix}(-1.0)(1) + (\ 0.5)(2)\\ (\ 0.5)(1) + (\ 0.1)(2)\\ (\ 0.2)(1) + (-2.0)(2)\end{pmatrix} + \begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}\right)$$

$$= \text{sigmoid}\left(\begin{pmatrix}0.0\\0.7\\-3.8\end{pmatrix} + \begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}\right) = \text{sigmoid}\begin{pmatrix}0.1\\0.8\\-3.7\end{pmatrix} = \begin{pmatrix}0.525\\0.690\\0.024\end{pmatrix}$$

$$\boldsymbol{y}(t) = \quad \sigma(\boldsymbol{V}^T\boldsymbol{h}(t) + c)$$

$$\boldsymbol{y}(1) = sigmoid(\boldsymbol{V}^T\boldsymbol{h}(1) + c) = \text{sigmoid}\left((2.0 \quad -1.5 \quad 0.2)\begin{pmatrix}0.525\\0.690\\0.024\end{pmatrix} + 0.1\right)$$

$$= \text{sigmoid}((2)(0.525) + (-1.5)(0.69) + (0.2)(0.024) + 0.1)$$

$$= \text{sigmoid}(0.120) = 0.530$$

Summary
- Output of the network:
  - y(1) = 0.530
  - y(2) = 0.337

At $t = 2$, $x(2) = \begin{pmatrix}-1.0\\1.0\end{pmatrix}$,

$$\boldsymbol{h}(t) = \quad \phi(\boldsymbol{U}^T\boldsymbol{x}(t) + \boldsymbol{W}^T\boldsymbol{h}(t-1) + \boldsymbol{b})$$

$$\boldsymbol{h}(2) = \text{sigmoid}(\boldsymbol{U}^T\boldsymbol{x}(2) + \boldsymbol{W}^T\boldsymbol{h}(1) + \boldsymbol{b})$$

$$= \text{sigmoid}\left(\begin{pmatrix}-1.0 & 0.5\\ 0.5 & 0.1\\ 0.2 & -2.0\end{pmatrix}\begin{pmatrix}-1.0\\1.0\end{pmatrix} + \begin{pmatrix}2.0 & 1.5 & -0.2\\ 1.3 & 0.0 & 1.5\\ -1.0 & -0.5 & 0.4\end{pmatrix}\begin{pmatrix}0.525\\0.690\\0.024\end{pmatrix} + \begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}\right)$$

$$= \text{sigmoid}\left(\begin{pmatrix}(-1.0)(-1) + (\ 0.5)(1)\\ (\ 0.5)(-1) + (\ 0.1)(1)\\ (\ 0.2)(-1) + (-2.0)(1)\end{pmatrix} + \begin{pmatrix}(\ 2.0)(0.525) + (1.5)(0.69) + (-0.2)(0.024)\\ (\ 1.3)(0.525) + (0.0)(0.69) + (\ 2.5)(0.024)\\ (-1.0)(0.525) + (-0.5)(0.69) + (0.4)(0.024)\end{pmatrix} + \begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}\right)$$

$$= \text{sigmoid}\left(\begin{pmatrix}1.5\\0.5\\-2.2\end{pmatrix} + \begin{pmatrix}2.080\\0.041\\-0.860\end{pmatrix} + \begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}\right) = \text{sigmoid}\begin{pmatrix}3.680\\0.641\\-2.960\end{pmatrix} = \begin{pmatrix}0.975\\0.655\\0.049\end{pmatrix}$$

$$\boldsymbol{y}(t) = \quad \sigma(\boldsymbol{V}^T\boldsymbol{h}(t) + c)$$

$$\boldsymbol{y}(2) = sigmoid(\boldsymbol{V}^T\boldsymbol{h}(2) + c) = \text{sigmoid}\left((2.0 \quad -1.5 \quad 0.2)\begin{pmatrix}0.975\\0.655\\0.049\end{pmatrix} + 0.1\right)$$

$$= \text{sigmoid}((0.2)(0.975) + (-1.5)(0.655) + +(0.2)(0.049) + 0.12)$$

$$= \text{sigmoid}(-0.678) = 0.337$$

4. (b) Briefly explain

(i) why recurrent neural networks are difficult to train to learn long-term dependencies.

(5 marks)

Answer

- RNNs are difficult to train on long-term dependencies because of the vanishing and exploding gradient problem.

- During backpropagation through time (BPTT), gradients are repeatedly multiplied by weights across many time steps.

- If the weights are small, gradients shrink exponentially (vanishing gradients), making it hard for the network to adjust weights based on distant past inputs.

- If weights are large, gradients grow exponentially (exploding gradients), causing instability.

- As a result, standard RNNs struggle to capture relationships between inputs and outputs that are far apart in the sequence.

(ii) how long short-term memory (LSTM) networks are able to learn long-term interactions.

(5 marks)

Answer

- LSTM networks overcome long-term dependency issues using a memory cell and a system of gates (input, forget, and output gates).

- The memory cell allows information to flow unchanged across many time steps unless explicitly modified by the gates.

- The forget gate decides what information to discard, the input gate controls what new information to store, and the output gate determines what part of the memory is output.

- This architecture maintains stable gradients during training and enables LSTMs to preserve and access information over long sequences, effectively learning long-term interactions.