Homework-4

Name: Chitradevi Maruthavanan
PSU_ID: 950828319

1. Consider the neural network (1) with a scalar input x and parameters.
   $W^H = \begin{bmatrix}1\\1\end{bmatrix}$, $b^H = \begin{bmatrix}-1\\-3\end{bmatrix}$, $W^O = [-1,2]$, $b^O = 0.5$
   using a hard threshold activation function (2) and threshold output function (4).

   (a)What is N_h, the number of hidden units? What is N_o, the number of output units?

   $W^H = N_h * N_i$, Since the parameter matrix W^H is a 2*1 matrix, so the number of hidden units N_h is 2. $W^o = No * N_h$, Since the parameter matrix W^0 is a 1*2 matrix, so the number of output units N_0 is 1.

   (b) Write z^H in terms of x. Show the functions for each component Z_j^H

   $z^H = w^H x + b^H$
   $z^H = \begin{bmatrix}1\\1\end{bmatrix}x + \begin{bmatrix}-1\\-3\end{bmatrix}$
   $z^H_1 = x-1$
   $z^H_2 = x-3$

   (c) Write u^H in terms of x. Show the functions for each component $\mathbf{u^H_j}$.

   $u^H = gact(z^H)$
   $u^H_1 = gact(z^H_1) = gact(x-1)$
   $u^H_2 = gact(Z^H_2) = gact(x-3)$
   $u^H_1 = 0$, if $x -1 < 0$; $u^H_1 = 1$, if $x -1 >=0$;
   $u^H_2 = 0$, if $x -3 < 0$; $u^H_2 = 1$, if $x -3 >=0$;
   To simplify,
   $u^H_1 = 0$, if $x < 1$; $u^H_1 = 1$, if $x >=1$;
   $u^H_2 = 0$, if $x < 3$; $u^H_2 = 1$, if $x >=3$;

(d) Write z^O in terms of x.

$z^O = w^o u^H + b^o$

$z^O = [-1\ \ 2]\ [u^H_1,\ \ u^H_2] + 0.5$

$z^O = [-1\ \ 2]\ [x < 1, 1 <= x < 3, x >= 3] + 0.5$

substituting u^H_1 and u^H_2 in terms of x, we get:

$z^O = [-1\ 2]\ [0\ 0] + 0.5$ if $x < 1$

$z^O = [-1\ 2]\ [1\ 0] + 0.5$ if $1 <= x < 3$

$z^O = [-1\ 2]\ [1\ 1] + 0.5$ if $x >= 3$

To Simplify we get,

$z^O = 0.5$ if $x < 1$

$z^O = -0.5$ if $1 <= x < 3$

$z^O = 1.5$ if $x >= 3$

(e) What is $y^\wedge$ in terms of x?

$\hat{y} =$ gout(z^o)

gout(z^o) = 1 if $z^o >= 0$

gout(z^o) = 0 if $z^o < 0$

Substituting z^O, we get:

$\hat{y} = 1$, if $[x < 1, x >= 3]$

$\hat{y} = 0$, if $[1 <= x < 3]$

2. Consider the data set for four points with scalar features xi and binary class labels $y_i = 0; 1$.

x_i	0	1	3	5
y_i	0	0	1	0

(a) Find a neural network with $N_h = 2$ units, $N_o = 1$ output units such that $\hat{y}_i = y_i$ on all four data points. Use a network similar in structure to the previous problem. Also, you want to find features that can distinguish between x = 3 and x = {0, 1;,5}. Since there are many features, use two features: whether x >= 2 and x >= 4. Use a hard threshold activation function (2) and threshold output function (4). State the weights and biases used in each layer.

To determine the weights and biases for this network,

For the hidden layer, we can set:

- **$W^H{}_1 = W^H{}_2 = 1$**
- **$b^H{}_1 = -2$ (so that gact(x - 2) = 1 when x >= 2, and gact(x - 2) = 0 otherwise)**

- **b^H_2 = -4 (so that gact(x - 4) = 1 when x >= 4, and gact(x - 4) = 0 otherwise)**

For the output layer, we can set:

- **w^O = [1, -1]**
- **b^O = -0.5**

Putting it all together, the neural network has the following weights and biases:

$$W^H = \begin{bmatrix}1\\1\end{bmatrix} \qquad b^H = \begin{bmatrix}-2\\-4\end{bmatrix} \qquad W^O = [1, -1] \qquad b^O = -0.5$$

The above weights and biases arrived on a trial and error basis by trying different x_i and making sure the correct ^yi was achieved. The detailed computation is described below:

(b) Compute the values of ^yi and all the intermediate variables z^H_i, u^H_i and z^O_i for each sample $x = x_i$.

We can compute the intermediate variables for each sample:

- **X = 0**

$z^H = w^H x + b^H$

$z^H = \begin{bmatrix}1\\1\end{bmatrix}x + \begin{bmatrix}-2\\-4\end{bmatrix}$

$z^H_1 = \begin{bmatrix}1\\1\end{bmatrix}0 + \begin{bmatrix}-2\\-4\end{bmatrix}$

$z^H_{11} = -2, \quad z^H_{12} = -4$

$u^H = gact(z^H)$

$u^H_{11} = gact(z^H_{11}) = gact(-2)$

$u^H_{12} = gact(Z^H_{12}) = gact(-4)$

$u^H_{11} = 0 \qquad u^H_{12} = 0$

$z^O = w^o u^H + b^o$

$z^O_1 = [1 \ \ -1]\ [u^H_{11},\ u^H_{12}] - 0.5$

$z^O_1 = [1 \ \ -1]\ [0,\ 0] - 0.5$

$z^O_1 = 0+0-0.5$

$z^O_1 = -0.5$

$y\hat{} = gout(z^o)$

$gout(z^o) = 1$ if $z^o >= 0$

$gout(z^o) = 0$ if $z^o < 0$

$y\hat{} = gout(-0.5) = 0$

- $X = 1$

  $z^{H}_{2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 1 + \begin{bmatrix} -2 \\ -4 \end{bmatrix}$

  $z^{H}_{21} = -1, \quad z^{H}_{22} = -3$

  $u^{H}_{21} = gact(-1), \quad u^{H}_{22} = gact(-3)$

  $u^{H}_{21} = 0, \quad u^{H}_{22} = 0$

  $z^{O}_{2} = [1 \ \ -1] \ [0, \ 0] - 0.5$

  $z^{O}_{2} = 0+0-0.5$

  $z^{O}_{2} = -0.5$

  $y\hat{} = gout(-0.5)$

  $y\hat{} = 0$

- $X = 3$

  $z^{H}_{3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 3 + \begin{bmatrix} -2 \\ -4 \end{bmatrix}$

  $z^{H}_{31} = 1, \quad z^{H}_{32} = -1$

  $u^{H}_{31} = gact(1), \quad u^{H}_{32} = gact(-1)$

  $u^{H}_{31} = 1, \quad u^{H}_{32} = 0$

  $z^{O}_{3} = [1 \ \ -1] \ [1, \ 0] - 0.5$

  $z^{O}_{3} = 1+0-0.5$

  $z^{O}_{3} = 0.5$

  $y\hat{} = gout(0.5)$

  $y\hat{} = 1$

- $X = 5$

  $z^{H}_{4} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 5 + \begin{bmatrix} -2 \\ -4 \end{bmatrix}$

  $z^{H}_{41} = 3, \quad z^{H}_{42} = 1$

  $u^{H}_{41} = gact(3), \quad u^{H}_{42} = gact(1)$

  $u^{H}_{41} = 1, \quad u^{H}_{42} = 1$

  $z^{O}_{4} = [1 \ \ -1] \ [1, \ 1] - 0.5$

  $z^{O}_{4} = 1-1-0.5$

$z^O{}_4$ = -0.5

$y\hat{} =$ gout(-0.5)
$y\hat{} =$ 0

(c) Now suppose we are given a new sample, x = 3:5. What does the network predict as ^y?

- X = 3.5

$z^H = \begin{bmatrix}1\\1\end{bmatrix}3.5 + \begin{bmatrix}-2\\-4\end{bmatrix}$

$z^H{}_1 = 1.5, \quad z^H{}_2 = -0.5$

$u^H{}_1$ = gact(1.5), $u^H{}_2$ = gact(-0.5)
$u^H{}_1$ = 1, $u^H{}_2$= 0

z^O = [1 -1] [1, 0] - 0.5
z^O = 1-0-0.5
z^O = 0.5

$y\hat{} =$ gout(z^o)
gout(0.5)
$y\hat{} =$ 1

3(a) Write the components of z^H and u^H as a function of (x_1, x_2). For each component j, indicate where in the (x_1; x_2) plane $u^H{}_j$= 1.

$z^H = w^H x + b^H$

$z^H = \begin{bmatrix}1 & 0\\0 & 1\\1 & 1\end{bmatrix}[x_1 \quad x_2] + \begin{bmatrix}0\\0\\-1\end{bmatrix}$

$z^H{}_1 = x_1$
$z^H{}_2 = x_2$
$z^H{}_3 = x_1+x_2-1$
$u^H = 1$ if $z^H >= 0$
$u^H = 0$ if $z^H < 0$

$u^H{}_1 = 1$ if $x_1 >= 0$; $u^H{}_1 = 0$ if $x_1 < 0$
$u^H{}_2 = 1$ if $x_2 >= 0$; $u^H{}_2 = 0$ if $x_2 < 0$
$u^H{}_3 = 1$ if $x_1+x_2 >= 1$; $u^H{}_3 = 0$ if $x_1+x_2 < 1$

(b) Write z^O as a function of $(x_1; x_2)$. In what region is $\hat{y} = 1$?

$z^O = w^o u^H + b^o$
$z^O = [1\ 1\ -1]\ [u^H_1\ \ u^H_2\ \ u^H_3] - 1.5$
$\hat{y} =$ gout(z^o)
gout(z^o) = 1 if $z^o >= 0$
$z^O = [1\ 1\ -1]\ [\ u^H_1\ ,\ u^H_2\ , u^H_3] - 1.5$
$z^O = [1\ 1\ -1]\ [1\ \ 1\ \ 0] - 1.5$ if $x_1 >= 0$, $x_2 >= 0$ and $x_1+x_2 < 1$
$z^O = [1+1-0]-1.5. = 0.5$
$\hat{y} =$ gout(0.5) = 1
$\hat{y} =$ 1 only when $x_1 >= 0$, $x_2 >= 0$, $x_1+x_2 < 1$

The other cases lead to $\hat{y} = 0$. They are:
$z^O = [0+0-0]-1.5. = -1.5$ if $x_1 < 0$, $x_2 < 0$ and $x_1+x_2 < 1$
$z^O = [1+0-0]-1.5. = -0.5$ if $x_1 >= 0$, $x_2 < 0$ and $x_1+x_2 < 1$
$z^O = [0+1-0]-1.5. = -0.5$ if $x_1 < 0$, $x_2 >= 0$ and $x_1+x_2 < 1$
$z^O = [0+0-1]-1.5. = -2.5$ if $x_1 < 0$, $x_2 < 0$ and $x_1+x_2 >= 1$
$z^O = [0+1-1]-1.5. = -1.5$ if $x_1 < 0$, $x_2 >= 0$ and $x_1+x_2 >= 1$
$z^O = [1+0-1]-1.5. = -1.5$ if $x_1 >= 0$, $x_2 < 0$ and $x_1+x_2 >= 1$
$z^O = [1+1-1]-1.5. = -0.5$ if $x_1 >= 0$, $x_2 >= 0$ and $x_1+x_2 >= 1$

$\hat{y} =$ 1 only when $x_1 >= 0$, $x_2 >= 0$, $x_1+x_2 < 1$

4. Architecture choices for different problems: For each problem, state possible selections for the dimensions N_i, N_h, N_o and the functions gact(.) and gout(.). Indicate which parameters are free to choose.

(a) One wants a neural network to take as an input a 20*20 gray scale image and determine which letter (`a' to `z') the image is of.

Possible choices for architecture are:

- **N_i = 400 (20 x 20 pixels)**
- **N_h = variable, free to choose**
- **N_o = 26 (one output per letter)**
- **gact(·) = ReLU or sigmoid (free to choose)**
- **gout(·) = softmax function for k- class classification, as we want a probability distribution over the 26 classes**

(b) One extracts 120 features of a sample of a speech recording (like the MFCCs). Based on the audio samples, the network is to determine if the speech is male or female.

Possible choices for architecture are:

- **Ni = 120 (number of features)**
- **Nh = variable, free to choose**
- **No = 1 (binary classification)**
- **gact(·) = ReLU or sigmoid (free to choose)**
- **gout(·) = sigmoid for binary classification, as we want to output a probability between 0 and 1 indicating the likelihood of the sample being from a male or female speaker**

(c) One wants a neural network to predict the stock price based on the average stock price of the last five days.

Possible choices for architecture are:

- **Ni = 5 (last 5 days)**
- **Nh = variable, free to choose**
- **No = 1 (stock price prediction)**
- **gact(·)= ReLU or sigmoid (free to choose)**
- **gout(·) = linear, as we want to output a real-valued prediction for the stock price**

5. Implementation in python: Write python code for implementing the following steps for a batch of samples:

(a) The hidden layer step (10a).

```
import numpy as np
def hidden_layer(X, WH, BH, gact):
    ZH = np.dot(X, WH.T) + BH
    UH = gact(ZH)
    return UH, ZH
```

(b) The output layer step (10b) for binary classification with a sigmoid output (3).

```
import numpy as np
def sigmoid(x):
    return 1 / (1 + np.exp(-x))
def binary_output_layer(UH, WO, BO):
    ZO = np.dot(UH, WO.T) + BO
    UO = sigmoid(ZO)
    return UO
```

(c) The output layer step (10b) for K-class classification with a softmax output (5).

```
import numpy as np
def softmax(x):
  exp_x = np.exp(x)
  return exp_x / np.sum(exp_x, axis=1, keepdims=True)
def output_layer_softmax(UH, WO, BO):
  ZO = np.dot(UH, WO.T) + BO
  UO = softmax(ZO)
  return UO
```

For all examples, avoid for-loops and instead use Python broadcasting.