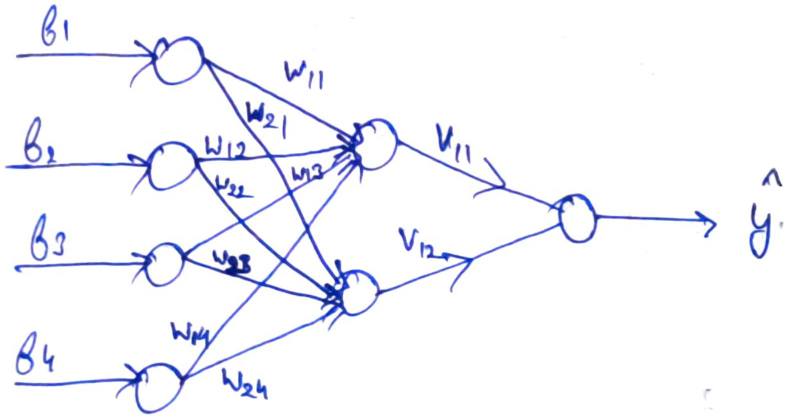
Name : NONIT GUPTA
Roll : 230712
Course : EE656 – Assignment 5

Training data →

b_1	b_2	b_3	b_4	y
0.5	0.1	0.5	11	1.1
0.1	0.3	0.7	13	2.3
0.3	0.7	0.9	10	1.5
0.0	0.8	0.1	11	1.4
0.5	0.1	0.3	16	3.3

One hidden layer (2 neurons)
Activation function → Sigmoid
Loss function → Mean Squared Error (MSE)
Learning rate (η) → 0.1



$$\begin{array}{cccc} & w_{1x} & w_{2x} & \\ w_{11} & \multicolumn{2}{c}{\begin{bmatrix} 0.1 & 0.2 \\ 0.2 & 0.1 \\ 0.5 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}} & w_{21} \end{array}$$

$$\begin{array}{c|cc|c} w_{11} & 0.1 & 0.2 & w_{21} \\ w_{12} & 0.2 & 0.1 & w_{22} \\ w_{13} & 0.5 & 0.1 & w_{23} \\ w_{14} & 0.1 & 0.1 & w_{24} \end{array}$$

$$\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.3 \end{bmatrix}$$

For first row,

Input x: [0.5, 0.1, 0.5, 11]

True Output: y = 1.1

Forward propagation →

Hidden layer → Neuron 1 →

$$z_1 = (0.5)(0.1) + (0.1)(0.2) + (0.5)(0.5) + (11)(0.1)$$
$$= 0.05 + 0.02 + 0.25 + 1.1 = 1.42$$
$$a_1 = \sigma(1.42) = \frac{1}{1+e^{-1.42}} = 0.805$$

Neuron 2 →

$$z_2 = (0.5)(0.2) + (0.1)(0.1) + (0.5)(0.1) + (11)(0.1)$$
$$= 0.1 + 0.01 + 0.05 + 1.1 = 1.26$$
$$a_2 = \sigma(1.26) = \frac{1}{1+e^{-1.26}} = 0.779$$

Output layer →

$$z_{out} = (0.5)(a_1) + (0.3)(a_2) = 0.6362$$

Loss (MSE) →

$$Loss = \frac{1}{2}(y-\hat{y})^2 = \frac{1}{2}(1.1 - 0.6362)^2 \approx 0.1075$$

~~Backward~~ Backpropagation →

Derivative of loss wrt output →

$$\frac{\partial L}{\partial \hat{y}} = \hat{y} - y = 0.6362 - 1.1 = -0.4638$$

Derivative wrt output weights →

For weight $w_1^{out} = 0.5$ linked to $a_1 = 0.805$

$$\frac{\partial L}{\partial w_1^{out}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1^{out}} = (-0.4638)(0.805)$$

$$= -0.3736$$

Similarly for w_2^{out}:

$$\frac{\partial L}{\partial w_2^{out}} = (-0.4638)(0.779) = -0.3613$$

Derivative wrt hidden activations

$$\frac{\partial L}{\partial a_1} = (-0.4638)(0.5) = -0.2319$$

$$\frac{\partial L}{\partial a_2} = (-0.4638)(0.3) = -0.1391$$

Derivative wrt hidden layer weights

$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$

Neuron 1

$$\sigma'(z_1) = (0.805)(1 - 0.805) \approx 0.157$$

$$\frac{\partial L}{\partial z_1} = (-0.2319)(0.157) = -0.0364$$

Now compute gradient wrt each weight in neuron 1

$$\frac{\partial L}{\partial w_{11}} = (-0.0364) \cdot x_1 = (-0.0364)(0.5) = -0.0182$$

$$\frac{\partial L}{\partial w_{12}} = -0.00364 \qquad \frac{\partial L}{\partial w_{13}} = -0.0182$$

$$\frac{\partial L}{\partial w_{14}} = -0.4004$$

Neuron 2 →

$$\sigma'(z_2) = 0.779(1-0.779) \approx 0.172$$

$$\frac{\partial L}{\partial z_2} = (-0.1391)(0.172) = -0.0239$$

$$\frac{\partial L}{\partial w_{21}} = (-0.0239)(0.5) = -0.01195$$

$$\frac{\partial L}{\partial w_{22}} = (-0.0239)(0.1) = -0.00239$$

$$\frac{\partial L}{\partial w_{23}} = (-0.0239)(0.5) = -0.01195$$

$$\frac{\partial L}{\partial w_{24}} = (-0.0239)(0.11) = -0.2633$$

Update Weight (Gradient Descent) →

Output layer:

$$w_1^{out} \leftarrow 0.5 - 0.1 \cdot (-0.3736) = 0.5 + 0.0374$$
$$= 0.5374$$

$$w_2^{out} \leftarrow 0.5 - 0.1 \cdot (-0.3613) = 0.3 + 0.0361$$
$$= 0.3361$$

Hidden layer:

Neuron 1:

$$w_{11} \leftarrow 0.1 + (0.1)(0.0182) = 0.1018$$

$$w_{12} \leftarrow 0.2 + 0.000364 = 0.2004$$

Neuron 2:

$$w_{21} \leftarrow 0.2 + 0.001195 = 0.2012$$

$$w_{22} \leftarrow 0.1 + 0.000239 = 0.1002$$

$$w_{23} \leftarrow 0.1 + 0.001195 = 0.1012$$

$w_{24} \leftarrow 0.1 + 0.02633 = 0.1263$

So, updated weights now are,

Neuron 1: $[0.1018, 0.2004, 0.5018, 0.1400]$

Neuron 2: $[0.2012, 0.1002, 0.1012, 0.1263]$

Output Neuron: $[0.5374, 0.3361]$

Row 2 → Input : $[0.1, 0.3, 0.7, 13]$, Target $= 2.3$

Forward pass →

Hidden Neuron 1: $z_1 = (0.1)(0.1018) + (0.3)(0.2004) + (0.7)(0.5018)$
$+ (0.1400)(13) = 2.2416$

$\Rightarrow a_1 = \sigma(2.24) = 0.904$

Hidden Neuron 2: $z_2 = (0.1)(0.2012) + (0.3)(0.1002) + (0.7)(0.1012)$
$+ (13)(0.1263) = 1.763$

$\Rightarrow a_2 = \sigma(1.763) = 0.854$

Output : $\hat{y} = (0.5374)(0.904) + (0.3361)(0.854) = 0.7724$

Loss : $\frac{1}{2}(2.3 - 0.7724)^2 = \frac{1}{2}(1.5276)^2 \approx 1.167$

Backpropagation: $\delta_{out} = 0.7724 - 2.3 = -1.5276$

Output Weights : $\Delta w_{o_1} = (-1.5276)(0.904) = -1.381$
$\Delta w_{o_2} = (-1.5276)(0.854) = -1.305$

Hidden Neuron 1:

$\sigma' = 0.904(1 - 0.904) = 0.087, \quad \delta_1 = (-1.5276)(0.5374)(0.087)$
$= -0.0712$

$\Delta w_1 = (-0.0712) \cdot [0.1, 0.3, 0.7, 13] = [-0.00712, -0.0214,$
$-0.0498, -0.9256]$

Hidden Neuron 2 →

$\sigma' = 0.854(1-0.854) = 0.125, \quad \delta_2 = (-1.5276)(0.3861)(0.125)$
$\doteq -0.0641$

$\Delta w_2 = -0.0641\,[0.1, 0.3, 0.7, 13] \approx \cancel{[-0.00712, -0.0214, -0.0498, -0.9256]}$
$= [-0.00641, -0.0192, -0.0449, -0.8333]$

Weight update →

Hidden Neuron 1:

$[0.1018 + 0.000712, 0.2004 + 0.00214, 0.5018 + 0.00498, 0.1400 + 0.09256]$
$= [0.1025, 0.2025, 0.5068, 0.2326]$

Hidden Neuron 2:

$[0.2018, 0.1021, 0.1057, 0.2096]$

Output Neuron → $[0.6755, 0.4666]$

Row 3 → Input : $[0.3, 0.7, 0.9, 10]$
Target : 1.5

Forward Pass →

Hidden Neuron 1: $z_1 = (0.3)(0.1025) + (0.7)(0.2025) + (0.9)(0.5068)$
$+ (10)(0.2326) = 2.9546$

$\Rightarrow a_1 = \sigma(2.9546) \approx 0.9504$

Hidden Neuron 2: $z_2 = (0.3)(0.2018) + (0.7)(0.1021) + (0.9)(0.1057)$
$+ (10)(0.2096) = 2.3231$

$\Rightarrow a_2 = \sigma(2.3231) \Rightarrow a_2 = \sigma(2.3231) = 0.9108$

Output Neuron → $\hat{y} = (0.6755)(0.9504) + (0.4666)(0.9108)$
$= 1.067$

Loss: $L = \frac{1}{2}(1.5 - 1.067)^2 = 0.0937$

Backpropagation: $\delta_{out} = \hat{y} - y = 1.067 - 1.5 = -0.433$

output layer gradients: $\Delta w_{a_1} = (-0.433)(0.9504) = -0.4117$
$\Delta w_{a_2} = (-0.433)(0.9108) = -0.3945$

Hidden Neuron 1: $\sigma'(z_1) = (0.9504)(1 - 0.9504) \approx 0.0471$
$\delta_1 = (-0.433)(0.6755)(0.0471) = -0.0137$
$\Delta w_1 = (-0.0137) \cdot [0.3, 0.7, 0.9, 10] = [-0.00411, -0.00959, -0.0123, -0.137]$

Hidden Neuron 2: $\sigma'(z_2) = (0.9108)(1 - 0.9108) = 0.0814$
$\delta_2 = (-0.433)(0.4666)(0.0814) = -0.0164$
$\Delta w_2 = (-0.0164) \cdot [0.3, 0.7, 0.9, 10] = [-0.00492, -0.01148, -0.0148, -0.164]$

Weight Update →

Hidden Neuron 1 → $[0.1029, 0.2035, 0.5080, 0.2463]$

Hidden Neuron 2 → $[0.2023, 0.1032, 0.1072, 0.2260]$

Output Weights → $[0.7167, 0.5060]$

Row 4 → Input : $[0.0, 0.8, 0.1, 11]$
Target : 1.4

Forward pass:

Hidden Neuron 1: $z_1 = (0)(0.1029) + (0.8)(0.2035) + (0.1)(0.5080) + (11)(0.2463) = 2.9226$

$a_1 = \sigma(2.9226) = 0.9486$

Hidden Neuron 2: $z_2 = 0 + (0.8)(0.1032) + (0.1)(0.1072) + (11)(0.2260) = 2.5793$

$a_2 = \sigma(2.5793) \approx 0.9296$

Output → $\hat{y} = (0.7167)(0.9486) + (0.5060)(0.9296) = 1.1502$

Loss: $L = \frac{1}{2}(1.4 - 1.1502)^2 = 0.0312$

Backpropagation →

$\delta_{out} = 1.1502 - 1.4 = -0.2498$

Output Gradients: $\Delta w_{a_1} = (-0.2498)(0.9486) = -0.2369$

$\Delta w_{a_2} = (-0.2498)(0.9296) = -0.2323$

Hidden Neuron 1: $\sigma' = (0.9486)(1 - 0.9486) \approx 0.0486$

$\delta_1 = (-0.2498)(0.7167)(0.0486) = -0.0087$

$\Delta w_1 = (-0.0087) \cdot [0.0, 0.8, 0.1, 11] = [0, -0.00696, -0.00087, -0.0957$

Hidden Neuron 2: $\sigma' = (0.9296)(1 - 0.9296) = 0.0655$

$\delta_2 = (-0.2498)(0.5060)(0.0655) = -0.0083$

$\Delta w_2 = (-0.0083) \cdot [0.0, 0.8, 0.1, 11] = [0, -0.00664, -0.00083, -0.0913$

Weight update:

Hidden Neuron 1: $[0.1029, 0.2042, 0.5081, 0.2559]$

Hidden Neuron 2: $[0.2023, 0.1039, 0.1073, 0.2351]$

Output Weights : $[0.7404, 0.5292]$

Row 5 → Input : $[0.5, 0.1, 0.3, 16]$
Target : 3.3

Forward pass:

Hidden Neuron 1: $z_1 = (0.5)(0.1029) + (0.1)(0.2042) + (0.3)(0.5081) + (16)(0.2559) = 4.3187$

$a_1 = \sigma(4.3187) \approx 0.9868$

Hidden Neuron 2: $z_2 = (0.5)(0.2023) + (0.1)(0.1039) + (0.3)(0.1073) + (16)(0.2351) = 3.9053$

$a_2 = \sigma(3.9053) \approx 0.9802$

Output : $\hat{y} = (0.7404)(0.9868) + (0.5292)(0.9802) = 1.2494$

Loss: $L = \frac{1}{2}(3.3 - 1.2494)^2 = 2.102$

Backpropagation: $\delta_{out} = 1.2494 - 3.3 = -2.0506$

Output Gradients :

$\Delta w_{a_1} = (-2.0506)(0.9868) = (-2.0234)$

$\Delta w_{a_2} = (-2.0506)(0.9802) = -2.0102$

Hidden Neuron 1 : $\sigma' = 0.9868(1 - 0.9868) \approx 0.013$

$\delta_1 = (-2.0506)(0.7404)(0.013) = -0.0197$

$\Delta w_1 = (-0.0197) \cdot [0.5, 0.1, 0.3, 16] = [-0.00985, -0.00197, -0.00591, -0.3152]$

Hidden Neuron 2: $\sigma' = 0.9802(1-0.9802) = 0.0195$

$\delta_2 = (-2.0506)(0.5292)(0.0195)$

$= -0.0211$

$\Delta w_2 = (-0.0211) \cdot [0.5, 0.1, 0.3, 16] = [-0.0106, -0.0021, -0.00633, -0.3376]$

Final weights →

Hidden Neuron 1: $[0.1039, 0.2044, 0.5087, 0.2874]$

- Hidden Neuron 2: $[0.2034, 0.1041, 0.1079, 0.2689]$

Output Weights: $[0.9427, 0.7302]$

Test input 1 → Input: $[0.9, 0.3, 0.2, 10]$

True value: 2.7

Hidden Neuron 1: $w_1 = [0.1039, 0.2044, 0.5087, 0.2874]$

$z_1 = (0.9)(0.1039) + (0.3)(0.2044) + (0.2)(0.5087) + (10)(0.2874) = 3.1305$

$a_1 = \sigma(3.1305) \approx 0.9582$

Hidden Neuron 2: $z_2 = (0.9)(0.2034) + (0.3)(0.1041) + (0.2)(0.1079) + (10)(0.2689)$

$= 2.9249$

$a_2 = \sigma(2.9242) \approx 0.9492$

Output predicted: → $\hat{y}_1 = (0.9427)(0.9582) + (0.7302)(0.9492)$

$= 1.5964$

$\text{Loss} = \frac{1}{2}(1.5964 - 2.7)^2 = 0.609$

Test input 2 → Input : [1.1, 0.3, 0.6, 13]

Neuron 1 → $z_1 = (1.1)(0.1039) + (0.3)(0.2044) + (0.6)(0.5087) + (13)(0.2874) = 4.217$

$a_1 = \sigma(4.217) \approx 0.9855$

Neuron 2 : $z_2 = (1.1)(0.2034) + (0.3)(0.1041) + (0.6)(0.1079) + (13)(0.2689) = 3.8153$

$a_2 = \sigma(3.8153) \approx 0.9785$

Output Prediction : → $\hat{y}_2 = (0.9427)(0.9855) + (0.7302)(0.9785) = 1.6494$

Error : $\frac{1}{2}(1.7 - 1.6494)^2 = 0.0016$

∴ Total MSE for ~~training~~ testing points $= 0.0016 + 0.61 = \boxed{0.6116}$

Total MSE for training points $= \boxed{1.0321}$