Problem Set 2:

Author: Xiaolan Li

Problem1:

$\because$ sigmoid $(x) = \frac{1}{1+e^{-x}}$

we have $\sigma(x) = \frac{1}{1+e^{-x}}$

derivation :

$$\frac{\partial \sigma(x)}{\partial x} = \frac{d}{dx}\sigma(x) = \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right]$$

$$= \frac{d}{dx}(1+e^{-x})^{-1}$$

$$= -(1+e^{-x})^{-2} \cdot \frac{d}{dx}(1+e^{-x})$$

$$= -(1+e^{-x})^{-2} \cdot (-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^{2}}$$

$$= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}} \cdot \frac{(1+e^{-x})-1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}} \cdot \left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right)$$

$$= \frac{1}{1+e^{-x}} \cdot \left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \sigma(x) \cdot (1-\sigma(x))$$

Problem2:

if we use the linear function as our activation function that $Thid_j = Whid_j$ and $Tout_k = Wout_k$

It means we use $y = x$ as activation function

We have $Whid_j = inp_i \times W_{ij} + bias\,H$

since $Thid_j = Whid_j$

$\therefore Wout_1 = [(inp_i \times W_{ij}) + biaH] \times W_{jk} + bias\,O$

since $Tout_k = Wout_k$

$\therefore Wout_1 = (W_{ij} \times W_{jk}) \times inp_i + (biaH \times W_{jk} + bias\,O)$

Which is equal to another linear function

$Wout_1 = w'\,inp + bias'$

so we don't need hidden layer anymore which is actually always computing the linear activation function in our network.

same as in $Wout_2$.

Problem3:

Part A:

$$\begin{aligned}
Hidden\ 1 &= inp1 \times W_{11} + inp2 \times W_{21} + inp3\ W_{31} + bias\,H \\
&= 4 \times 0.1 + 8 \times (-0.4) + 6 \times (-0.1) + 0.2 \\
&= -3.2
\end{aligned}$$

$$\begin{aligned}
Hidden\ 2 &= inp1 \times W_{12} + inp2 \times W_{22} + inp3\ W_{32} + bias\,H \\
&= 4 \times 0.3 + 8 \times 0.1 + 6 \times (-0.2) + 0.2 \\
&= 1
\end{aligned}$$

$$\begin{aligned}
Hidden\ 3 &= inp1 \times W_{13} + inp2 \times W_{23} + inp3\ W_{33} + bias\,H \\
&= 4 \times (-0.2) + 8 \times 0.2 + 6 \times 0.4 + 0.2 \\
&= 3.4
\end{aligned}$$

$\because sigmoid(x) = \frac{1}{1+e^{-x}}$

Thus $\begin{cases} sigmoid(-3.2) \approx 0.039 \\ sigmoid(1) \approx 0.731 \\ sigmoid(3.4) \approx 0.968 \end{cases}$

$$W_{out1} = T_{hid1} \times W_{11} + T_{hid2} \times W_{21} + T_{hid3} \times W_{31} + bias_O$$
$$= 0.039 \times 0.5 + 0.731 \times (-0.3) + 0.968 \times 0.2 + 0.3$$
$$= 0.2938$$

$$W_{out2} = T_{hid1} \times W_{12} + T_{hid2} \times W_{22} + T_{hid3} \times W_{32} + bias_O$$
$$= 0.039 \times 0.2 + 0.731 \times (-0.3) + 0.968 \times 0.1 + 0.3$$
$$= \cancel{0.2555}\ 0.1853$$

$$T_{out1} = \text{sigmoid}(0.2938) \approx 0.573$$

$$T_{out2} = \text{sigmoid}(0.2555) \approx \cancel{0.564}\ 0.546$$

input1	input2	input3	hidden1 ~~output1~~	hidden2	hidden3	output1	output2
4	8	6	0.039	0.731	0.968	0.573	~~0.564~~ 0.546

Part B:

$$\because E = 0.5 \times \sum_k (targ_k - T_{outk})^2$$

For node out1 : $0.65 - 0.573 = 0.077$

For node out2 : $0.4 - 0.546 = -0.146$

$$\therefore E = 0.5 \times (0.077^2 + (-0.146)^2) = 0.0136$$

Part C:

We have the gradient from hidden to output node.

$$g(E, W_{jk}) = (T_{outk} - targ_k) \times T_{outk} \times (1 - T_{outk}) \times T_{hidj}.$$

$$\therefore g(E, W_{11}) = (0.573 - 0.65) \times 0.573 \times (1 - 0.573) \times 0.039 = -0.00073$$

$$g(E, W_{12}) = (0.546 - 0.4) \times 0.546 \times (1 - 0.546) \times 0.039 = 0.00141$$

$$g(E, W_{21}) = (0.573 - 0.65) \times 0.573 \times (1 - 0.573) \times 0.731 = \cancel{0.71237}\ -0.0137$$

$$g(E, W_{22}) = (0.546 - 0.4) \times 0.546 \times (1 - 0.546) \times 0.731 = 0.02646$$

$$g(E, W_{31}) = (0.573 - 0.65) \times 0.573 \times (1 - 0.573) \times 0.968 = -0.01819$$

$$g(E, W_{32}) = (0.546 - 0.4) \times 0.546 \times (1 - 0.546) \times 0.968 = 0.03503$$

We have the gradient for the bias term for output node.

$$g(E, bias_O) = \sum_k [(T_{outk} - targ_k) \times T_{outk} \times (1 - T_{outk})] = (0.573 - 0.65) \times 0.573 \times (1 - 0.573) + (0.546 - 0.4) \times 0.546 \times (1 - 0.546)$$

$$= 0.01740$$

We have for all input nodes, inp_i and hidden nodes hid_j.

$g(E, W_{ij}) = g(E, Thid_j) \times Thid_j \times (1 - Thid_j) \times inp_i$

We have ~~g(E, Thid_1)~~ = $(Tout_1 - targ_1) \times Tout_1 \times (1 - Tout_1) = -0.01879$

$(Tout_2 - targ_2) \times Tout_2 \times (1 - Tout_2) = 0.03619$

$Thid_1 \times (1 - Thid_1) = 0.0375$

$Thid_2 \times (1 - Thid_2) = 0.1966$

$Thid_3 \times (1 - Thid_3) = 0.0310$

Thus $g(E, Thid_1) = (-0.01879) \times 0.5 + 0.03619 \times 0.2 = -0.0094 + 0.0072 = \cancel{0.0166}\ -0.0022$

$g(E, Thid_2) = (-0.01879) \times (-0.3) + 0.03619 \times (-0.3) = 0.0056 - 0.0109 = -0.0053$

$g(E, Thid_3) = (-0.01879) \times 0.2 + 0.03619 \times 0.1 = -0.00376 + 0.0036 = -0.0002$

Thus, $g(E, W_{11}) = \overset{-0.0022}{\cancel{0.0166}} \times 0.0375 \times 4 = \cancel{0.00025}\ -0.00033$

$g(E, W_{12}) = -0.0053 \times 0.1966 \times 4 = -0.00417$

$g(E, W_{13}) = -0.0002 \times 0.0310 \times 4 = -0.00002$

$g(E, W_{21}) = -0.0022 \times 0.0375 \times 8 = -0.00066$

$g(E, W_{22}) = -0.0053 \times 0.1966 \times 8 = -0.00834$

$g(E, W_{23}) = -0.0002 \times 0.0310 \times 8 = -0.00005$

$g(E, W_{31}) = -0.0022 \times 0.0375 \times 6 = -0.0005$

$g(E, W_{32}) = -0.0053 \times 0.1966 \times 6 = -0.00625$

$g(E, W_{33}) = -0.0002 \times 0.0310 \times 6 = -0.00004$

We have the bias term for the hidden nodes, gradient is

$g(E, bias\ H) = \sum_k [(Tout_k - targ_k) \times Tout_k \times (1 - Tout_k) \times w_{jk}] \times Thid_j \times (1 - Thid_j)$

$= \sum_j g(E, Thid_j) \times Thid_j \times (1 - Thid_j)$

$= -0.0022 \times 0.0375 + (-0.0053) \times 0.1966 + (-0.0002) \times 0.031$

$= -0.00008 - 0.00104 - 0.0000062$

$= -0.0011$

Link from node	Link to node	Gradient
input1	Hidden 1	-0.00033
input1	Hidden 2	-0.00417
input 1	Hidden 3	-0.00002
input 2	Hidden 1	-0.00066
input 2	Hidden 2	-0.00834
input 2	Hidden 3	-0.00005
input 3	Hidden 1	-0.0005
input 3	Hidden 2	-0.00625
input 3	Hidden3	-0.00004
Hidden1	output 1	-0.00073
Hidden 1	output 2	0.00141
Hidden2	output 1	-0.0137
Hidden2	output 2	0.02646
Hidden3	output1	-0.01819
Hidden3	output2	0.03503
	layer	(Bias)
	hidden	-0.0011
	output	0.0174

Part D:

Link from node	Link to node	Weight
input1	Hidden 1	0.100066
input1	Hidden 2	0.30083
input 1	Hidden 3	−0.19917
input 2	Hidden 1	−0.4
input 2	Hidden 2	0.10167
input 2	Hidden 3	0.20001
input 3	Hidden 1	−0.0999
input 3	Hidden 2	−0.19875
input 3	Hidden3	0.4
Hidden1	output 1	0.50015
Hidden 1	output 2	0.19972
Hidden2	output1	−0.29726
Hidden2	output 2	−0.30529
Hidden3	output1	0.20364
Hidden3	output 2	0.093
	layer	Weight
	hidden	0.20022
	output	0.29652

Part A Redux: Forward Propagation:

$$\text{Hidden}_1 = \text{inp}1 \times W_{11} + \text{inp}2 \times W_{21} + \text{inp}3\ W_{31} + \text{bias}\ H$$
$$= 4 \times 0.100066 + 8 \times (-0.4) + 6 \times (-0.0999) + 0.20022$$
$$= -3.198936$$

$$\text{Hidden}_2 = \text{inp}1 \times W_{12} + \text{inp}2 \times W_{22} + \text{inp}3 \times W_{32} + \text{bias}\ H$$
$$= 4 \times 0.30083 + 8 \times 0.10167 + 6 \times (-0.19875) + 0.20022$$
$$= 1.0244$$

$$\text{Hidden}_3 = \text{inp}1 \times W_{13} + \text{inp}2 \times W_{23} + \text{inp}3\ W_{33} + \text{bias}\ H$$
$$= 4 \times (-0.19917) + 8 \times 0.20001 + 6 \times 0.4 + 0.20022$$
$$= 3.40362$$

Thus
$$\begin{cases} \text{sigmoid}(-3.198936) \approx 0.0392058 \\ \text{sigmoid}(1.0244) \approx 0.7358288 \\ \text{sigmoid}(3.40362) \approx 0.9678175 \end{cases}$$

$$W_{out1} = T_{hid1} \times W_{11} + T_{hid2} \times W_{21} + T_{hid3} \times W_{31} + \text{bias}\ O$$
$$= 0.0392058 \times 0.50015 + 0.7358288 \times (-0.29726) + 0.9678175 \times 0.20364 + 0.29652$$
$$= 0.01960878 - 0.21873247 + 0.19708636 + 0.29652 = 0.29448$$

$$W_{out2} = T_{hid1} \times W_{12} + T_{hid2} \times W_{22} + T_{hid3} \times W_{32} + \text{bias}\ O$$
$$= 0.0392058 \times 0.19972 + 0.7358288 \times (-0.30529) + 0.9678175 \times 0.093 + 0.29652$$
$$= 0.00783018 - 0.22464117 + 0.09000703 + 0.29652 = 0.169716$$

~~Wout 3 = Thid1 ×~~

$$\begin{cases} T_{out1} = \text{sigmoid}(W_{out1}) = 0.5730926 \\ T_{out2} = \text{sigmoid}(W_{out2}) = 0.5423275 \end{cases}$$

input1	input2	input3	hidden1	hidden2	hidden3	out1	out2
4	8	6	0.0392058	0.7358288	0.9678175	0.5730926	0.5423275

Part B Redux: Determine the Error

$$\because E = 0.5 \times \sum_K (\text{targ}K - T_{outK})^2$$

For node out1 : $0.65 - 0.5730926 = 0.0769074$

For node out2 : $0.4 - 0.5423275 = -0.1423275$

$$\therefore E = 0.5 \times (0.0769074^2 + (-0.1423275)^2)$$
$$= 0.5 \times (0.005914748 + 0.0202571)$$
$$= 0.0130859$$

After compared to the previous error, the new error decrease from 0.0136 to 0.0130859 that has 3.8% decrease.