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

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

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

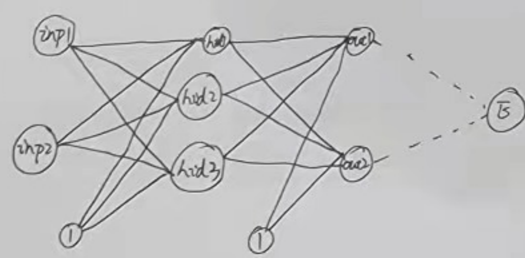
$$= \frac{1}{(1+e^{-x})}\cdot\frac{e^{-x}}{1+e^{-x}}$$

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

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

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

② problem 2



When we use identity function as activation function

$f(x) = x$

$$
\begin{aligned}
Tout1 = Wout1 &= W_{11} \cdot Thid1 + W_{21} \cdot Thid2 + W_{31} \cdot Thid3 + bias\ O \\
&= W_{11} \cdot Whid1 + W_{21} \cdot Whid2 + W_{31} \cdot Whid3 + bias\ O \\
&= W_{11} \cdot (w_{11} \cdot inp1 + w_{21} \cdot inp2 + bias\ H) \\
&+ W_{21} \cdot (w_{12} \cdot inp1 + w_{22} \cdot inp2 + bias\ H \\
&+ W_{31} \cdot (w_{13} \cdot inp1 + w_{23} \cdot inp2 + bias\ H) + bias\ O
\end{aligned}
$$

$$
\begin{aligned}
Tout2 = Wout2 &= W_{12}\, Thid1 + W_{22}\, Thid2 + W_{32}\, Thid3 + bias\ O \\
&= W_{12}\, Whid1 + W_{22}\, Whid2 + W_{32}\, Whid3 + bias\ O \\
&= W_{12}\, (w_{11}\, inp1 + w_{21}\, inp2 + bias\ H) \\
&+ W_{22}\, (w_{12}\, inp1 + w_{22}\, inp2 + bias\ H) \\
&+ W_{32}\, (w_{13}\, inp1 + w_{23}\, inp2 + bias\ H) + bias\ O
\end{aligned}
$$

③ problem 3

part A:

input1	input2	input3	hidden1	hidden2	hidden3	output1	output2
4	8	6	0.0392	0.7311	0.9677	0.5779 ~~0.5792~~	0.5462

part B:

$E \approx 0.01366$

④

part C:

$$E = Toutk - targk$$

$$g(E, W_{jk}) = \frac{\partial E}{\partial W_{jk}} = \frac{\partial(Toutk - \cancel{targ} \ targk)}{\partial W_{jk}}$$

$$= (Toutk - targk)\frac{\partial Toutk}{\partial W_{jk}}$$

$$= (Toutk - targk) \cdot \frac{\partial\left(\frac{1}{1+e^{-Woutk}}\right)}{\partial W_{jk}}$$

$$= (Toutk - targk) \cdot \frac{1}{1+e^{-Woutk}} \cdot \left(1 - \frac{1}{1+e^{-Woutk}}\right) \cdot \frac{\partial(Woutk)}{\partial W_{jk}}$$

$$= \cancel{\partial}\,(Toutk - targk) \cancel{\times} \, Toutk \cdot (1 - Toutk) \cdot \frac{\partial(W_{1k}Thid_1 + W_{2k}Thid_2 + W_{3k}Thid_3 + biaso)}{\partial W_{jk}}$$

$$= (Toutk - targk) \, Toutk \cdot (1 - Toutk) \cdot Thidj.$$

5

$$g(E, bias0) = \frac{\partial E}{\cancel{\partial W_{ij}}\ \partial bias0}$$

$$= \sum_k \left[\frac{\partial (Tout_k - targ_k)}{\partial bias0} \right]$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot \frac{\partial Tout_k}{\partial bias0} \right]$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot \frac{1}{1+e^{-Wout_k}} \cdot \left(1 - \frac{1}{1+e^{-Wout_k}}\right) \frac{\partial (Wout_k)}{\partial bias0} \right]$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k (1 - Tout_k) \cdot \frac{\partial (W_{1k} Thid_1 + W_{2k} Thid_2 + W_{3k} Thid_3 + bias0)}{\partial bias0} \right]$$

$$= \sum_k \left[(Tout_k - targ_k)\, Tout_k \cdot (1 - Tout_k) \cdot \cancel{Thid_j} \cdot 1 \right]$$

$$\cancel{g(E, W_{ij}) = \sum_k \partial(0.5 * (targ_k}$$

$$g(E, w_{ij}) = \sum_k \frac{\partial (0.5 * (Tout_k - targ_k)^2}{\partial \cancel{(Tout_k)}\ \partial w_{ij}}$$

$$= 0.5 \times 2 \times \sum_k (Tout_k - targ_k) \cdot \frac{\partial Tout_k}{\partial w_{ij}}$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot \frac{\partial \left(\frac{1}{1+e^{-Wout_k}}\right)}{\partial W_{jk}} \right]$$

$$\cancel{= \sum_k (Tout_k - targ_k) \cdot \frac{\partial (W_{1k} Thid_1 + W_{2k} Thid_2 + W_{3k} Thid_3 + bias0)}{\partial W_{jk}}}$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot \frac{\ }{\cancel{\partial W_{jk}}\ \partial w_{ij}} \frac{1}{1+e^{-Wout_k}} \cdot \left(1 - \frac{1}{1+e^{Wout_k}}\right) \cdot \frac{\partial (Wout_k)}{\partial w_{ij}} \right]$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot \frac{\partial (W_{1k} Thid_1 + W_{2k} Thid_2 + W_{3k} Thid_3 + bias0)}{\partial W_{ij}} \right]$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot W_{jk} \right] \cdot \frac{\partial Thid_j}{\partial w_{ij}}$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot W_{jk} \right] \cdot \frac{1}{1+e^{-Whid_j}} \cdot \left(1 - \frac{1}{1+e^{-Whid_j}}\right) \frac{\partial Whid_j}{\partial w_{ij}}$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot W_{jk} \right] \cdot Thid_j \cdot (1 - Thid_j) \cdot \frac{\partial (w_{1j} \cdot inp_1 + \cancel{w_{2j} \cdot inp_2} + bias_H)}{\partial w_{ij}}$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot W_{jk} \right] \cdot Thid_j \cdot (1 - Thid_j) \cdot inp_i$$

⑥

$$g(E, biasH) = \sum_k \frac{\partial (0.5 * Tout_k - targ_k)^2}{\partial\, biasH}$$

$$= 0.5 \times 2 * \sum_k (Tout_k - targ_k) \cdot \frac{\partial \left(\frac{1}{1+e^{-Wout_k}} \right)}{\partial\, biasH}$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot \frac{1}{1+e^{-Wout_k}} \cdot \left(1 - \frac{1}{1+e^{Wout_k}}\right) \cdot \frac{\partial (Wout_k)}{\partial\, biasH} \right]$$

$$= \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot \frac{\partial (W_{1k} Thid_1 + W_{2k} Thid_2 + W_{3k} Thid_3 + biasO)}{\partial\, biasH} \right]$$

$$= \sum_j \sum_k \left[(Tout_k - targ_k) \cdot Tout_k (1 - Tout_k) \cdot W_{jk} \right] \cdot \frac{\partial\, Thid_j}{\partial\, biasH}$$

$$= \sum_j \sum_k \left[(Tout_k - targ_k) \cdot Tout_k \cdot (1 - Tout_k) \cdot W_{jk} \right] \cdot \frac{1}{1+e^{-Whid_j}} \cdot \left(1 - \frac{1}{1+e^{-Whid_j}}\right) \cdot \frac{\partial\, Whid_j}{\partial\, biasH}$$

$$= \sum_j \sum_k \left[(Tout_k - targ_k) \cdot Tout_k (1 - Tout_k) \cdot W_{jk} \right] \cdot Thid_j (1 - Thid_j) \cdot$$

link from node	link to node	gradient
input 1	hidden 1	-0.0003289
input 1	hidden 2	-0.004098
input 1	hidden 3	-0.0000186
input 2	hidden 1	-0.00065
input 2	hidden 2	-0.008196
input 2	hidden 3	~~3.~~ -0.0000373
input 3	hidden 1	-0.000494
input 3	hidden 2	~~6.~~ -0.006148
input 3	hidden 3	-0.00002799
hidden 1	out 1	-0.0007
hidden 1	out 2	0.0014
hidden 2	out 1	-0.0138
hidden 2	out 2	0.0265
hidden 3	out 1	-0.0183
hidden 3	out 2	0.0351

	layer	Bias
	Hidden	-0.0011
	output	0.0174

(8) part D

new weights		
link from node	link to node	weight
input1	hidden1	0.10006584
input1	hidden2	0.30081968
input1	hidden3	-0.19999627
input2	hidden1	-0.39986833
input2	hidden2	0.10163937
input2	hidden3	0.20000746
input3	hidden1	-0.09990124
input3	hidden2	-0.19877048
input3	hidden3	0.4000056
hidden1	out1	0.5001479
hidden1	out2	0.19971589
hidden2	out1	-0.29724152
hidden2	out2	-0.30529871
hidden3	out1	0.20365118
hidden3	out2	0.09298651

new BIAS weights		
	layer	weight
	Hidden	0.2002
	Output	0.2965

⑨ Part A Redux

input1	input2	input3	hidden1	hidden2	hidden3	out1	out2
4	8	6	0.0392	0.7357	0.9677	0.5140	0.5432

Part B Redux

$E = 0.013147$

The new error ~~decress~~ deerease ~~de~~ due to we use ~~The new~~ the gradient to optimize the weight and let it predict more ~~a~~ accuracy. So the error ~~decress~~. decrease.