Neural Network and optimization of calculation

LSI Design Contest
This is an example of Neural network calculation.
The neural network structure used here is 3-layer structure.
It consist of 2 input units, 3 hidden units and 2 output units.
• Students need to pass both Mathematics and Physics examination in order to pass [1,0] the spring semester
• Passing mark for each exam is 6 and total mark is 10
• Students who only manage to pass only one exam is considered as failed [0,1]
Score sample

Score sample ①

Score sample ②

Score sample ③

Score sample ④
• From above example, the result for four students are as below.

<table>
<thead>
<tr>
<th>Students</th>
<th>Math. Score $K_1^i$</th>
<th>Physics Score $K_2^i$</th>
<th>Supervisor $[t_1, t_2]$</th>
<th>Results</th>
<th>Output value $[a_1^3, a_2^3]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>8</td>
<td>[1,0]</td>
<td>Pass</td>
<td>$[0.76, 0.69]$</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>5</td>
<td>[0,1]</td>
<td>Fail</td>
<td>$[0.75, 0.68]$</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>8</td>
<td>[0,1]</td>
<td>Fail</td>
<td>$[0.75, 0.67]$</td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>5</td>
<td>[0,1]</td>
<td>Fail</td>
<td>$[0.72, 0.66]$</td>
</tr>
</tbody>
</table>

• $z_i^2 = w_{11}^2 k_1 + w_{21}^2 k_2 + b_1^2$

• square error: $C_i = \frac{1}{2} \{(a_1^3[i] - t_1)^2 + (a_2^3[i] - t_2)^2\}$

• cost function: $C = C_1 + C_2 + \cdots + C_i + \cdots$

• $\delta_1^i = (a_1^3 - t_1) a' (z_i^2)$

• $\delta_2^i = (\delta_1^3 w_{11}^3 + \delta_2^3 w_{21}^3 + \cdots) a' (z_i^2)$

• $\frac{\partial C}{\partial w_{ij}^m} = \delta_j^m a_j^{m-1}$

• $\frac{\partial C}{\partial b_j^m} = \delta_j^m$
Equation of z and a

\[ z_i^2 = w_1^2 i k_1 + w_2^2 i k_2 + b_i^2 \]

\[ a_i^2 = a(z_i^2) = \frac{1}{1+e^{-z_i^2}} \]

\[ a'(z_i^2) = \frac{e^{-z_i^2}}{(e^{-z_i^2} + 1)^2} \]

\[ z_i^3 = w_1^3 i a_i^2 + w_2^3 i a_i^2 + w_3^3 i a_i^2 + b_i^3 \]

\[ a_i^3 = a(z_i^3) = \frac{1}{1+e^{-z_i^3}} \]

\[ a'(z_i^3) = \frac{e^{-z_i^3}}{(e^{-z_i^3} + 1)^2} \]
Gradient descent

- C is all learning data’s error.
⇒ You find weight and bias that make minimize C to differentiate C with respect to weight and bias.
⇒ Gradient descent

\( (\Delta w_{11}^2, \Delta w_{11}^3, \Delta b_1^2, \Delta b_1^3) = -\eta \left( \frac{\partial C}{\partial w_{11}^2}, \frac{\partial C}{\partial w_{11}^3}, \frac{\partial C}{\partial b_1^2}, \frac{\partial C}{\partial b_1^3} \right) \)

If this numerical expression satisfy, C is the most smaller.

But it is too difficult to calculate this parameter.
If numbers of input units: 10, hidden units: 10, output units: 3, numbers of parameter is 10*10 + 10(bias) + 10*3 + 3(bias) = 143.
⇒ want to calculate at all once, if it’s possible decrease difference…
⇒ Back propagation.
Solve Gradient descent

- Partial differential term can be generalized by below method.

\[
\frac{\partial C}{\partial w_i} = \frac{\partial C}{\partial z_1} \frac{\partial z_1}{\partial w_1} = \frac{\partial C}{\partial z_1} k_1 \\
(\because z_1 = w_1 k_1 + w_2 k_2 + w_3 k_3 + \cdots + b) \\
\text{define } \frac{\partial C}{\partial z_1} = \delta_1 \rightarrow \frac{\partial C}{\partial w_1} = \delta_1 k_1 \\
\text{while } \frac{\partial C}{\partial b_1} = \frac{\partial C}{\partial z_1} \frac{\partial z_1}{\partial b_1} = \delta_1 \\
\therefore \frac{\partial C}{\partial w_{ij}} = \delta_j a_j^{m-1}, \frac{\partial C}{\partial b_j} = \delta_j \quad (m=2 \text{ or } 3) \\
(a_j = k_j)
\]
Back propagation is optimization of calculation.

Calculate all parameters to return output to input.
Calculation of $\delta$

- $\delta$ is errors. It is unit’s error. If you can express $\delta$ in known parameter, you don’t have to differential the calculation.

It is easy to calculate output layer errors.

$$\delta^3_1 = \frac{\partial C}{\partial z^3_1} = \frac{\partial C}{\partial a^3_1} \frac{\partial a^3_1}{\partial z^3_1} = \frac{\partial C}{\partial a^3_1} a'(z^3_1)$$

$\therefore\delta^3_1 = (a^3_1 - t_1)a'(z^3_1)$

$\because C$ is made of the difference between $t$ and $a_3$.

But hidden layer errors is more difficult to calculate

$$\delta^2_1 = \frac{\partial C}{\partial z^2_1} = \frac{\partial C}{\partial a^2_1} \frac{\partial a^2_1}{\partial z^2_1} + \frac{\partial C}{\partial z^2_2} \frac{\partial a^2_1}{\partial z^2_2} + \cdots$$

$$(\frac{\partial C}{\partial z^2_1} = \delta^3_1, \frac{\partial C}{\partial z^2_2} = \delta^3_2, \cdots \frac{\partial z^3_1}{\partial a^3_1} = w^3_{11}, \frac{\partial z^3_2}{\partial a^3_1} = w^3_{21}, \cdots)$$

$$\delta^2_1 = \delta^3_1w^3_{11}a'(z^2_1) + \delta^3_2w^3_{21}a'(z^2_1) + \cdots$$

$\therefore\delta^2_1 = (\delta^3_1w^3_{11} + \delta^3_2w^3_{21} + \cdots)a'(z^2_1)$
Mathematics score

Physics score

$z_1^2 \rightarrow a_1^2$

$z_2^2 \rightarrow a_2^2$

$z_3^2 \rightarrow a_3^2$

Bias=-1

Bias=-1

Bias=-1

$z_1^3 \rightarrow a_1^3$

$z_2^3 \rightarrow a_2^3$

$z_3^3 \rightarrow a_3^3$

ans

ans

ans
Score: [8,8] (z2,a2,z3,a3)

\[ z_i^2 = w_1^2 a_i^2 + w_2^2 a_i^2 + b_i^2 \]

\[ a_i^2 = a(z_i^2) = \frac{1}{1 + e^{-z_i^2}} \]

\[ z_i^3 = w_1^3 a_i^3 + w_2^3 a_i^3 + w_3^3 a_i^3 + b_i^3 \]

\[ a_i^3 = a(z_i^3) = \frac{1}{1 + e^{-z_i^3}} \]
Mathematic score = 8

Physics score = 8

\[ \delta_1^2 = -4.811e-5 \]

\[ \delta_2^2 = 2.9164e-5 \]

\[ \delta_3^2 = 0.0010 \]

\[ \delta_1^3 = -0.0438 \]

\[ \delta_2^3 = 0.1478 \]

\[ \delta_3^3 = 0.0438 \]

\[ \delta b_i^2 = \delta_i^2 \]

\[ \delta b_i^3 = \delta_i^3 \]
Score: [8, 5] (z2, a2, z3, a3)

Mathematic score = 8

Physics score = 5

\[ z_i^2 = w_i^2 a_i^1 + w_i^2 a_i^2 + w_i^2 a_i^3 + b_i^2 \]

\[ a_i^2 = a(z_i^2) = \frac{1}{1 + e^{-z_i^2}} \]

\[ z_i^3 = w_i^3 a_i^1 + w_i^3 a_i^2 + w_i^3 a_i^3 + b_i^3 \]

\[ a_i^3 = a(z_i^3) = \frac{1}{1 + e^{-z_i^3}} \]
Score : [8, 5] \( (\delta_1^2, \delta_2^2, \delta_3^2, \delta_1^3, \delta_2^3) \)

\[ C_1 = 0.2635 \]
\[ C_2 = 0.3302 \]
\[ c = 0.5937 \]

\[ \delta_1^3 = (a_1^3 - t_1) a'(z_1^3) \]
\[ \delta_1^2 = (\delta_1^3 w_1^3 + \delta_2^3 w_2^3) a'(z_1^2) \]
\[ \delta_2^3 = \delta w_1^3 \]
\[ \delta_2^2 = \delta w_2^2 \]
\[ \delta b_1^3 = b_1^3 + \delta_1^3 \]
\[ \delta b_1^2 = b_1^2 + \delta_1^2 \]

Mathe matic score = 8

Physics score = 5

\( \delta_1^2 = 0.0103 \)
\( \delta_2^2 = -1.3233e - 4 \)
\( \delta_3^2 = 0.0014 \)
\( \delta_1^3 = 0.1413 \)
\( \delta_2^3 = -0.0701 \)

\( \delta_1' z_1^3 \)
\( \delta_2' z_1^2 \)
\( \delta w_1^3 \)
\( \delta w_2^2 \)
\( \delta b_1^3 \)
\( \delta b_1^2 \)
Score: [5, 8] (z2, a2, z3, a3)

\[ z_i^2 = w_1^2 i k_1 + w_2^2 i k_2 + b_i^2 \]

\[ a_i^2 = a(z_i^2) = \frac{1}{1 + e^{-z_i^2}} \]

\[ z_i^3 = w_1^3 i a_i^2 + w_2^3 i a_i^2 + w_3^3 i a_i^2 + b_i^3 \]

\[ a_i^3 = a(z_i^3) = \frac{1}{1 + e^{-z_i^3}} \]
Mathematical score = 5

Physics score = 5

Bias = -1

\[ z_i^2 = w_{1i}k_1 + w_{2i}k_2 + b_i^2 \]

\[ a_i^2 = a(z_i^2) = \frac{1}{1 + e^{-z_i^2}} \]

\[ z_i^3 = w_{1i}a_i^2 + w_{2i}a_i^2 + w_{3i}a_i^2 + b_i^3 \]

\[ a_i^3 = a(z_i^3) = \frac{1}{1 + e^{-z_i^3}} \]
Score : [5,5] $\left( \delta_1^2, \delta_2^2, \delta_3^2, \delta_1^3, \delta_2^3 \right)$

Mathe
matic
score
=5

0.1
0.4
1.5
0.8176
0.7240

$\delta_1^2 = 0.0128$

$\delta_2^2 = -4.2792e^{-4}$

$\delta_3^2 = 0.0073$

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$

 Physics
score
=5

0.6
0.1
2.5
0.9241

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$

$\delta_1^2 = 0.0128$

$\delta_2^2 = -4.2792e^{-4}$

$\delta_3^2 = 0.0073$

$\delta_1^2 = 0.0128$

$\delta_2^2 = -4.2792e^{-4}$

$\delta_3^2 = 0.0073$

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$

$\delta_1^2 = 0.0128$

$\delta_2^2 = -4.2792e^{-4}$

$\delta_3^2 = 0.0073$

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$

$\delta_1^2 = 0.0128$

$\delta_2^2 = -4.2792e^{-4}$

$\delta_3^2 = 0.0073$

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$

$\delta_1^2 = 0.0128$

$\delta_2^2 = -4.2792e^{-4}$

$\delta_3^2 = 0.0073$

$\delta_1^3 = 0.1447$

$\delta_2^3 = -0.0768$

$\delta_3^3 = 0.1447$
From the above result, we can see that the output value is far from the state value that we are finding. So, to obtain a suitable output value, we use Back Propagation.
New parameter (after 1\textsuperscript{st} renewal)

\[
\begin{align*}
\begin{array}{c}
\text{Mathematic score} \\
\text{Physics score}
\end{array}
\end{align*}
\]

\[
\begin{align*}
&z_1^2 \quad a_1^2 \\
&z_2^2 \quad a_2^2 \\
&z_3^2 \quad a_3^2
\end{align*}
\]

\[
\begin{align*}
&z_1^3 \quad a_1^3 \\
&z_2^3 \quad a_2^3 \\
&z_3^3 \quad a_3^3
\end{align*}
\]

- (new)\(w_{ij}^3 = (\text{old})w_{ij}^3 - \eta \times \delta w_{ij}^3\)
- (new)\(w_{ij}^2 = (\text{old})w_{ij}^2 - \eta \times \delta w_{ij}^2\)
- (new)\(b_{ij}^3 = (\text{old})b_{ij}^3 - \eta \times \delta b_{ij}^3\)
- (new)\(b_{ij}^2 = (\text{old})b_{ij}^2 - \eta \times \delta b_{ij}^2\)

\[\therefore \eta = 0.1\]
New parameter (after 10000\textsuperscript{th} renewal)

- \((new)w_{i,j}^3 = (old)w_{i,j}^3 - \eta \cdot \delta w_{i,j}^3\)
- \((new)w_{i,j}^2 = (old)w_{i,j}^2 - \eta \cdot \delta w_{i,j}^2\)
- \((new)b_{i,j}^3 = (old)b_{i,j}^3 - \eta \cdot \delta b_{i,j}^3\)
- \((new)b_{i,j}^2 = (old)b_{i,j}^2 - \eta \cdot \delta b_{i,j}^2\)

\[ \therefore \eta = 0.1 \]
After some calculation, we can see that the value of weight and bias is changing simultaneously with the input value. This calculation is repeated until the different value between output and supervisor value become smaller.