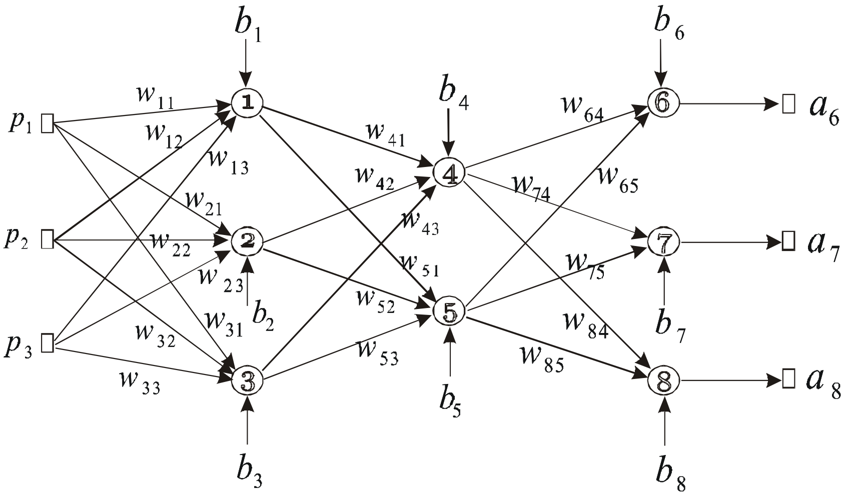
Metoda “delta” uczenia wielowarstwowych sieci neuronowych

za pomocą wstecznej propagacji błędów

Jacek Kluska

Politechnika Rzeszowska

Przykładowa architektura sieci neuronowej



Rysunek: Sieć trójwarstwowa. Oznaczenia: w_{ij} – waga i-go neuronu od j-go wejścia, b_i - przesunięcie i-go neuronu.

Poszukiwane parametry sieci neuronowej

1. Dla neuronów pierwszej warstwy należy wyznaczyć:

$$W_1 = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}, \qquad B_1 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

2. Dla neuronów drugiej warstwy należy wyznaczyć:

$$W_2 = \begin{bmatrix} w_{41} & w_{42} & w_{43} \\ w_{51} & w_{52} & w_{53} \end{bmatrix}, \qquad B_2 = \begin{bmatrix} b_4 \\ b_5 \end{bmatrix}$$

3. Dla neuronów trzeciej warstwy należy wyznaczyć:

$$W_3 = \begin{bmatrix} w_{64} & w_{65} \\ w_{74} & w_{75} \\ w_{84} & w_{85} \end{bmatrix}, \qquad B_3 = \begin{bmatrix} b_6 \\ b_7 \\ b_8 \end{bmatrix}$$

Poszukujemy więc $(3 \times 3 + 3) + (2 \times 3 + 2) + (3 \times 2 + 3) = 29$ parametrów.

Postawienie problemu

Problem

Dla danych uczących:

$$\left\{ \left(\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \begin{bmatrix} t_6 \\ t_7 \\ t_8 \end{bmatrix} \right)_1, \ldots \left(\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \begin{bmatrix} t_6 \\ t_7 \\ t_8 \end{bmatrix} \right)_Q \right\} \subset \mathbb{R}^3 \times \mathbb{R}^3$$

chcemy znaleźć taki algorytm uczenia wszystkich parametrów w_{ij} oraz b_i, aby

$$E(k) = \frac{1}{2} \sum_{r \in \{6,7,8\}} (t_r - a_r)^2 \longrightarrow \min \quad , \qquad \forall\, k$$

Uczenie parametrów neuronu z warstwy trzeciej

$$w_{ij}(t+1) = w_{ij}(t) - \gamma \frac{\partial E}{\partial w_{ij}} \Leftrightarrow w'_{ij} = w_{ij} - \gamma \frac{\partial E}{\partial w_{ij}}, \quad \gamma > 0.$$

Rozpatrzmy wagę z warstwy trzeciej, np. w_{84}:

$$w'_{84} = w_{84} - \gamma \frac{\partial E}{\partial w_{84}} = w_{84} - \gamma \frac{\partial E}{\partial a_8} \frac{\partial a_8}{\partial n_8} \frac{\partial n_8}{\partial w_{84}}$$

gdzie

$$\frac{\partial E}{\partial a_8} = -(t_8 - a_8) = -e_8,$$

$$\frac{\partial a_8}{\partial n_8} = \frac{\partial f_8(n_8)}{\partial n_8} = f_8'(n_8) = f_8',$$

$$\frac{\partial n_8}{\partial w_{84}} = \frac{\partial (w_{84} a_4 + w_{85} a_5 + b_8)}{\partial w_{84}} = a_4.$$

Uczenie parametrów neuronu z warstwy trzeciej - c.d.

$$w'_{84} = w_{84} + \gamma \underbrace{e_8 f'_8\left(n_8\right)}_{\delta_8} a_4 = w_{84} + \gamma \delta_8 a_4$$

Łatwo zauważyć, że

$$b'_8 = b_8 - \gamma \frac{\partial E}{\partial b_8} = b_8 - \gamma \frac{\partial E}{\partial a_8} \frac{\partial a_8}{\partial n_8} \frac{\partial n_8}{\partial b_8} = b_8 + \gamma \delta_8$$

Obowiązuje następująca procedura uczenia parametrów warstwy trzeciej

$$\boxed{\left.\begin{aligned} e_i &= t_i - a_i \\ \delta_i &= e_i f'_i\left(n_i\right) \\ b'_i &= b_i + \gamma \delta_i \\ w'_{ij} &= w_{ij} + \gamma \delta_i a_j \end{aligned}\right\} \quad \forall\left(i, j\right) \in \{6, 7, 8\} \times \{4, 5\}}$$

Dwie uwagi

1. Wzory

$$b_8' = b_8 - \gamma \frac{\partial E}{\partial a_8} \frac{\partial a_8}{\partial n_8} \frac{\partial n_8}{\partial b_8}, \;\; w_{84}' = w_{84} - \gamma \frac{\partial E}{\partial a_8} \frac{\partial a_8}{\partial n_8} \frac{\partial n_8}{\partial w_{84}}$$

wynikają bezpośrednio ze związków przyczynowo-skutkowych

$$b_8 \quad \rightarrow \quad n_8 \quad \rightarrow \quad a_8 \quad \rightarrow \quad E \;, \quad w_{84} \quad \rightarrow \quad n_8 \quad \rightarrow \quad a_8 \quad \rightarrow \quad E$$

2. Funkcje aktywacji muszą być różniczkowalne, np.

   1. $$f(n) = \frac{1}{1 + \exp(-pn)}, \;\; p > 0$$

   2. $$f(n) = \frac{1}{\pi} \left(\frac{\pi}{2} + \operatorname{arc\,tg}(pn) \right), \;\; p > 0.$$

Uczenie parametrów neuronu z warstwy drugiej

Poszukujemy w'_{41}. Czyli $w'_{41} = w_{41} - \gamma \frac{\partial E}{\partial w_{41}}$. Graf przyczynowo-skutkowy

$$
\begin{array}{ccccccccccc}
 & & & & & & n_6 & \rightarrow & a_6 & & \\
 & & & & & \nearrow & & & & \searrow & \\
w_{41} & \rightarrow & n_4 & \rightarrow & a_4 & \rightarrow & n_7 & \rightarrow & a_7 & \rightarrow & E \\
 & & & & & \searrow & & & & \nearrow & \\
 & & & & & & n_8 & \rightarrow & a_8 & &
\end{array}
$$

Czyli

$$
\begin{aligned}
\frac{\partial E}{\partial w_{41}} &= \frac{\partial E}{\partial a_6}\frac{\partial a_6}{\partial n_6}\frac{\partial n_6}{\partial a_4}\frac{\partial a_4}{\partial n_4}\frac{\partial n_4}{\partial w_{41}} + \frac{\partial E}{\partial a_7}\frac{\partial a_7}{\partial n_7}\frac{\partial n_7}{\partial a_4}\frac{\partial a_4}{\partial n_4}\frac{\partial n_4}{\partial w_{41}} + \\
&\quad + \frac{\partial E}{\partial a_8}\frac{\partial a_8}{\partial n_8}\frac{\partial n_8}{\partial a_4}\frac{\partial a_4}{\partial n_4}\frac{\partial n_4}{\partial w_{41}} \\
&= -\left(t_6 - a_6\right) f_6'\left(n_6\right) w_{64} f_4'\left(n_4\right) a_1 - \left(t_7 - a_7\right) f_7'\left(n_7\right) w_{74} f_4'\left(n_4\right) a_1 + \\
&\quad - \left(t_8 - a_8\right) f_8'\left(n_8\right) w_{84} f_4'\left(n_4\right) a_1
\end{aligned}
$$

Uczenie parametrów neuronu z warstwy drugiej - c.d.

Otrzymujemy

$$\frac{\partial E}{\partial w_{41}} = -\underbrace{\left[e_6 f_6'(n_6)\, w_{64} + e_7 f_7'(n_7)\, w_{74} + e_8 f_8'(n_8)\, w_{84}\right] f_4'(n_4)}_{\delta_4} a_1 = -\delta_4 a_1,$$

$$\frac{\partial E}{\partial b_4} = -\underbrace{\left[e_6 f_6'(n_6)\, w_{64} + e_7 f_7'(n_7)\, w_{74} + e_8 f_8'(n_8)\, w_{84}\right] f_4'(n_4)}_{\delta_4} = -\delta_4,$$

Można wykazać, że

$$\delta_i = \left(\sum_{k \in \{6,7,8\}} e_k f_k'(n_k)\, w_{ki} \right) f_i'(n_i), \quad i \in \{4, 5\}$$

Uczenie parametrów neuronu z warstwy drugiej - c.d.

Obowiązuje następująca procedura uczenia parametrów warstwy drugiej

$$\boxed{\left.\begin{array}{l} e_i = t_i - a_i \\ \delta_i = \left(\sum\limits_{k \in \{6,7,8\}} e_k f_k' \left(n_k \right) w_{ki} \right) f_i' \left(n_i \right) \\ b_i' = b_i + \gamma \delta_i \\ w_{ij}' = w_{ij} + \gamma \delta_i a_j \end{array} \right\} \ \forall \ (i, j) \in \{4, 5\} \times \{6, 7, 8\}}$$

Uczenie parametrów neuronu z warstwy pierwszej

Poszukujemy w'_{32}. Czyli $w'_{32} = w_{32} - \gamma \frac{\partial E}{\partial w_{32}}$.
Graf przyczynowo-skutkowy

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & & & & n_6 & \rightarrow & a_6 & & \\
 & & & & & & & & & \nearrow & & & & \searrow & \\
 & & & & & & n_4 & \rightarrow & a_4 & \rightarrow & n_7 & \rightarrow & a_7 & \rightarrow & E \\
 & & & & & \nearrow & & & & \searrow & & & & \nearrow & \\
w_{32} & \rightarrow & n_3 & \rightarrow & a_3 & & & & & & n_8 & \rightarrow & a_8 & & \\
 & & & & & \searrow & & & & \nearrow & & & & \searrow & \\
 & & & & & & n_5 & \rightarrow & a_5 & \rightarrow & n_7 & \rightarrow & a_7 & \rightarrow & E \\
 & & & & & & & & & \searrow & & & & \nearrow & \\
 & & & & & & & & & & n_6 & \rightarrow & a_6 & &
\end{array}
$$

Uczenie parametrów neuronu z warstwy pierwszej - c.d.

Otrzymujemy

$$\frac{\partial E}{\partial w_{32}} = \frac{\partial a_4}{\partial n_4}\frac{\partial n_4}{\partial a_3}\frac{\partial a_3}{\partial n_3}\frac{\partial n_3}{\partial w_{32}}\left(\frac{\partial E}{\partial a_6}\frac{\partial a_6}{\partial n_6}\frac{\partial n_6}{\partial a_4} + \frac{\partial E}{\partial a_7}\frac{\partial a_7}{\partial n_7}\frac{\partial n_7}{\partial a_4} + \frac{\partial E}{\partial a_8}\frac{\partial a_8}{\partial n_8}\frac{\partial n_8}{\partial a_4}\right)$$

$$+ \frac{\partial a_5}{\partial n_5}\frac{\partial n_5}{\partial a_3}\frac{\partial a_3}{\partial n_3}\frac{\partial n_3}{\partial w_{32}}\left(\frac{\partial E}{\partial a_6}\frac{\partial a_6}{\partial n_6}\frac{\partial n_6}{\partial a_5} + \frac{\partial E}{\partial a_7}\frac{\partial a_7}{\partial n_7}\frac{\partial n_7}{\partial a_5} + \frac{\partial E}{\partial a_8}\frac{\partial a_8}{\partial n_8}\frac{\partial n_8}{\partial a_5}\right)$$

Uczenie parametrów neuronu z warstwy pierwszej - c.d.

Następnie

$$\begin{aligned}\frac{\partial E}{\partial w_{32}} = & -f_4'(n_4)\,w_{43}f_3'(n_3)\,p_2e_6f_6'(n_6)\,w_{64}+\\ & -f_4'(n_4)\,w_{43}f_3'(n_3)\,p_2e_7f_7'(n_7)\,w_{74}+\\ & -f_4'(n_4)\,w_{43}f_3'(n_3)\,p_2e_8f_8'(n_8)\,w_{84}+\\ & -f_5'(n_5)\,w_{53}f_3'(n_3)\,p_2e_6f_6'(n_6)\,w_{65}+\\ & -f_5'(n_5)\,w_{53}f_3'(n_3)\,p_2e_7f_7'(n_7)\,w_{75}+\\ & -f_5'(n_5)\,w_{53}f_3'(n_3)\,p_2e_8f_8'(n_8)\,w_{85}\end{aligned}$$

Definiujemy

$$\begin{aligned}\delta_3 = & f_4'(n_4)\,w_{43}f_3'(n_3)\,e_6f_6'(n_6)\,w_{64} + f_4'(n_4)\,w_{43}f_3'(n_3)\,e_7f_7'(n_7)\,w_{74}\\ & + f_4'(n_4)\,w_{43}f_3'(n_3)\,e_8f_8'(n_8)\,w_{84} + f_5'(n_5)\,w_{53}f_3'(n_3)\,e_6f_6'(n_6)\,w_{65}\\ & + f_5'(n_5)\,w_{53}f_3'(n_3)\,e_7f_7'(n_7)\,w_{75} + f_5'(n_5)\,w_{53}f_3'(n_3)\,e_8f_8'(n_8)\,w_{85}\end{aligned}$$

Uczenie parametrów neuronu z warstwy pierwszej - c.d.

Pozostałe "delty" w warstwie pierwszej

$$\begin{aligned}\delta_1 = {} & e_6 f_6'(n_6) f_4'(n_4) f_3'(n_3) w_{64} w_{41} + e_7 f_7'(n_7) w_{74} f_4'(n_4) f_3'(n_3) w_{41} \\ & + e_8 f_8'(n_8) w_{84} f_4'(n_4) f_3'(n_3) w_{41} + e_6 f_6'(n_6) f_5'(n_5) f_3'(n_3) w_{65} w_{51} \\ & + e_7 f_7'(n_7) w_{75} f_5'(n_5) f_3'(n_3) w_{51} + e_8 f_8'(n_8) w_{85} f_5'(n_5) f_3'(n_3) w_{51}\end{aligned}$$

$$\begin{aligned}\delta_2 = {} & e_6 f_6'(n_6) f_4'(n_4) f_3'(n_3) w_{64} w_{42} + e_7 f_7'(n_7) w_{74} f_4'(n_4) f_3'(n_3) w_{42} \\ & + e_8 f_8'(n_8) w_{84} f_4'(n_4) f_3'(n_3) w_{42} + e_6 f_6'(n_6) f_5'(n_5) f_3'(n_3) w_{65} w_{52} \\ & + e_7 f_7'(n_7) w_{75} f_5'(n_5) f_3'(n_3) w_{52} + e_8 f_8'(n_8) w_{85} f_5'(n_5) f_3'(n_3) w_{52}.\end{aligned}$$

Uczenie parametrów neuronu z warstwy pierwszej - c.d.

Obowiązuje następująca procedura uczenia parametrów warstwy pierwszej

$$\boxed{\left.\begin{array}{l} e_i = t_i - a_i \\ \delta_i = \ \ldots \\ b_i' = b_i + \gamma\delta_i \\ w_{ij}' = w_{ij} + \gamma\delta_i p_j \end{array}\right\} \ \forall\ (i,j) \in \{1,2,3\} \times \{1,2,3\}}$$