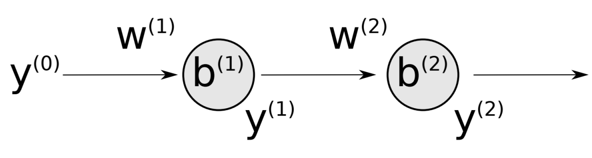
Sprawozdanie - elementy inteligencji obliczeniowej (EIO)	
Prowadzący: *mgr inż. Anna Labijak-Kowalska*	Grupa: ***L9***
Temat Ćwiczeń: ***SSN2*** *– Wsteczna propagacja błędu – Część I*	
Autorzy: *Daniel Zdancewicz [145317]*	

Przy zadaniach przyjęto operacje macierzowe domyślne na tych samych wymiarach jako działania hadamarda

1 Zadanie 1 – Sieć nr 1

Schemat sieci neuronowej:



Zbiór uczący:

$$X_1 = (0, \hat{y} = 0),\ X_2 = (4, \hat{y} = 2)$$

Funkcja aktywacji:
$ReLU = ReLU(x) :=$
$max(0, x) = \begin{cases} x & x > 0, \\ 0 & x \leq 0 \end{cases}$

$\frac{\partial ReLU}{\partial x} := \begin{cases} 1 & x > 0, \\ 0 & x \leq 0 \end{cases}$

Funkcja błędu:
$MSE = MSE(x, \hat{x})$
$:= mean(sum(square(x - \hat{x})))$

Aktualizacja wagi:
$next(\omega_i) := \omega_i - \mu \frac{\partial E}{\partial \omega_i}$

Aktualizacja bias:
$next(b_i) := b_i - \mu \frac{\partial E}{\partial b_i}$

a) Korzystając z reguły łańcuchowej wyprowadź wzory na obliczenie pochodnych cząstkowych

$$\frac{\partial E}{\partial \omega^{(2)}} = \frac{\partial E}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial w^{(2)}} = \frac{\partial E}{\partial y^{(2)}} \cdot \frac{\partial y^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial w^{(2)}} = \left(y^{(2)} - \hat{y}^{(2)}\right) \cdot \begin{cases} 1 & z^{(2)} > 0, \\ 0 & z^{(2)} \leq 0 \end{cases} \cdot y^{(1)} = \delta^{(2)} \cdot y^{(1)}$$

$$\frac{\partial E}{\partial y^{(2)}} = \frac{\partial \frac{1}{2}\left(y^{(2)} - \hat{y}^{(2)}\right)^2}{\partial y^{(2)}} = \frac{1}{2} 2\left(y^{(2)} - \hat{y}^{(2)}\right) = y^{(2)} - \hat{y}^{(2)}$$

$$\frac{\partial y^{(2)}}{\partial z^{(2)}} = \frac{\partial ReLU(z^{(2)})}{\partial z^{(2)}} = \begin{cases} 1 & z^{(2)} > 0, \\ 0 & z^{(2)} \leq 0 \end{cases}$$

$$\frac{\partial z^{(2)}}{\partial w^{(2)}} = \frac{\partial\left(y^{(1)}\omega^{(2)} + b^{(2)}\right)}{\partial w^{(2)}} = y^{(1)}$$

$$\delta^{(1)} = \frac{\partial E}{\partial y^{(2)}} \frac{\partial y^{(2)}}{\partial z^{(2)}}$$

$$\frac{\partial E}{\partial \omega^{(1)}} = \delta^{(2)} \cdot \frac{\partial z^{(2)}}{\partial y^{(1)}} \cdot \frac{\partial y^{(1)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial w^{(1)}} = \delta^{(1)} \cdot y^{(0)}$$

$$\frac{\partial E}{\partial y^{(1)}} = \mathbf{q.v.} = y^{(1)} - \hat{y}^{(1)}$$

$$\frac{\partial y^{(1)}}{\partial z^{(1)}} = \mathbf{q.v.} = \begin{cases} 1 & z^{(1)} > 0, \\ 0 & z^{(1)} \leq 0 \end{cases}$$

$$\frac{\partial z^{(1)}}{\partial w^{(1)}} = \mathbf{q.v.} = y^{(0)}$$

$$\delta^{(1)} = \delta^{(2)}\omega^{(2)} \cdot \frac{\partial y^{(1)}}{\partial z^{(1)}}$$

$$\frac{\partial E}{\partial b^{(2)}} = \frac{\partial E}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial b^{(2)}} = \frac{\partial E}{\partial y^{(2)}} \cdot \frac{\partial y^{(2)}}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial b^{(2)}} = \delta^{(2)}$$

$$\frac{\partial z^{(2)}}{\partial b^{(2)}} = \frac{\partial\left(y^{(1)}\omega^{(2)} + b^{(2)}\right)}{\partial b^{(2)}} = 1$$

$$\frac{\partial E}{\partial b^{(1)}} = \delta^{(2)} \cdot \frac{\partial y^{(2)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial b^{(1)}} = \delta^{(1)}$$

$$\frac{\partial z^{(1)}}{\partial b^{(1)}} = \mathbf{q.v.} = 1$$

b) Oblicz zaktualizowane wagi dla $\mu = 0.1$ oraz następujących wartości wag

* operacje górnej macierzy są niezależne od dolnej, po prostu zapisane 2 aktualizacje by się nie duplikować. Zadanie wykonane przy założeniu, że chodzi o podanie nowych wag niezależnie jako 2 przypadki.

$$\omega^{(1)} = 0.2,\ b^{(1)} = 0.5,\ \omega^{(2)} = 0.5,\ b^{(2)} = -0.5$$

b).1 Aktualizacja 1 – X_1

Aktualizacja poprzednich wag o daną testową X_1

Dane:

$$y^{(0)} = 0,\ \hat{y} = 0$$

Propagacja w przód:

$$z^{(1)} = y^{(0)}\omega^{(1)} + b^{(1)} = 0 \cdot 0.2 + 0.5 = 0.5$$

$y^{(1)} = ReLU\left(y^{(1)}\right) = 0.5$

$z^{(2)} = y^{(1)}\omega^{(2)} + b^{(2)} = 0.5 \cdot 0.5 - 0.5 = -0.25$

$y^{(2)} = ReLU\left(y^{(1)}\right) = 0$

Propagacja wstecz:

$\delta^{(2)} = \left(y^{(2)} - \hat{y}\right) ReLU'\left(z^{(2)}\right) = 0 \cdot 0 = 0$

$\delta^{(1)} = \delta^{(2)}\omega^{(2)} ReLU'\left(z^{(1)}\right) = 0$ Stop. Brak zmiany.

Wagi po aktualizacji takie same, brak aktualizacji.

b).2 Aktualizacja 2 – X_2

Aktualizacja poprzednich wag o daną testową X_2

Dane:

$y^{(0)} = 4,\ \hat{y} = 2$

Propagacja w przód:

$z^{(1)} = y^{(0)}\omega^{(1)} + b^{(1)} = 4 \cdot 0.2 + 0.5 = 1.3$

$y^{(1)} = ReLU\left(y^{(1)}\right) = 1.3$

$z^{(2)} = y^{(1)}\omega^{(2)} + b^{(2)} = 1.3 \cdot 0.5 - 0.5 = -0.15$

$y^{(2)} = ReLU\left(y^{(1)}\right) = 0.15$

Propagacja wstecz:

$\delta^{(2)} = \left(y^{(2)} - \hat{y}\right) ReLU'\left(z^{(2)}\right) = (0.15 - 2)1 = -1.85$

$next\left(\omega^{(2)}\right) = \omega^{(2)} - \mu\delta^{(2)}y^{(1)} = 0.5 - 0.1 \cdot -1.85 \cdot 1.3 = 0.7405$

$next\left(b^{(2)}\right) = \_(2) - \mu\delta^{(2)} = -0.5 - 0.1 \cdot -1.85 = 0.685$

$\delta^{(1)} = \delta^{(2)}\omega^{(2)} ReLU'\left(z^{(1)}\right) = -(1.85 \cdot 0.5)1 = -0.925$

$next\left(\omega^{(1)}\right) = \omega^{(1)} - \mu\delta^{(1)}y^{(0)} = 0.2 - 0.1 \cdot -0.925 \cdot 0.15 = 0.213875$

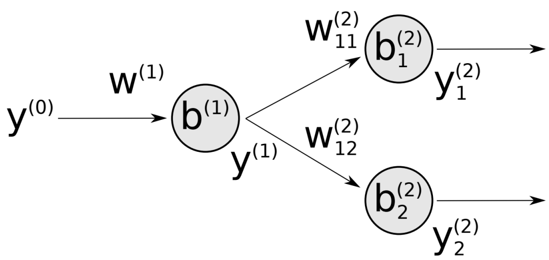
$next\left(b^{(1)}\right) = b^{(1)} - \mu\delta^{(1)} = 0.5 - 0.1 \cdot -0.925 = 0.5925$

Wagi po aktualizacji:

$\omega^{(2)} = 0.7405$
$b^{(2)} = 0.685$
$\omega^{(1)} = 0.213875$
$b^{(1)} = 0.5925$

2 Zadanie 2 – Sieć nr 2

Schemat sieci neuronowej:



Zbiór uczący:

$X_1 = (0, \hat{y} = (0, 1))$

$X_2 = (4, \hat{y} = (2, 5))$

Funkcja aktywacji:

$$ReLU = ReLU(x) := max(0, x) = \begin{cases} x & x > 0, \\ 0 & x \leq 0 \end{cases}$$

$$\frac{\partial ReLU}{\partial x} := \begin{cases} 1 & x > 0, \\ 0 & x \leq 0 \end{cases}$$

Funkcja błędu:
$MSE = MSE(x, \hat{x})$
$:= mean(sum(square(x - \hat{x})))$

Aktualizacja wagi:
$next(\omega_i) := \omega_i - \mu \frac{\partial E}{\partial \omega_i}$

Aktualizacja bias:
$next(b_i) := b_i - \mu \frac{\partial E}{\partial b_i}$

a) Korzystając z reguły łańcuchowej wyprowadź wzory na obliczenie pochodnych cząstkowych

$$\frac{\partial E}{\partial \omega^{(2)}_{11}} = \frac{\partial E}{\partial y^{(2)}_1} \cdot \frac{\partial y^{(2)}_1}{\partial z^{(2)}_1} \cdot \frac{\partial z^{(2)}_1}{\partial \omega^{(2)}_{11}} = \mathbf{q}.\mathbf{v}. = \delta^{(2)}_1 y^{(1)}$$

$$\frac{\partial E}{\partial y^{(2)}_1} = \frac{\partial \frac{1}{2}(y^{(2)}_1 - \hat{y})^2}{\partial y^{(2)}_1} = \mathbf{q}.\mathbf{v}. = y^{(2)}_1 - \hat{y}$$

$$\frac{\partial y^{(2)}_1}{\partial z^{(2)}_1} = \frac{\partial ReLU\left(z^{(2)}_1\right)}{\partial z^{(2)}_1} = \begin{cases} 1 & z^{(2)}_1 > 0, \\ 0 & z^{(2)}_1 \leq 0 \end{cases}$$

$$\frac{\partial z^{(2)}_1}{\partial w^{(2)}_{11}} = \frac{\partial\left(y^{(1)} w^{(2)}_{11} + b^{(2)}_1\right)}{\partial w^{(2)}_{11}} = y^{(1)}$$

$$\delta^{(2)}_1 = \frac{\partial E}{\partial y^{(2)}_1} \cdot \frac{\partial y^{(2)}_1}{\partial z^{(2)}_1}$$

$$\frac{\partial E}{\partial \omega^{(2)}_{12}} = \frac{\partial E}{\partial y^{(2)}_2} \cdot \frac{\partial y^{(2)}_2}{\partial z^{(2)}_2} \cdot \frac{\partial z^{(2)}_2}{\partial \omega^{(2)}_{12}} = \mathbf{q}.\mathbf{v}. = \delta^{(2)}_2 y^{(1)}$$

$$\frac{\partial E}{\partial y^{(2)}_2} = \mathbf{q}.\mathbf{v}. = y^{(2)}_2 - \hat{y}$$

$$\frac{\partial y_2^{(2)}}{\partial z_2^{(2)}} = \mathbf{q}.\mathbf{v}. = \begin{cases} 1 & z_2^{(2)} > 0, \\ 0 & z_2^{(2)} \leq 0 \end{cases}$$

$$\frac{\partial z_2^{(2)}}{\partial w_{12}^{(2)}} = \mathbf{q}.\mathbf{v}. = y^{(1)}$$

$$\delta_2^{(2)} = \frac{\partial E}{\partial y_2^{(2)}} \cdot \frac{\partial y_2^{(2)}}{\partial z_2^{(2)}}$$

$$\frac{\partial E}{\partial \omega^{(1)}} = \frac{\partial E}{\partial y_1^{(2)}} \cdot \frac{\partial y_1^{(2)}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial y^{(1)}} \cdot \frac{\partial y^{(1)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial w^{(1)}} + \frac{\partial E}{\partial y_2^{(2)}} \cdot \frac{\partial y_2^{(2)}}{\partial z_2^{(2)}} \cdot \frac{\partial z_2^{(2)}}{\partial y^{(1)}} \cdot \frac{\partial y^{(1)}}{\partial z^{(1)}} \cdot \frac{\partial z^{(1)}}{\partial w^{(1)}} =$$
$$\left(\delta_1^{(2)}\omega_1^{(2)} + \delta_2^{(2)}\omega_2^{(2)}\right) \frac{\partial y^{(1)}}{\partial z^{(1)}} y^{(0)} = \delta^{(1)} y^{(0)}$$

$$\frac{\partial y^{(1)}}{\partial z^{(1)}} = \begin{cases} 1 & z^{(1)} > 0, \\ 0 & z^{(1)} \leq 0 \end{cases}$$

$$\frac{\partial z^{(1)}}{\partial w^{(1)}} = \mathbf{q}.\mathbf{v}. = y^{(0)}$$

$$\delta^{(1)} = \left(\delta_1^{(2)}\omega_1^{(2)} + \delta_2^{(2)}\omega_2^{(2)}\right) \frac{\partial y^{(1)}}{\partial z^{(1)}}$$

$$\frac{\partial E}{\partial b^{(1)}} = \mathbf{q}.\mathbf{v}. = \delta^{(1)}$$

$$\frac{\partial z^{(1)}}{\partial b^{(1)}} = \mathbf{q}.\mathbf{v}. = 1$$

Wyznaczane tak samo tylko ostatnim elementem jest $\frac{\partial z^{(1)}}{\partial b^{(1)}}$ zamiast $\frac{\partial z^{(1)}}{\partial w^{(1)}}$

b) Oblicz zaktualizowane wagi dla $\mu = 0.1$ oraz następujących wartości wag

* operacje górnej macierzy są niezależne od dolnej, po prostu zapisane 2 aktualizacje by się nie duplikować. Zadanie wykonane przy założeniu, że chodzi o podanie nowych wag niezależnie jako 2 przypadki.

$$\omega^{(1)} = 0.2,\ b^{(1)} = 0.5,\ \omega_{11}^{(2)} = 0.5,\ b_1^{(2)} = -0.5,\ \omega_{12}^{(2)} = 0.0,\ b_2^{(2)} = 0.5$$

b).1 Aktualizacja 1 – X_1

Aktualizacja poprzednich wag o daną testową X_1

Dane:

$$y^{(0)} = 0,\ \hat{y} = (0, 1)$$

Propagacja w przód:

- $z^{(1)} = y^{(0)}\omega^{(1)} + b^{(1)} = 0 \cdot 0.2 + 0.5 = 0.5$
- $y^{(1)} = ReLU\left(z^{(1)}\right) = 0.5$
- $z_1^{(2)} = y^{(1)}\omega_{11}^{(2)} + b_1^{(2)} = 0.5 \cdot 0.5 - 0.5 = -0.25$

- $y_1^{(2)} = ReLU\left(z_1^{(2)}\right) = 0$
- $z_2^{(2)} = y^{(1)}\omega_{12}^{(2)} + b_2^{(2)} = 0.5 \cdot 0.0 + 0.5 = 0.5$
- $y_2^{(2)} = ReLU\left(z_2^{(2)}\right) = 0.5$

Propagacja wstecz:

- $\delta_1^{(2)} = \left(y_1^{(2)} - \hat{y}\right) ReLU'\left(z_1^{(2)}\right) = 0 - 0 = 0$
- $next\left(\omega_{11}^{(2)}\right) = \omega_{11}^{(2)} - \mu\delta_1^{(2)}y^{(1)} = 0.5 - 0.1 \cdot 0 = 0.5$
- $next\left(b_1^{(2)}\right) = b_1^{(2)} - \mu\delta_1^{(2)} = -0.5 \cdot 0.1 \cdot 0.0 = 0$
- $\delta_2^{(2)} = \left(y_2^{(2)} - \hat{y}\right) ReLU'\left(z_2^{(2)}\right) = (0.5 - 1)1 = -0.5$
- $next\left(\omega_{12}^{(2)}\right) = \omega_{12}^{(2)} - \mu\delta_2^{(2)}y^{(1)} = 0 + 0.5 \cdot 0.1 \cdot 0.5 = 0.025$
- $next\left(b_2^{(2)}\right) = b_2^{(2)} - \mu\delta_2^{(2)} = 0.5 + 0.5 \cdot 0.1 = 0.55$
- $\delta^{(1)} = \left(\delta_1^{(2)}\omega_{11}^{(2)} + \delta_2^{(2)}\omega_{12}^{(2)}\right) ReLU'\left(z^{(1)}\right) = 0 + -0.5 \cdot 0 = 0$
- $next\left(\omega^{(1)}\right) = \omega^{(1)} - \mu\delta^{(1)}y^{(0)} = \omega^{(1)} - 0 = 0.2$
- $next\left(b^{(1)}\right) = b^{(1)} - \mu\delta^{(1)} = b^{(1)} - 0 = 0.5$

Wagi po aktualizacji:

- $\omega^{(1)} = 0.2$ (brak zmiany)
- $b^{(1)} = 0.5$ (brak zmiany)
- $\omega_{11}^{(2)} = 0.5$ (brak zmiany)
- $b_1^{(2)} = -0.5$ (brak zmiany)
- $\omega_{12}^{(2)} = 0.025$
- $b_2^{(2)} = 0.55$

b).2 Aktualizacja 2 – X_2

Aktualizacja poprzednich wag o daną testową X_2

Dane:

$$y^{(0)} = 4,\ \hat{y} = (2, 5)$$

Propagacja w przód:

- $z^{(1)} = 0.2 \cdot 4 + 0.5 = 1.3$
- $y^{(1)} = ReLU\left(z^{(1)}\right) = 1.3$

- $z_1^{(2)} = y^{(1)}\omega_{11}^{(2)} + b_1^{(2)} = 1.3 \cdot 0.5 - 0.5 = 0.15$
- $y_1^{(2)} = ReLU\left(z_1^{(2)}\right) = 0.15$
- $z_2^{(2)} = y^{(1)}\omega_{12}^{(2)} + b_2^{(2)} = 1.3 \cdot 0.025 + 0.55 = 0.5825$
- $y_2^{(2)} = ReLU\left(z_2^{(2)}\right) = 0.5825$

Propagacja wstecz:

- $\delta_1^{(2)} = \left(y_1^{(2)} - \hat{y}\right) ReLU'\left(z_1^{(2)}\right) = (0.15 - 2)\, 1 = -1.85$
- $next\left(\omega_{11}^{(2)}\right) = \omega_{11}^{(2)} - \mu\delta_1^{(2)}y^{(1)} = 0.5 - 0.1 \cdot -1.85 \cdot 1.3 = 0.7405$
- $next\left(b_1^{(2)}\right) = b_1^{(2)} - \mu\delta_1^{(2)} = -0.5 - 0.1 \cdot -1.85 = -0.315$
- $\delta_2^{(2)} = \left(y_2^{(2)} - \hat{y}\right) ReLU'\left(z_2^{(2)}\right) = (0.5825 - 5)\, 1 = -4.4175$
- $next\left(\omega_{12}^{(2)}\right) = \omega_{12}^{(2)} - \mu\delta_2^{(2)}y^{(1)} = 0.025 - 0.1 \cdot -4.4175 \cdot 1.3 = 0.599275$
- $next\left(b_2^{(2)}\right) = b_2^{(2)} - \mu\delta_1^{(2)} = 0.55 - 0.1 \cdot -4.4175 = -0.99175$
- $\delta^{(1)} = \left(\delta_1^{(2)}\omega_{11}^{(2)} + \delta_2^{(2)}\omega_{12}^{(2)}\right) ReLU'\left(z^{(1)}\right) = (-1.85 \cdot -0.5 - 4.4175 \cdot 0.025)\, 1 = 0.8145625$
- $next\left(\omega^{(1)}\right) = \omega_{12}^{(2)} - \mu\delta^{(1)}y^{(0)} = 0.2 - 0.1 \cdot 0.8145625 \cdot 4 = -0.125825$
- $next\left(b^{(1)}\right) = b_2^{(2)} - \mu\delta^{(1)} = 0.55 - 0.1 * 0.8145625 = 0.46854375$

Wagi po aktualizacji:

- $\omega^{(1)} = -0.125825$
- $b^{(1)} = 0.46854375$
- $\omega_{11}^{(2)} = 0.7405$
- $b_1^{(2)} = -0.315$
- $\omega_{12}^{(2)} = 0.599275$
- $b_2^{(2)} = -0.99175$

3 Wnioski

Na podstawie ćwiczeń wytworzył się schemat propagacji wstecznej polegający na początkowym przejściu "do przodu"wzdłuż sieci, zdobywając informacje o kolejnych wartościach aktualizacji i następnie w drugim kroku przejściu "wstecz"wzdłuż sieci kolejno aktualizując wagi oraz bias'y sieci. Jako osoba z dyskalkulia siedziałem nad tym zdecydowanie zbyt długo (8 godzin), a i tak pewnie popełniłem błędy. Dzieci lubią sieci.