NLP

Wydział Biochemii, Biofizyki i Biotechnologii
Adrian Kania

Działania na macierzach

Dodawanie macierzy

Dodawać można tylko macierze o tych samych wymiarach i robi to się następująco:

$$\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{bmatrix} + \begin{bmatrix}0 & 3 & 1\\ -3 & 7 & 8\end{bmatrix} = \begin{bmatrix}1+0 & 2+3 & 3+1\\ 4-3 & 5+7 & 6+8\end{bmatrix} = \begin{bmatrix}1 & 5 & 4\\ 1 & 12 & 14\end{bmatrix}$$

Mnożenie macierzy przez liczbę

Jeśli mnożymy macierz przez liczbę, to każdy jej wyraz mnożymy przez tę liczbę:

$$2 \cdot \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = \begin{bmatrix}2 & 4\\ 6 & 8\end{bmatrix}$$

Mnożenie dwóch macierzy

Mnożenie macierzy jest trochę bardziej skomplikowane. Można je wykonać tylko wtedy gdy pierwsza macierz ma tyle samo kolumn ile druga wierszy (w szczególności więc widać, że mnożenie macierzy nie jest przemienne, bo iloczyn w odwrotnej kolejności może w ogóle nie istnieć). Obliczając iloczyn dwóch macierzy mnożymy skalarnie każdy wiersz pierwszej macierzy przez każdą kolumnę drugiej macierzy:

$$\begin{array}{cc} & \begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{bmatrix} \\ \begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{bmatrix} & \begin{bmatrix} & \downarrow & \\ \rightarrow & \bullet & \end{bmatrix}\end{array}$$

W miejscu kropki znajduje się iloczyn wiersza przy strzałce poziomej i kolumny przy strzałce pionowej, czyli:

$4 \cdot 2 + 5 \cdot 5 + 6 \cdot 8 = 81$

Ile wynosi pochodna $\frac{dy}{dx}$, gdzie $y = x^4$?

Ile wynosi pochodna $\frac{dy}{dx}$, gdzie $y = x^4$?

$$\frac{dy}{dx} = 4x^3$$

Ile wynosi pochodna $\frac{dy}{dx}$, gdzie $y = 5x^4$?

$$\frac{dy}{dx} = 5 \cdot 4x^3 = 20x^3$$

Ile wynosi pochodna $\frac{dy}{dx}$, gdzie $y = 5x^4$?

$$\frac{dy}{dx} = 5 \cdot 4x^3 = 20x^3$$

Twierdzenie: Pochodna sumy funkcji jest równa sumie pochodnych tych funkcji.
Przykład: $y = 6x^5 + 8x^3 + 7x + 1$.

Ile wynosi pochodna $\frac{dy}{dx}$, gdzie $y = 5x^4$?

$$\frac{dy}{dx} = 5 \cdot 4x^3 = 20x^3$$

Twierdzenie: Pochodna sumy funkcji jest równa sumie pochodnych tych funkcji.

Przykład: $y = 6x^5 + 8x^3 + 7x + 1$. Wtedy $\frac{dy}{dx} = 30x^4 + 24x^2 + 7 + 0$.

Pochodne cząstkowe. Dana jest funkcja $f(x,y) = yx^4$. Ile wynosi $\frac{\partial f}{\partial x}$?

Pochodne cząstkowe. Dana jest funkcja $f(x, y) = yx^4$. Ile wynosi $\frac{\partial f}{\partial x}$?

$$\frac{\partial f}{\partial x} = y \cdot 4x^3 = 4yx^3.$$

Kiedy licze pochodną po zmiennej x traktuje y jak stałą.

Pochodne cząstkowe. Dana jest funkcja $f(x, y) = y^2x^4 + 3xy$. Wtedy:

- $\frac{\partial f}{\partial x} = 4y^2x^3 + 3y$
- $\frac{\partial f}{\partial y} = 2yx^4 + 3x$

Pochodne cząstkowe. Dana jest funkcja $f(x, y) = y^2x^4 + 3xy$. Wtedy:

- $\frac{\partial f}{\partial x} = 4y^2x^3 + 3y$
- $\frac{\partial f}{\partial y} = 2yx^4 + 3x$

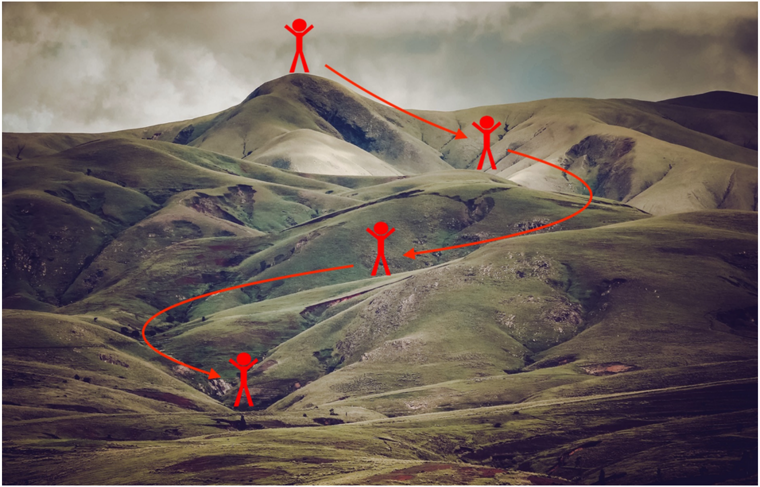
Zadanie: Wyznacz $\frac{\partial f}{\partial x}$ oraz $\frac{\partial f}{\partial y}$ gdzie:

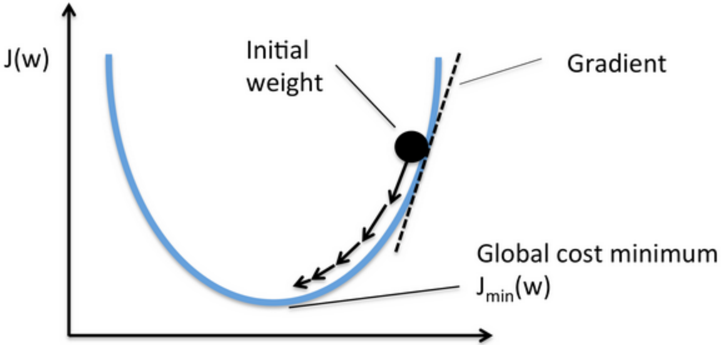
- $f(x, y) = y^3x^5 + 7yx^2 + 12x$.

Zadanie: Wyznacz $\frac{\partial f}{\partial x_1}$, gdzie:

- $f(x_1, x_2, x_3) = 2x_1^5x_2^7 - 4x_1^5x_3^{10}$

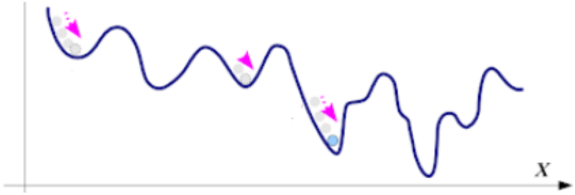
Poszukiwanie minimum





- Jeżeli szukamy minimum funkcji $J(w)$, to wybieramy jakiś punkt początkowy w_0 a następnie poruszamy się w kierunku $-\frac{\partial J}{\partial w}$.
- W kolejnym kroku aktualizujemy w_0 wg przepisu: $w_0 = w_0 - \alpha \frac{\partial J}{\partial w}$, gdzie α - pewien współczynnik.

Funkcja na ogół ma wiele minimów lokalnych ... dlatego możemy nie dotrzeć do minimum globalnego.



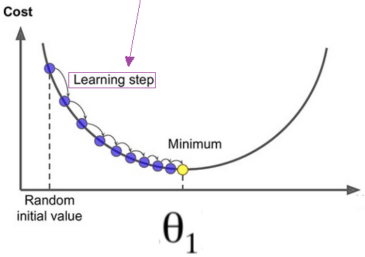
Warto wystartować z różnych punktów.

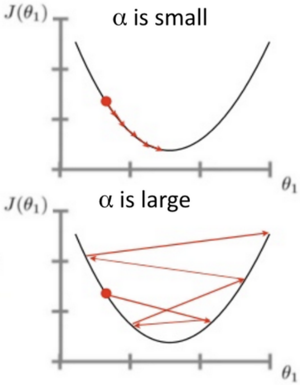
Współczynnik uczenia (α)

repeat until convergence {

$$\theta_j := \theta_j - \boxed{\alpha}\frac{\partial}{\partial\theta_j}J(\theta_0,\theta_1)$$

}





Regresja liniowa - wyznaczanie współczynników modelu

$$y = ax + b$$

- definiujemy parametry początkowe a i b
- definiujemy funkcje forward (postać modelu)
- definiujemy funkcje kosztu **loss** (średni błąd pomiędzy przewidywaną a oczekiwaną wartością)
$$L(a, b) = \frac{1}{n} \sum_i (ax_i + b - y_i)^2$$
- definiujemy gradient (pochodną z funkcji kosztu po a i b)
$$\frac{\partial L}{\partial a} = \frac{1}{n} \sum_i 2x_i(ax_i + b - y_i)$$
$$\frac{\partial L}{\partial b} = \frac{1}{n} \sum_i 2(ax_i + b - y_i)$$
- definiujemy współczynnik uczenia (α)
- definujemy liczbę kroków (ile razy będziemy powtarzać aktualizację nowych parametrów a i b).