Inteligenţă artificială - laborator

Metoda gradientului

Gradientul într-un punct al unei funcţii reprezintă direcţia creşterii maxime a acelei funcţii. Gradientul reprezintă tangenta la curbă pe direcţia de creştere şi este nul în punctele de maxim şi minim. Având funcţia $E(\mathbf{x})$, unde $\mathbf{x}$ este un vector cu n componente, gradientul se defineşte prin:

$$\nabla E_{\mathrm{x}} = \begin{bmatrix} \frac{\partial E}{\partial x_1} \\ \vdots \\ \frac{\partial E}{\partial x_n} \end{bmatrix}$$

Metoda gradientului este un algoritm iterativ de optimizare care constă în ***găsirea maximului sau a minimului local al unei funcţii*** folosind gradientul, respectiv gradientul său negat.

Minimul unei funcţii se poate găsi folosind metoda gradientului astfel: $\mathbf{x}_{n+1} = \mathbf{x}_n - c\nabla f(\mathbf{x}_n)$, unde c este o constantă pozitivă, suficient de mică astfel încât $f(\mathbf{x}_0) \geq f(\mathbf{x}_1) \geq \ldots \geq f(\mathbf{x}_n) \geq f(\mathbf{x}_{n+1})$. Algoritmul foloseşte o iniţializare aleatoare a lui $\mathbf{x}_0$. Constanta de învăţare c determină o convergenţă a valorilor $\mathbf{x}_n$ către optim mai rapidă şi mai brută daca este mare sau mai lentă şi mai fină dacă este mică. Se poate alege, de exemplu, c=0,01, dar se poate opta şi pentru modificarea acestei valori chiar în timpul rulării programului. Algoritmul continuă până când modificările componentelor lui $\mathbf{x}_n$ scad sub un prag acceptabil, de exemplu 0,00001.

Pentru funcţii $f(x)$ care depind de o singură variabilă, relaţia de mai sus se rescrie astfel:

$$x_{n+1} = x_n - cf'(x_n).$$

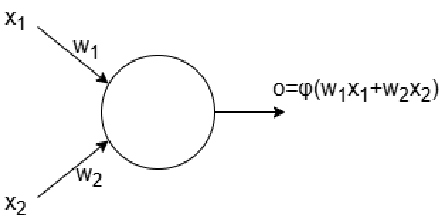
Pentru funcţii $f(x,y)$ care depind de două variabile, relaţia se rescrie prin:

$$\begin{cases} x_{n+1} = x_n - c\dfrac{\partial f(x_n)}{\partial x} \\ y_{n+1} = y_n - c\dfrac{\partial f(y_n)}{\partial y} \end{cases}$$

Exercițiul 1. Scrieţi un program care găseşte minimul funcţiei $f(x) = 3x^2-6x+2$ prin metoda gradientului. Folosiţi diverse valori pentru c. Afişaţi valoarea lui $f(x)$ după fiecare iteraţie.

Exercițiul 2. Instruirea unui neuron constă în ajustarea treptată a ponderilor pentru a minimiza eroarea dintre ieșirea produsă și cea așteptată, pe baza datelor de instruire. Scopul este obținerea unor ponderi optime care reduc această eroare la un nivel minim.

Neuronul de mai jos are intrările x_1și x_2, ponderile w_1și w_2 și ieșirea o. Pentru acest exercițiu, considerăm că funcția de activare este liniară, $\varphi(z) = z$.



Determinați ponderile acestui neuron minimizând eroarea pătratică medie $E(w_1, w_2) = \frac{1}{2}(y - o)^2$ prin metoda gradientului, astfel încât ieșirea să fie $y = 5$ pentru intrările $x_1 = 1$ și $x_2 = 2$.