Napredni algoritmi i strukture podataka

1. jesenski rok

a.g. 2019/2020

Zadatak 2 (+10)

OZNAKE: **neuronska mreža** **backpropagation** **adaline** **tanh**

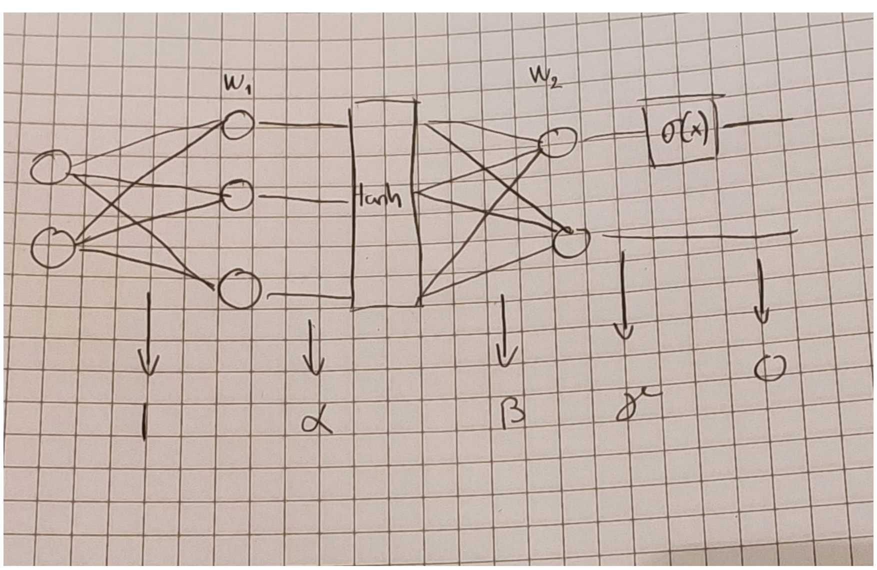
Potpuno povezana, unaprijedna (feedforward) troslojna neuronska mreža (ANN; *Artificial Neural Network*) ima strukturu $2 \times 3 \times 2$, pri čemu je sloju tangens hiperbolni (tanh), dok je u izlaznom sloju aktivacijska funkcija za izlaz 1 sigmoid, a za izlaz 2 je Adaline. Provedite prvi korak uvježbavanja te mreže (jednom osvježiti sve parametare) algoritmom koračnog uvježbavanja (*on-line learning*) ako se podatci za uvježbavanje uzimaju redom iz sljedeće tablice.

ulaz 1	ulaz 2	izlaz 1	izlaz 2
-1	3	0.4	-2
-1	6	0.2	-6
-9	4	-0.4	8
5	-3	0.4	-9

Početne vrijednosti svih parametara mreže postavite na nula, a zatrebaju li Vam još neke veličine, pridijelite im vrijednosti po vlastitom nahođenju, samo jasno navedite svoj izbor i kratko objasnite ulogu te veličine.

Napomena: $tanh(x) = 2\sigma(2x) - 1$

Prvo skicirajmo mrežu:



Pojedine tokove u mreži označili smo s I (input), α, β, γ i O (output). Znamo da nam trebaju gradijenti za svaki tok izuzev ulaza, kao i gradijenti za svaki skup parametara. Osim označenih W_1 i W_2 imamo i b_1 i b_2. Umjesto da pišemo parcijalne derivacije, pisat ćemo gradijent prefiksiran s nablom. Na primjer, umjesto $\frac{\partial \mathcal{L}}{\partial O}$ pišemo ∇O.

Dakle, potrebni su nam sljedeći gradijenti:

- ∇O
- $\nabla \gamma$
- ∇W_2, ∇b_2
- $\nabla \beta$
- $\nabla \alpha$
- ∇W_1, ∇b_1

Podrazumijevana funkcija gubitka $\mathcal{L}$ je MSE:

$$\mathcal{L}(\text{target}, \text{prediction}) = \frac{1}{2} (\text{target} - \text{prediction})^2 \tag{1}$$

Prvi gradijent kojeg možemo izračunati je ∇O. On je derivacija gubitka po izlazu (predictionu) je

$$\nabla O = prediction - target \tag{2}$$

Ako ovo pretvorimo u matrični oblik, dobivamo:

$$\nabla O = \begin{bmatrix} O_0 - Y_0 \\ O_1 - Y_1 \end{bmatrix} \tag{3}$$

nazovemo li prediction vektor $\vec{O}$, a target vektor $\vec{Y}$.

S obzirom na to da je $\nabla \gamma$ kompozitna funkcija, morat ćemo $\frac{\partial o}{\partial \gamma}$ raspisati matrično:

$$\frac{\partial O}{\partial \gamma} = \begin{bmatrix} \sigma(\gamma_0)\,(1 - \sigma(\gamma_0)) \\ 1 \end{bmatrix} = \begin{bmatrix} O_0\,(1 - O_0) \\ 1 \end{bmatrix} \tag{4}$$

Stoga uz ulančano pravilo vrijedi

$$\nabla \gamma = \begin{bmatrix} \nabla O_0 \cdot O_0\,(1 - O_0) \\ \nabla O_1 \end{bmatrix} \tag{5}$$

Sada nas zanimaju gradijenti parametara (∇W_2 i ∇b_2). Prvo trebamo $\frac{\partial \gamma}{\partial W_2}$ i $\frac{\partial \gamma}{\partial b_2}$, a oni su istog oblika kao i W_2 i b_2:

$$\frac{\partial \gamma}{\partial W_2} = \begin{bmatrix} \beta_0 & \beta_0 \\ \beta_1 & \beta_1 \\ \beta_2 & \beta_2 \end{bmatrix} \tag{6}$$

$$\frac{\partial \gamma}{\partial b_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{7}$$

pa kad primijenimo ulančano pravilo dobivamo

$$\nabla W_2 = \begin{bmatrix} \beta_0 \nabla \gamma_0 & \beta_0 \nabla \gamma_1 \\ \beta_1 \nabla \gamma_0 & \beta_1 \nabla \gamma_1 \\ \beta_2 \nabla \gamma_0 & \beta_2 \nabla \gamma_1 \end{bmatrix} \tag{8}$$

$$\nabla b_2 = \begin{bmatrix} \nabla \gamma_0 \\ \nabla \gamma_1 \end{bmatrix} \tag{9}$$

Zatim tražimo $\nabla\beta$, a to je težinska suma težina po retcima sa zadnjim gradijentom toka (u našem slučaju s γ):

$$\nabla \beta = \begin{bmatrix} W_2^{0,0} \nabla \gamma_0 + W_2^{0,1} \nabla \gamma_1 \\ W_2^{1,0} \nabla \gamma_0 + W_2^{1,1} \nabla \gamma_1 \\ W_2^{2,0} \nabla \gamma_0 + W_2^{2,1} \nabla \gamma_1 \end{bmatrix} \tag{10}$$

Slično kao i prije, trebamo pomnožiti naš gradijent elementwise s gradijentom aktivacijske funkcije. Uz zadatak smo dobili hint:

$$tanh(x) = 2\sigma(2x) - 1 \tag{11}$$

pa stoga možemo reći

$$\frac{\partial tanh(x)}{\partial x} = 4\sigma(x) \cdot^* (1 - \sigma(x)) \tag{12}$$

intuitivno, kada gledate $tanh(x)$, to je sigmoida koja je samo duplo izdužena u visinu. Duplo povećanje u visinu će kvadratno povećati gradijent, a $2^2 = 4$.

Sada možemo dobiti i $\nabla\alpha$. Uzevši u obzir da vrijedi

$$\frac{\partial \beta}{\partial \alpha} = 4\sigma(\alpha) \cdot^* (1 - \sigma(\alpha)) \tag{13}$$

uz ulančano pravilo možemo pisati

$$\nabla \alpha = \begin{bmatrix} 4\sigma(\alpha_0)\,(1 - \sigma(\alpha_0))\,\nabla \beta_0 \\ 4\sigma(\alpha_1)\,(1 - \sigma(\alpha_1))\,\nabla \beta_1 \\ 4\sigma(\alpha_2)\,(1 - \sigma(\alpha_2))\,\nabla \beta_2 \end{bmatrix} \tag{14}$$

Finalno, ponovimo sve slično kao u jednadžbama (6) i (7):

$$\frac{\partial \gamma}{\partial W_1} = \begin{bmatrix} I_0 & I_0 & I_0 \\ I_1 & I_1 & I_1 \end{bmatrix} \tag{15}$$

$$\frac{\partial \gamma}{\partial b_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{16}$$

pa uz ulančano pravilo dobivamo

$$\nabla W_1 = \begin{bmatrix} I_0 \nabla \alpha_0 & I_0 \nabla \alpha_1 & I_0 \nabla \alpha_2 \\ I_1 \nabla \alpha_0 & I_1 \nabla \alpha_1 & I_1 \nabla \alpha_2 \end{bmatrix} \tag{17}$$

$$\nabla b_1 = \begin{bmatrix} \nabla \alpha_0 \\ \nabla \alpha_1 \\ \nabla \alpha_2 \end{bmatrix} \tag{18}$$

Time smo izračunali sve što nam treba pa možemo krenuti na prvi korak učenja.

Forward pass

Sve težine su na 0, tj. vrijedi:

$$W_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{19}$$

$$b_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{20}$$

$$W_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{21}$$

$$b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{22}$$

Kada propustimo prvi primjerak kroz mrežu, dobivamo sljedeće tokove:

$$I = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \tag{23}$$

$$\alpha = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{24}$$

$$\beta = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{25}$$

$$\gamma = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{26}$$

$$O = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} \tag{27}$$

Backward pass

Sukladno izračunatim gradijentima, pišemo:

$$\nabla O = \begin{bmatrix} 0.5 - 0.4 \\ 0 - (-2) \end{bmatrix} = \begin{bmatrix} 0.1 \\ 2 \end{bmatrix} \tag{28}$$

$$\nabla \gamma = \begin{bmatrix} 0.1 \cdot 0.5 \cdot 0.5 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.025 \\ 2 \end{bmatrix} \tag{29}$$

$$\nabla W_2 = \begin{bmatrix} 0 \cdot 0.025 & 0 \cdot 2 \\ 0 \cdot 0.025 & 0 \cdot 2 \\ 0 \cdot 0.025 & 0 \cdot 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{30}$$

$$\nabla b_2 = \begin{bmatrix} 0.025 \\ 2 \end{bmatrix} \tag{31}$$

$$\nabla \beta = \begin{bmatrix} 0 \cdot 0.1 + 0 \cdot 2 \\ 0 \cdot 0.1 + 0 \cdot 2 \\ 0 \cdot 0.1 + 0 \cdot 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{32}$$

$$\nabla \alpha = \begin{bmatrix} 4 \cdot 0.5 \cdot 0.5 \cdot 0 \\ 4 \cdot 0.5 \cdot 0.5 \cdot 0 \\ 4 \cdot 0.5 \cdot 0.5 \cdot 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{33}$$

$$\nabla W_1 = \begin{bmatrix} -1 \cdot 0 & -1 \cdot 0 & -1 \cdot 0 \\ 3 \cdot 0 & 3 \cdot 0 & 3 \cdot 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{34}$$

$$\nabla b_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{35}$$

Ažuriranje parametara

Pravilom

$$\theta^k = \theta^{k-1} - \eta \nabla \theta^{k-1} \tag{36}$$

ažuriramo težine uz stopu učenja 1 (tj. uz $\eta = 1$):

$$W_1' = \begin{bmatrix} 0-0 & 0-0 & 0-0 \\ 0-0 & 0-0 & 0-0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{37}$$

$$b_1' = \begin{bmatrix} 0-0 \\ 0-0 \\ 0-0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{38}$$

$$W_2' = \begin{bmatrix} 0-0 & 0-0 \\ 0-0 & 0-0 \\ 0-0 & 0-0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{39}$$

$$b_2' = \begin{bmatrix} 0 - 0.025 \\ 0 - 2 \end{bmatrix} = \begin{bmatrix} -0.025 \\ -2 \end{bmatrix} \tag{40}$$