VIR

Daniel Kubišta

8. listopadu 2021

1 Zadání

Máme k dispozici data (x_i, y_i) pro $i = 1, 2, 3$. Chceme najít afinní funkci, která aproximuje y. Konkrétně najít parametry w_0, w_1 funkce $\widetilde{y}(x) = f(x) = w_0 + w_1 x$, která minimalizuje výraz $\sum_{i=1}^{3} (y_i - \widetilde{y}(x_i))^2$. Za ztrátovou funkci volíme L2. Data, která máme k dispozici, jsou následující

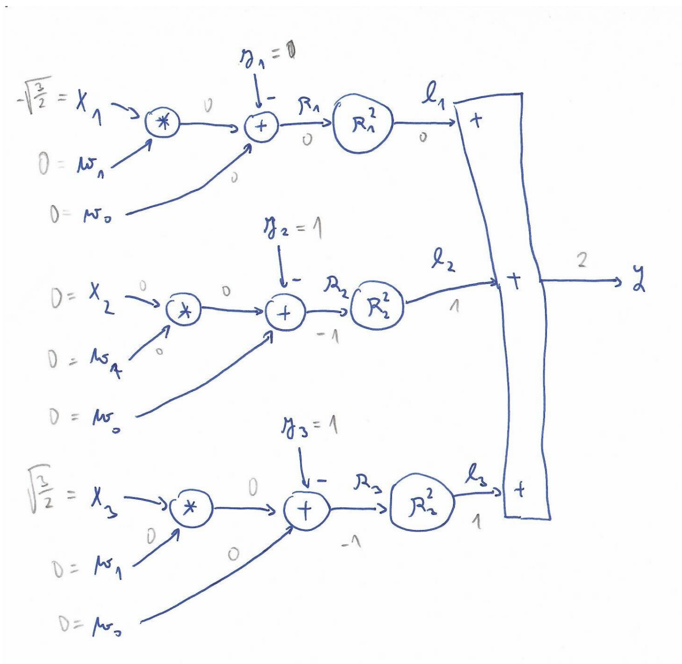
i	x_i	y_i
1	$-\sqrt{3/2}$	0
2	0	1
3	$\sqrt{3/2}$	1

1.1 Výpočetní graf a forward pass

Nakreslete výpočetní graf problému a proveďte forward pass s hodnotami vah $w_0 = w_1 = 0$.

Řešení

Výpočetní graf



Obrázek 1: Výpočetní graf.

Forward pass získáme dosazením, pro ztrátovou funkci získáme:

$$\mathcal{L} = \sum_{i=1}^{3} (w_0 + w_1 x_i - y_i)^2 = 0 + 1 + 1 = 2 \tag{1}$$

1.2 Gradient a update vah

Spočítejte gradient ztrátové funkce a proveďte jeden update vah s learning rate $\alpha = 0.1$.

Řešení

Nejdříve spočítáme složky gradientu

$$\frac{\partial \mathcal{L}}{\partial w_0} = 2 \sum_{i=1}^{3} (w_0 + w_1 x_i - y_i) = 0 - 2 - 2 = -4 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = 2 \sum_{i=1}^{3} x_i (w_0 + w_1 x_i - y_i) = 0 + 0 + 2\sqrt{3/2}(-1) = -\sqrt{6} \tag{3}$$

Poté aktualizujeme váhy

$$\mathbf{w}^n = \mathbf{w} - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{w}}^T = \begin{bmatrix} 0 + 0.4 \\ 0 + 0.1\sqrt{6} \end{bmatrix} \doteq \begin{bmatrix} 0.4 \\ 0.245 \end{bmatrix} \tag{4}$$

1.3 Bonus

Vyřešte problém pomocí pseudoinverze.
Hint: Napište problém ve tvaru $\mathbf{A}\mathbf{w} = \mathbf{y}$. Dokážete vyjádřit matici $\mathbf{A}$ jako součin skaláru s maticí, která má ortonormální sloupce?

Řešení

Problém vyjádříme ve tvaru $\mathbf{A}\mathbf{w} = \mathbf{y}$ jako

$$\begin{bmatrix} 1 & -\sqrt{3/2} \\ 1 & 0 \\ 1 & \sqrt{3/2} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \tag{5}$$

Matici $\mathbf{A}$ napíšeme jako součin skaláru t s maticí, která má ortonormální sloupce $\mathbf{U}$

$$\mathbf{A} = t\mathbf{U} = \sqrt{3} \begin{bmatrix} \sqrt{1/3} & -\sqrt{1/2} \\ \sqrt{1/3} & 0 \\ \sqrt{1/3} & \sqrt{1/2} \end{bmatrix} \tag{6}$$

Matice $\mathbf{A}$ je úzká a má lineárně nezávislé sloupce, řešení tedy najdeme pomocí levé pseudoinverze jako $\mathbf{w} = \mathbf{A}^+\mathbf{b}$, kde $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = (t^2\mathbf{U}^T\mathbf{U})^{-1}t\mathbf{U}^T$. Díky vlastnosti matice s ortonormálními sloupci $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ se pseudoinverze zjednoduší na $\mathbf{A}^+ = \mathbf{U}^T/t$. Optimální hodnota vah je tedy

$$\mathbf{w} = \mathbf{U}^t\mathbf{b}/t = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \tag{7}$$