Градиент на функцията softmax

26 ноември 2025 г.

Функцията softmax : $\mathbb{R}^n \to \mathbb{R}^n$ се дефинира като:

$$\text{softmax}(\boldsymbol{z})_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

или векторно:

$$\text{softmax}(\boldsymbol{z}) = \frac{e^{\boldsymbol{z}}}{\sum_j e^{z_j}}$$

Градиентът ѝ е матрица $\in \mathbb{R}^{n \times n}$, чиито елементи са:

$$\begin{aligned}
\left(\frac{\partial\ \text{softmax}(\boldsymbol{z})}{\partial\ z}\right)_{ik} &= \frac{\partial\ \text{softmax}(\boldsymbol{z})_i}{\partial\ z_k} \\
&= \frac{\left(\frac{\partial e^{z_i}}{\partial z_k}\right)\left(\sum_j e^{z_j}\right) - e^{z_i}\left(\frac{\partial \sum_j e^{z_j}}{\partial z_k}\right)}{\left(\sum_j e^{z_j}\right)^2} \\
&= \frac{e^{z_i}\delta_{ik}\left(\sum_j e^{z_j}\right) - e^{z_i}e^{z_k}}{\left(\sum_j e^{z_j}\right)^2} \\
&= \frac{e^{z_i}}{\sum_j e^{z_j}}\left(\delta_{ik} - \frac{e^{z_k}}{\sum_j e^{z_j}}\right) \\
&= \text{softmax}(\boldsymbol{z})_i\left(\delta_{ik} - \text{softmax}(\boldsymbol{z})_k\right)
\end{aligned}$$

Същата сметка може да се напише за целия ред на градиента:

$$\begin{aligned}
\frac{\partial\ \text{softmax}(\boldsymbol{z})_i}{\partial\ \boldsymbol{z}} &= \frac{\left(\frac{\partial e^{z_i}}{\partial \boldsymbol{z}}\right)\left(\sum_j e^{z_j}\right) - e^{z_i}\left(\frac{\partial \sum_j e^{z_j}}{\partial \boldsymbol{z}}\right)}{\left(\sum_j e^{z_j}\right)^2} \\
&= \frac{e^{z_i}\text{eye}(i)\left(\sum_j e^{z_j}\right) - e^{z_i}e^{\boldsymbol{z}}}{\left(\sum_j e^{z_j}\right)^2} \\
&= \frac{e^{z_i}}{\sum_j e^{z_j}}\left(\text{eye}(i) - \frac{e^{\boldsymbol{z}}}{\sum_j e^{z_j}}\right) \\
&= \text{softmax}(\boldsymbol{z})_i\left(\text{eye}(i) - \text{softmax}(\boldsymbol{z})\right)
\end{aligned}$$

, където eye(i) е вектор с 1 на позиция i и 0 навсякъде другаде (numpy.eye(n, k=i)). Така може да напишем и матричния израз за градиента на softmax:

$$\frac{\partial\ \text{softmax}(\boldsymbol{z})}{\partial\ \boldsymbol{z}} = \text{softmax}(\boldsymbol{z})^T \cdot (\text{eye}(n) - \text{softmax}(\boldsymbol{z}))$$