Função de ativação: Softmax

Resumo teórico

Função:

$$\pi_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

$$\frac{\partial \pi}{\partial \mathbf{x}} = \frac{e^{\mathbf{x}}}{\sum_{j=1}^{n} e^{x_j}}$$

Como a função de probabilidade é uma função vetorial, a sua derivada é uma matriz jacobiana:

$$\mathbf{J}_{\pi}(\mathbf{x}) = \begin{pmatrix} \frac{\partial \pi_1}{\partial x_1} & \frac{\partial \pi_1}{\partial x_2} & \cdots & \frac{\partial \pi_1}{\partial x_n} \\ \frac{\partial \pi_2}{\partial x_1} & \frac{\partial \pi_2}{\partial x_2} & \cdots & \frac{\partial \pi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_m}{\partial x_1} & \frac{\partial \pi_m}{\partial x_2} & \cdots & \frac{\partial \pi_m}{\partial x_n} \end{pmatrix}$$

Usando a derivada logarítmica:

$$\frac{\partial \pi_i}{\partial x_j} = \pi_i \cdot \frac{\partial log(\pi_i)}{\partial x_j}$$

$$log(\pi_i) = log(\frac{e^{x_i}}{\sum_{l=1}^{n} e^{x_l}}) = x_i - log(\sum_{l=1}^{n} e^{x_l})$$

$$\frac{\partial log(\pi_i)}{\partial x_j} = \frac{\partial x_i}{\partial x_j} - \frac{\partial log(\sum_{l=1}^{n} e^{x_l})}{\partial x_j}$$

$$\frac{\partial x_i}{\partial x_j} = \begin{cases} 1 : i = j \\ 0 : i \neq j \end{cases}$$

$$\frac{\partial log(\sum_{l=1}^{n} e^{x_l})}{\partial x_j} = \frac{e^{x_j}}{\sum_{l=1}^{n} e^{x_l}}$$

$$\frac{\partial log(\pi_i)}{\partial x_j} = 1(i = j) - \frac{e^{x_j}}{\sum_{l=1}^{n} e^{x_l}}$$

$$\frac{\partial log(\pi_i)}{\partial x_j} = 1(i = j) - \pi_j$$

$$\frac{\partial \pi_i}{\partial x_j} = \pi_i \cdot \frac{\partial log(\pi_i)}{\partial x_j} = \pi_i \cdot (1(i=j) - \pi_j)$$

$$\frac{\partial \pi_i}{\partial x_j} = \pi_i(i=j) - \pi_i \cdot \pi_j$$

$$\frac{\partial \pi_i}{\partial x_j} = \begin{cases} \pi_i \cdot (1-\pi_j) : i = j \\ -\pi_i \cdot \pi_j : i \neq j \end{cases}$$

Revisitando a matriz jacobiana:

$$\mathbf{J}_\pi(\mathbf{x}) = \begin{pmatrix} \pi_1 \cdot (1-\pi_1) & -\pi_1 \cdot \pi_2 & \cdots & -\pi_1 \cdot \pi_n \\ -\pi_2 \cdot \pi_1 & \pi_2 \cdot (1-\pi_2) & \cdots & -\pi_2 \cdot \pi_n \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_m \cdot \pi_1 & -\pi_m \cdot \pi_2 & \cdots & \pi_m \cdot (1-\pi_n) \end{pmatrix}$$

Para a implementação é possível criar a Jacobiana a partir de uma matrix identidade:

$$\mathbf{J}_\pi(\mathbf{x}) = \pi \cdot \mathbf{I} - \pi^T \cdot \pi$$

E finalmente, o valor da derivada da função vetorial, é calculado com a multiplicação da entrada com a jacobiana:

$$\pi\text{'}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{J}_\pi(\mathbf{x})$$

$$\pi\text{'}(\mathbf{x}) = \mathbf{x} \cdot (\pi \cdot \mathbf{I} - \pi^T \cdot \pi)$$