Homework 1 Question3

December 12, 2025

1 3.3

(3) Comparação entre Softmax, Sparsemax e Relumax

Consideremos o caso $K = 2$, com logits

$$z = [0,\ t].$$

Softmax. A probabilidade da segunda classe é

$$\text{softmax}(z)_2 = \frac{e^t}{e^0 + e^t} = \frac{e^t}{1 + e^t}.$$

A derivada em relação a t é

$$\frac{d}{dt}\,\text{softmax}(z)_2 = \frac{(1 + e^t)e^t - e^t e^t}{(1 + e^t)^2} = \frac{e^t}{(1 + e^t)^2}.$$

Sparsemax. A sparsemax é definida por

$$\text{sparsemax}(z)_i = [z_i \tau]_+, \quad \text{com } \sum_i [z_i - \tau]_+ = 1.$$

Para $z = [0, t]$ com $t > 1$, obtemos

$$\text{sparsemax}(z) = [0,\ 1].$$

Logo,

$$\frac{d}{dt}\,\text{sparsemax}(z)_2 = 1.$$

Relumax. Definimos a relumax como

$$\text{relumax}(z)_i = \frac{\text{ReLU}(z_i - \max(z) + b)}{\sum_j \text{ReLU}(z_j - \max(z) + b)}.$$

Para $z = [0, t]$, temos $\max(z) = t$, logo:

$$\text{relumax}(z)_2 = \frac{\text{ReLU}(b)}{\text{ReLU}(-t+b) + \text{ReLU}(b)}.$$

Se $b > 0$ e $-t + b > 0$, então

$$\text{relumax}(z)_2 = \frac{b}{(-t+b)+b} = \frac{b}{-t+2b}.$$

A derivada em relação a t é então

$$\frac{\partial}{\partial t}\,\text{relumax}(z)_2 = \begin{cases} \dfrac{-b}{(-t+2b)^2}, & -t+b > 0 \\ 0, & -t+b \leq 0 \end{cases}$$

e a derivada em relação a b é

$$\frac{\partial}{\partial b}\,\text{relumax}(z)_2 = \begin{cases} \dfrac{-t}{(-t+2b)^2}, & -t+b > 0 \\ 0, & -t+b \leq 0. \end{cases}$$

Resumo.

- O softmax é suave e tem derivada sempre positiva.
- A sparsemax é linear por partes e pode ter gradiente constante.
- A relumax introduz uma normalização baseada em ReLU e depende do parâmetro b.