Homework 1 Question3

December 12, 2025

1 3.2

(a) Shift-invariance e formas fechadas

Softmax. Seja $z \in \mathbb{R}^K$. O softmax é definido por

$$\operatorname{softmax}(z)_i \;=\; \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}.$$

Para qualquer constante $c \in \mathbb{R}$, verifica-se a invariância por translação:

$$\operatorname{softmax}(z + c\mathbf{1})_i = \frac{e^{z_i + c}}{\sum_{j=1}^{K} e^{z_j + c}} = \frac{e^{z_i} e^c}{\sum_{j=1}^{K} e^{z_j} e^c} = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} = \operatorname{softmax}(z)_i.$$

Sparsemax (forma fechada). A sparsemax é definida como a projeção Euclidiana no simplex:

$$\operatorname{sparsemax}(z) = \arg\min_{p \in \Delta_K} \|p - z\|^2, \qquad \Delta_K = \Big\{ p \in \mathbb{R}^K : \; p \geq 0, \; \sum_{i=1}^{K} p_i = 1 \Big\}.$$

A solução tem a forma fechada

$$\operatorname{sparsemax}(z)_i = \big[\, z_i - \gamma(z) \,\big]_+, \qquad [t]_+ := \max(t, 0),$$

onde o escalar $\gamma(z)$ é escolhido de modo a satisfazer

$$\sum_{i=1}^{K} \big[\, z_i - \gamma(z) \,\big]_+ = 1.$$

Se $z_{(1)} \geq \ldots \geq z_{(K)}$ denotam os logits ordenados por ordem decrescente, e

$$k = \max\Big\{ r \in \{1, \ldots, K\} : 1 + r z_{(r)} > \sum_{j=1}^{r} z_{(j)} \Big\},$$

então

$$\operatorname{sparsemax}(z{+}c)_i = \big[\, z_i{+}c{-}\gamma(z{+}c) \,\big]_+ = \big[\, z_i{+}c{-}\gamma(z){+}c \,\big]_+ = \big[\, z_i{-}\gamma(z) \,\big]_+ = \operatorname{sparsemax}(z)_i$$

Relumax. Definimos a relumax como uma normalização baseada em ReLU:

$$\operatorname{relumax}(z)_i = \frac{\operatorname{ReLU}\big(z_i - \max(z) + b\big)}{\sum_{j=1}^{K} \operatorname{ReLU}\big(z_j - \max(z) + b\big)}, \qquad \operatorname{ReLU}(t) = \max(t, 0).$$

Para qualquer constante $c \in \mathbb{R}$,

$$\operatorname{relumax}(z + c\mathbf{1})_i = \frac{\operatorname{ReLU}\big(z_i + c - \max(z + c\mathbf{1}) + b\big)}{\sum_{j=1}^{K} \operatorname{ReLU}\big(z_j + c - \max(z + c\mathbf{1}) + b\big)}.$$

Como $\max(z + c\mathbf{1}) = \max(z) + c$, obtemos

$$\operatorname{relumax}(z + c\mathbf{1})_i = \frac{\operatorname{ReLU}\big(z_i - \max(z) + b\big)}{\sum_{j=1}^{K} \operatorname{ReLU}\big(z_j - \max(z) + b\big)} = \operatorname{relumax}(z)_i.$$