Exercise Sheet 5

Arsen Nuramatov
Pylyp Filippov

1.

a) $\sigma(\beta^T x) = \dfrac{1}{1+\exp(-\beta^T x)} = \dfrac{1}{1+\exp(-\beta_i x_i)}$

$$\frac{\partial \sigma(\beta^T x)}{\partial x_j} = \frac{-1}{[1+\exp(-\beta_i x_i)]^2} \exp(-\beta_i x_i) \, -\frac{\partial}{\partial x_j}(\beta_i x_j) =$$

$$= \frac{-1}{[1+\exp(-\beta^T x)]^2} \exp(-\beta^T x)(-\beta_j) = \frac{\beta_j}{[1+\exp(-\beta^T x)]^2} \exp[-\beta^T x]$$

$$\frac{\partial \sigma(\beta^T x)}{\partial x} = \beta^T \frac{e^{-\beta^T x}}{[1+e^{-\beta^T x}]^2} = \beta^T \sigma(\beta^T x)[1 - \sigma(\beta^T x)]$$

$$\frac{1}{\;} \qquad \frac{e^{-\beta^T x}}{1+e^{-\beta^T x}}$$

b)

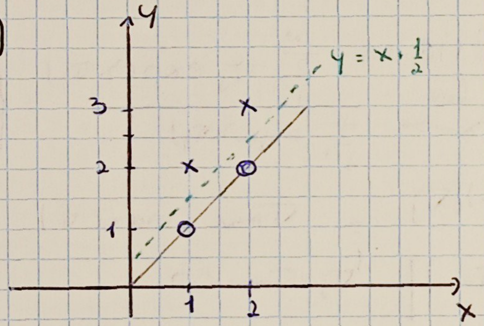
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1-e^{-2x}}{1+e^{-2x}} = -1 + \frac{2}{1+e^{-2x}} = \frac{2}{1+e^{-2x}} - 1 =$$

$$\left[\frac{1-e^{-2x}}{1+e^{-2x}} \; \frac{1+e^{-2x}}{-1} \; 2 \right] - \; 2 \quad = 2\sigma(2x) - 1$$

We then can write

$\tanh(x) = 2\sigma(2x) - 1$ → $\tanh(x)$ is just scaled and shifted version of the logistic sigmoid

c)



class 1 → (1,1), (2,2)

class 2 → (1,2), (2,3)

I want to give the sigmoid a value 0 for class 1 and 1 for 2. Then

$w^T x + b < 0$ for class 1

$w^T x + b > 0$ for class 2

Eq.

$$\left.\begin{aligned} w_1 + w_2 + b &< 0 \\ 2w_1 + 2w_2 + b &< 0 \\ w_1 + w_2 + \tfrac{b}{2} &< 0 \end{aligned}\right]$$

$$\left.\begin{aligned} w_1 + 2w_2 + b &> 0 \\ 2w_1 + 3w_2 + b &> 0 \end{aligned}\right]$$

Let's try $w_1 = -w_2$ → $b < 0$ If $w_2 = 5$, $w_1 = -5$

$w_2 + b > 0$

$\cancel{-w_2 + b > 0}$

$b > -5$

$b = -3$ (for example)

Then we get $\vec{w} = 5(-1, 1)$, $b = -3$

$w^T x + b = 5(y - x) - 3$

(1,1) → −3
(2,2) → −3
(1,2) → 2
(2,3) → 2

To make it a bit smoother we could choose $w = (-1, 1)$ & $b = -\frac{1}{2}$

⇒ $W^T = (-1, 1)$ & $b = -1/2$ works to separate both classes.

3.

a) We start by calculating

$$\text{soft}(\arg)\max(\bar{\alpha} + \beta\vec{\sigma}; \lambda)_k = \frac{\exp[\lambda[\cancel{\alpha +} \beta\sigma_k]]}{\sum_{j=1}^{K} \exp[\lambda(\alpha + \beta\sigma_j)]} =$$

$$\bar{\alpha} = \cancel{\alpha(1,1,1)} = \cancel{(\alpha, \dots, \alpha)} = \alpha(1, 1, \dots, 1)$$

$$= \frac{\exp(\lambda\alpha)\exp(\lambda\beta\sigma_k)}{\sum_{j=1}^{N} \exp[\lambda\alpha]\exp[\lambda\beta\sigma_j]} = \frac{\exp(\lambda\beta\sigma_k)}{\sum_{j=1}^{N} \exp[\lambda\beta\sigma_j]}$$

As we can see, soft(arg)max is invariant under a constant offset but not rescaling of its output. We conclude that soft(arg)max will give the same results for $\vec{\sigma}^1 = (1,2,3)^T$ & $\vec{\sigma}^2 = (11, 12, 13)$, since the difference is a constant offset $(10, 10, 10)$.

Note: for the offset, we add the same value α to all the components of $\vec{\sigma}$.

d) $$\frac{\partial \text{lse}(\vec{\sigma}; \lambda)}{\partial \sigma_k} = \frac{1}{\lambda} \frac{1}{\sum_{j=1}^{K} \exp(\lambda\sigma_j)} \times \sum_{j=1}^{K} \lambda\delta_{kj}\exp(\lambda\sigma_j) =$$

$$= \frac{\exp(\lambda\sigma_k)}{\sum_{j=1}^{N} \exp(\lambda\sigma_j)} = \text{soft}(\arg)\max(\vec{\sigma}, \lambda)_k.$$

e) $$\lim_{\lambda\to\infty} \frac{1}{\lambda}\log\left(\sum_{j=1}^{K} \exp(\lambda\sigma_j)\right) = \frac{\infty}{\infty} \overset{\text{l'Hôpital}}{=} \lim_{\lambda\to\infty} \frac{\sum_{j=1}^{N} \sigma_j \exp(\lambda\sigma_j)}{\sum_{j=1}^{K} \exp(\lambda\sigma_j)} = \left\{ \begin{array}{l} \text{Multiply} \\ \text{by} \\ \exp(-\lambda\sigma_{max}) \end{array} \right\}$$

$$= \lim_{\lambda\to\infty} \frac{\sigma_{max} + \sum_{j=1, j\neq j_{max}}^{N} \exp[\lambda(\sigma_j - \sigma_{max})]\sigma_j}{1 + \sum_{j=1, j\neq j_{max}}^{N} \exp[\lambda(\sigma_j - \sigma_{max})]} \underset{\sigma_j - \sigma_{max} \leq 0}{=} \sigma_{max} = \max(\vec{\sigma}).$$