

Exercises 2.5

Exercise 2.5.1

- Show that the logistic function σ satisfies the inequality $0 < \sigma'(x) \leq \frac{1}{4}$, for all $x \in \mathbb{R}$.
- How does the inequality change in the case of the functions σ_c ?

Exercise 2.5.2

Let $S(x)$ and $H(x)$ denote the bipolar step function and the Heaviside function, respectively. Show that:

- $S(x) = 2H(x) - 1$
- $\text{ReLU}(x) = \frac{1}{2}x(S(x) + 1)$

Exercise 2.5.3

Show that the softplus function, $sp(x)$, satisfies the following properties:

- $sp'(x) = \sigma(x)$, where $\frac{1}{1+e^{-x}}$
- Show that $sp(x)$ is invertible with inverse $sp^{-1}(x) = \ln(e^x - 1)$
- Use the softplus function to show the formula $\sigma(x) = 1 - \sigma(-x)$

Exercise 2.5.4

Show that $\tanh(x) = 2\sigma(2x) - 1$

Exercise 2.5.5

Show that the softsign function, $so(x)$, satisfies the following properties:

- It is strictly increasing;
- It is onto $(-1, 1)$, with the inverse $so^{-1}(x) = \frac{1}{1-|x|}$, for $|x| < 1$.
- $so(|x|)$ is subadditive, i.e., $so(|x + y|) \leq so(|x|) + so(|y|)$.

Exercise 2.5.6

Show that the softmax function is invariant with respect to the addition of constant vectors $\mathbf{c} = (c_1 \dots c_n)^T$, i.e.,

$$\text{softmax}(\mathbf{y} + \mathbf{c}) = \text{softmax}(\mathbf{y}).$$

This property is used in practice by replacing $\mathbf{c} = -\max_i y_i$, a fact that leads to a more stable numerically variant of this function.

Exercise 2.5.7

Let $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined by $\rho(y) \in \mathbb{R}^n$, with $\rho(y)_i = \frac{y_i^2}{\|y\|}$. Show that:

- a. $0 \leq \rho(y)_i \leq 1$ and $\sum_i \rho(y)_i = 1$.
- b. The function ρ is invariant with to multiplication by nonzero constant, i.e., $\rho(\lambda y) = \rho(y)$ for any $\lambda \in \mathbb{R}/0$. Taking $\lambda = \frac{1}{\max_i y_i}$ leads in practice to a more stable version of this function.

Exercise 2.5.8 (cosine squasher)

Show that the function $\varphi(x) = \frac{1}{2}(1 + \cos(x + \frac{3\pi}{2}))1_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(x) + 1_{(\frac{\pi}{2}, \infty)}(x)$ is a squashing function.

Exercise 2.5.9

- a. Show that any squashing function is a sigmoidal function.
- b. Give an example of a sigmoidal function which is not a squashing function.

SOLUTIONS

2.5.1 (a)

Computing the derivative of σ we find: $\sigma'(x) = \frac{d}{dx} \frac{1}{1+e^{-x}} = \frac{d}{dx} \frac{e^x}{1+e^x} = \frac{e^x}{(1+e^x)^2}$. From the inequality $1 \leq (1+e^x)^2$ and the non-negativeness of the exponential function follows that $0 \leq \frac{e^x}{(1+e^x)^2}$.

Now let's prove that in $x = 0$ the function has a local maximum in $[-1, 1]$, this will imply $0 \leq \frac{e^x}{(1+e^x)^2} \leq \sigma'(0)$, $\sigma'(0) = \frac{1}{4}$. By computing the first derivative of σ' we find: $\sigma''(x) = e^x \frac{1-e^x}{(1+e^x)^3}$. The critical will be found by solving the equation $\sigma''(x) = 0$.

From $\sigma''(x) = e^x \frac{1-e^x}{(1+e^x)^3} = 0$ follows that $1 - e^x = 0$, it is straightforward to check that the solution is $x = 0$. It rests to determine the nature of the extremizing point. To achieve this goal is necessary to calculate the second derivative of σ' .

$$\begin{aligned}\sigma'''(x) &= \frac{d}{dx} \frac{e^x - e^{2x}}{(1+e^x)^3} = \frac{(e^x - 2e^{2x})(1+e^x)^3 - 3(1+e^x)^2 e^x (e^x - e^{2x})}{(1+e^x)^6} \\ &= \frac{e^x \{1 - 4e^x + e^{2x}\} (1+e^x)^2}{(1+e^x)^6} = \frac{e^x \{1 - 4e^x + e^{2x}\}}{(1+e^x)^4}\end{aligned}$$

We clearly have $\sigma'''(0) < 0$, then $x = 0$ is a local maximum for σ' , i.e. $\forall x \in [-1, 1]$, $\sigma'(x) \leq \frac{1}{4}$. On the other hand, the function σ' decreases on the intervals $(-\infty, -1)$ and $(1, \infty)$ this implies that:

$\sup_{x \in (1, \infty)} \sigma'(x) = \frac{e}{(1+e)^2} = \frac{e^{-1}}{(1+e^{-1})^2} = \sup_{x \in (-\infty, -1)} \sigma'(x)$. From the fact that $\frac{e}{(1+e)^2} < \frac{1}{4}$ follows that $0 \leq \sigma'(x) \leq \frac{1}{4}$ is valid $\forall x \in \mathbb{R}$.