

Supplementary Materials

1 Proof of Lemma 4.1

Lemma 4.1. The scaled soft-plus function $\gamma_s(x) = s \log(1 + \exp(x/s))$ ($s > 0$) is convex and $\log(\gamma_s(x))$ is concave.

Proof. Since s is a positive constant, we only need to show that the soft-plus function $\gamma(x) = \gamma_1(x)$ is convex and log concave. Then it is straightforward to show that the scaled version is also convex and log concave. To this end, we first observe that

$$\gamma(x) = \log(1 + \exp(x)) = -\log(\sigma(-x))$$

where $\sigma(x) = 1/(1 + \exp(-x))$ is the sigmoid activation function. We then take the gradient of $\gamma(x)$,

$$\frac{d\gamma(x)}{dx} = -\frac{1}{\sigma(-x)}\sigma(-x)(1 - \sigma(-x))(-1) = \sigma(x). \quad (1)$$

Note that we have used a known fact that $\frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$. Next, we take the second derivative,

$$\frac{d^2\gamma(x)}{dx^2} = \sigma(x)(1 - \sigma(x)).$$

Since $\forall x \in \mathbb{R}$, we have $0 \leq \sigma(x) \leq 1$, we must have $\frac{d^2\gamma(x)}{dx^2} \geq 0$. Therefore, $\gamma(x)$ is convex.

Now, let us look at $h(x) = \log(\gamma(x))$. First, we can derive the first derivative based on (1),

$$\frac{dh(x)}{dx} = \frac{1}{\gamma(x)} \frac{d\gamma(x)}{dx} = \frac{\sigma(x)}{\gamma(x)}.$$

Then, the second derivative is

$$\frac{d^2h(x)}{dx^2} = \frac{\frac{d\sigma(x)}{dx}\gamma(x) - \sigma(x)\frac{d\gamma(x)}{dx}}{(\gamma(x))^2} = \frac{\sigma(x) \cdot g(x)}{(\gamma(x))^2} \quad (2)$$

where

$$g(x) = (1 - \sigma(x))\gamma(x) - \sigma(x).$$

From (2), we can see that $\sigma(x) \geq 0$ and $(\gamma(x))^2 \geq 0$. Therefore, we only need to check if $g(x) \leq 0$ to show the concavity of $h(\cdot)$. Since $\gamma(x) = -\log(\sigma(-x)) = -\log(1 - \sigma(x))$, we can view $g(x)$ as a function of $t = 1 - \sigma(x)$, namely,

$$g(x) = g(t) = -t \log(t) - (1 - t) = t(1 - \log(t)) - 1,$$

and $0 \leq t \leq 1$. Note that $g(t) = 0$ when $t = 1$. We take the derivative of $g(\cdot)$ w.r.t t ,

$$\frac{dg(t)}{dt} = 1 - \log(t) + t(-\frac{1}{t}) = -\log(t) \geq 0.$$

Therefore, $g(t)$ is monotonically increasing with t . Since $0 \leq t \leq 1$, we always have $g(t) \leq g(t = 1) = 0$. Hence, $\forall x, g(x) \leq 0$. From (2), we have $\frac{d^2 h(x)}{dx^2} \leq 0$, and hence the log soft-plus function is concave. □