Supplementary Materials

1 Proof of Lemma 4.1

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

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

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

$$\frac{\mathrm{d}\gamma(x)}{\mathrm{d}x} = -\frac{1}{\sigma(-x)}\sigma(-x)(1-\sigma(-x))(-1) = \sigma(x). \tag{1}$$

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

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

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

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

$$\frac{\mathrm{d}h(x)}{\mathrm{d}x} = \frac{1}{\gamma(x)}\frac{\mathrm{d}\gamma(x)}{\mathrm{d}x} = \frac{\sigma(x)}{\gamma(x)}.$$

Then, the second derivative is

$$\frac{\mathrm{d}^2h(x)}{\mathrm{d}x^2} = \frac{\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}\gamma(x) - \sigma(x)\frac{\mathrm{d}\gamma(x)}{\mathrm{d}x}}{\big(\gamma(x)\big)^2} = \frac{\sigma(x)\cdot g(x)}{\big(\gamma(x)\big)^2} \tag{2}$$

where

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

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

$$\frac{\mathrm{d}g(t)}{\mathrm{d}t} = 1 - \log(t) + t(-\frac{1}{t}) = -\log(t) \geq 0.$$

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 $\frac{\mathrm{d}^2h(x)}{\mathrm{d}x^2} \leq 0$, and hence the log soft-plus function is concave.

□

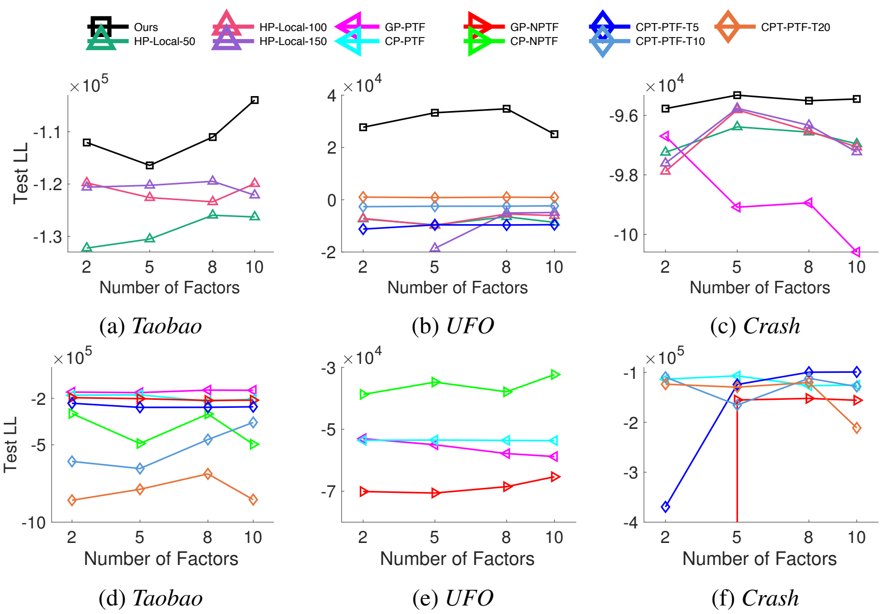

Figure 1: Test log-likelihood (LL) on real-world datasets. HP-Local-{50, 100, 150} means running HP-Local with window size 50, 100 and 150. CPT-PTF-{5,10,20} are CPT-PTF with 5, 10 and 20 time steps.

2 Complete Test Log-Likelihood Results

In Fig. 1, we report the test log-likelihood (LL) of all the methods in the three real-word datasets examined in Section 6.1 of the main paper. Note that the first row are the same as Fig. 1 in the main paper. The second row shows the prediction accuracy of the remaining methods, which are much worse than the results in the first row.