CP − Temporal − PP

We introduce a Time max $T : d_T \times R$

We partition the events into time buckets, then use PP

Now we have $(U^1, U^2, \ldots U^K)$

$U^1, \ldots U^{K-1}$ are latents factors for tensor dimensions

U^K is the time factors

$$p(U^k) = N(\cancel{U^k}\, vec(U^k) \,|\, 0, I) \qquad 1 \le k \le K-1$$

for U^K, i.e. time factors

regarding SGD, refer to CP-PP derivation, the only difference is the grad. w.r.t prior

$$U_1^K \sim N(U_1^K \,|\, 0, I)$$

$$U_i^K \,|\, U_{i-1}^K \sim N(U_i^K \,|\, U_{i-1}^K, s^{-1} I)$$

$$s \sim Gamma(s \,|\, a_0, b_0)$$

~~for~~ for any entry $i = (i_1, i_2, \ldots i_{K-1}, i_K)$

$$p(Y_i | U) = PP(Y_i \,|\, \underbrace{e^{\langle U^1_{i_1} \circ U^2_{i_2} \cdots \circ U^K_{i_K} \rangle}}_{\lambda_i})$$

log(joint)

$$= \sum_{k=1}^{K-1} -\frac{1}{2} tr(U^{k^T} U^k)$$

$$+ \cancel{\ } -\frac{1}{2}\|U_1^k\|_2^2 + \sum_{j=2}^{d_T} \log N(U_j^k \mid U_{j-1}^k, s^{-1} \cdot I)$$

$$+ \log Gamma(s \mid a_0, b_0)$$

$$+ \sum_{i \in Observed} (-\lambda_i \cdot \Delta T_{i_K} + Y_i \log \lambda_i)$$

ΔT is the time interval for $U^k_{i_K}$ it can be identical for all $1 \ldots d_T$ or can be differents

$$= \log(Prior) + \sum_{\substack{j \\ \text{mini-batch}}} \frac{N}{|B_j|} \sum_{i \in B_j} (-\lambda_i T + Y_i \log \lambda_i)$$

SGD is the same as CP-PP, except for prior terms

we need cal. grad. w.r.t U^k & s

$$\log N(U_j^k \mid U_{j-1}^k, s^{-1} I) = \log |2\pi s^{-1} I|^{-\frac{1}{2}} \exp(-\frac{1}{2} s \|U_j^k - U_{j-1}^k\|_2^2)$$

$$= \frac{R}{2} \log s - \frac{1}{2} s (U_j^k - U_{j-1}^k)^T (U_j^k - U_{j-1}^k)$$

$$\frac{d \log N(U_j^k \mid U_{j-1}^k, s^{-1} I)}{ds} = \frac{R}{2} \frac{1}{s} - \frac{1}{2}\|U_j^k - U_{j-1}^k\|_2^2$$

$$\frac{d \log N(U_j^k \mid U_{j-1}^k, s^{-1} I)}{d U_j^k} = - s \cdot (U_j^k - U_{j-1}^k), \quad \frac{d \log N(U_j^k \mid U_{j-1}^k, s^{-1} I)}{d U_{j-1}^k} = - s \cdot (U_{j-1}^k - U_j^k)$$

$$\frac{d \log p(\text{joint})}{ds} = \frac{R(d_T - 1)}{2} \cdot \frac{1}{s} - \frac{1}{2} \sum_{j=2}^{d_T} \| U_j^k - U_{j-1}^k \|_2^2 + \frac{d \log \text{Gamma}(s \mid a_0, b_0)}{ds}$$

$$\log \frac{b_0^{a_0}}{\Gamma(a_0)} e^{-b_0 s} s^{a_0 - 1} = -b_0 s + (a_0 - 1) \log s$$

$$-b_0 + \frac{a_0 - 1}{s}$$

$$\boxed{\frac{d \log p(\text{joint})}{ds} = \frac{1}{s}\left(a_0 - 1 + \frac{R}{2}(d_T - 1)\right) - \left(b_0 + \frac{1}{2} \sum_{j=2}^{d_T} \| U_j^k - U_{j-1}^k \|_2^2 \right)}$$

$$\boxed{\frac{d \log p(\text{prior})}{dU_1^k} = -U_1^k - s \cdot (U_1^k - U_2^k)}$$

$$\boxed{\frac{d \log p(\text{Prior})}{dU_j^k} = -s(U_j^k - U_{j-1}^k) - s(U_j^k - U_{j+1}^k) \qquad 2 \le j \le d_T - 1}$$

$$\boxed{\frac{d \log p(\text{Prior})}{dU_{d_T}^k} = -s(U_{d_T}^k - U_{d_T - 1}^k)}$$

$$\frac{df}{dy} \cdot \frac{dy}{dx}$$

$$y = e^x$$

$$\frac{dy}{dx} = e^x = y.$$

$$\frac{d \log(-)}{d \log s} = \frac{d \log(-)}{ds} \cdot \frac{ds}{d \log s} = \frac{d \log(-)}{ds} \cdot s$$