Session 8 : Paper - Yu et al., 2009 3/21/2022

(I'm presenting on figure 3)

Gaussian-Process Factor Analysis for low-dimensional, single-trial analysis of neural population activity

"two-stage methods":

1. smooth firing-rate profiles for each neuron on a single trial
2. apply a "static" dimensionality-reduction technique (e.g. PCA)
   - i.e. do not account for the relationship of the data points over time

Motivation for GPFA

- extension of FA w/ smoothing & dimensionality reduction in a common probabilistic framework
- GPFA is a set of factor analyzers linked together in the low-dim. state space by a GP
  - one FA per time point w/ identical parameters

Mathematical description of GPFA

- bin spike counts & sqrt transform
  - but no pre-smoothing
- $y_{:,t} \in \mathbb{R}^{q \times 1}$: high-dim. vector of sqrt. spike counts for time point $t: 1, \dots, T$ where q is the number of neurons
- $x_{:,t} \in \mathbb{R}^{p \times 1}$: "latent neural state" w $p < q$ latent dimensions

$$X = [x_{:,1}, \dots, x_{:,T}] \in \mathbb{R}^{p \times T}$$

$$Y = [y_{:,1}, \dots, y_{:,T}] \in \mathbb{R}^{q \times T}$$

- linear Gaussian relationship b/w observations y & neural states x:

"emission model"

$$y_{:,t} \mid x_{:,t} \sim N(Cx_{:,t} + d, R) \qquad \text{eq. 1}$$

$C \in \mathbb{R}^{q \times p}$, $d \in \mathbb{R}^{q \times 1}$, $R \in \mathbb{R}^{q \times q}$

- constrained to be diagonal like in FA

- the neural states $\mathbf{x}_{:,t}$ across t are related through a GP
  - a separate GP for each latent dim. $i = 1, \dots, p$

eq. 2 $\quad \mathbf{x}_{i,:} \sim N(0, K_i) \qquad K_i \in \mathbb{R}^{T \times T}$: cov. mat. for i^{th} GP

"transition model"

- used "squared exponential" (SE) covariance fxn :

eq 3 $\quad K_i(t_1, t_2) = \sigma_{f,i}^2 \exp\left[-(t_1 - t_2)^2 / (2\tau_i^2)\right] + \sigma_{n,i}^2 \cdot \delta_{t_1, t_2}$

$\sigma_{f,i}^2$: signal variance $\in \mathbb{R}_+$ (ie. >0)
τ_i : characteristic timescale $\in \mathbb{R}_+$
$\sigma_{n,i}^2$: GP noise variance $\in \mathbb{R}_+$

- non-identifiability b/w X and C
  - solution : fix the scale of X and leave C unconstrained
  - fix $K_i(t,t) = 1$ which sets the prior $\mathbf{x}_{:,t} \sim N(0, \mathbf{I})$ at each t
    - achieved by setting $\sigma_{f,i}^2 = 1 - \sigma_{n,i}^2$ w/ $0 < \sigma_{n,i}^2 \leq 1$
    - set $\sigma_{n,i}^2 = 10^{-3}$ (some small number) — to keep $\sigma_{f,i}^2 > 0$
  - result : τ_i is the only parameter of the cov. fxn that is learned

<u>Fitting the GPFA model</u>

- use EM
- parameters to fit : $\theta = \{C, d, R, \tau_1, \dots, \tau_T\}$
- maximize prob. of the observed data Y
- E : update the neural trajectories $\mathbf{x}$
  - use the recent parameter updates to evaluate the prob. of neural trajectories
  - $P(X|Y)$
- M : update model parameters θ using $P(X|Y)$ from E

Appendix: GPFA model fitting

E : compute probabilities $P(X|Y)$ for all possible neural trajectories X given observed activity Y

- find joint dist. $P(X,Y)$ which will be Gaussian
- re-expressions of Eq. 1 & 2

$$\bar{x} \sim N(0, \bar{K}) \quad \boxed{a1}$$

$$\bar{y} \mid \bar{x} \sim N(\bar{C}\bar{x} + \bar{d} \mid \bar{R}) \quad \boxed{a2}$$

$$\bar{x} = [x'_{:,1}, \dots, x'_{:,T}]' \in \mathbb{R}^{pT \times 1}$$

$$\bar{y} = [y'_{:,1}, \dots, y'_{:,T}]' \in \mathbb{R}^{qT \times 1}$$

$\bar{C} \in \mathbb{R}^{qT \times pT}$
$\bar{R} \in \mathbb{R}^{qT \times qT}$ } block diagonal matrices

$\bar{d} \in \mathbb{R}^{qT \times 1}$: concatenation of T copies of d

$$\bar{K} = \begin{bmatrix} \bar{K}_{11} & \cdots & \bar{K}_{1T} \\ \vdots & \ddots & \vdots \\ \bar{K}_{T1} & \cdots & \bar{K}_{TT} \end{bmatrix} \in \mathbb{R}^{pT \times pT} \quad \boxed{a3}$$

$$\bar{K}_{t_1 t_2} = \underbrace{\mathrm{diag}}_{\text{makes a diagonal matrix}}\{K_1(t_1, t_2), \dots, K_p(t_1, t_2)\} \in \mathbb{R}^{p \times p}$$

joint dist of $\bar{x}$ & $\bar{y}$: (from a1 & a2)

$$P(X,Y): \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \bar{d} \end{bmatrix}, \begin{bmatrix} \bar{K} & \bar{K}\bar{C}'^{\,\text{Transpose!}} \\ \bar{C}\bar{K} & \bar{C}\bar{K}\bar{C}' + \bar{R} \end{bmatrix} \right) \quad \boxed{a4}$$

- w/ joint dist (a4), get conditional $P(X|Y)$

$$\bar{x} \mid \bar{y} \sim N\left(\bar{K}\bar{C}'(\bar{C}\bar{K}\bar{C}' + \bar{R})^{-1}(\bar{y} - d),\ \bar{K} - \bar{K}\bar{C}'(\bar{C}\bar{K}\bar{C}' + \bar{R})^{-1}\bar{C}\bar{K}\right)$$

$$\boxed{E[\bar{x} \mid \bar{y}] = \bar{K}\bar{C}'(\bar{C}\bar{K}\bar{C}' + \bar{R})^{-1}(\bar{y} - d)} \qquad \boxed{\bar{y} \sim N(\bar{d}, \bar{C}\bar{K}\bar{C}' + R)}$$

(M) · maximizing $\mathcal{E}(\theta) = E[\log P(X, Y \mid \theta)]$ where $\theta = \{C, d, R, \tau_1, \dots, \tau_p\}$

· notation:

$$\langle x_{:,t} \rangle = E[x_{:,t} \mid Y] \in \mathbb{R}^{p \times 1}$$

$$\langle x_{:,t} x'_{:,t} \rangle = E[x_{:,t} x'_{:,t} \mid Y] \in \mathbb{R}^{p \times p}$$

$$\langle x'_{i,:} x_{i,:} \rangle = E[x'_{i,:} x_{i,:} \mid Y] \in \mathbb{R}^{T \times T}$$

· maximize $\mathcal{E}(\theta)$ w.r.t C & d:

$$[C \;\; d] = \left(\sum_{t=1}^{T} y_{:,t} \cdot [\langle x_{:,t} \rangle' \; 1] \right) \left(\sum_{t=1}^{T} \begin{bmatrix} \langle x_{:,t} x'_{:,t} \rangle & \langle x_{:,t} \rangle \\ \langle x_{:,t} \rangle' & 1 \end{bmatrix} \right)^{-1} \quad \boxed{a8}$$

· maximize $\mathcal{E}(\theta)$ w.r.t R: (using values of C & d from a8)

$$R = \frac{1}{T} \underbrace{\operatorname{diag}}_{\text{zeros all off-diagonal elements}} \left\{ \sum_{t=1}^{T} (y_{:,t} - d)(y_{:,t} - d)' - \left(\sum_{t=1}^{T} (y_{:,t} - d) \langle x_{:,t} \rangle' \right) C' \right\}$$

update time scale parameters τ_i w/ gradient optimization

$$\frac{\partial \mathcal{E}(\theta)}{\partial \tau_i} = \operatorname{tr}\left(\left[\frac{\partial \mathcal{E}(\theta)}{\partial K_i} \right]' \frac{\partial K_i}{\partial \tau_i} \right)$$

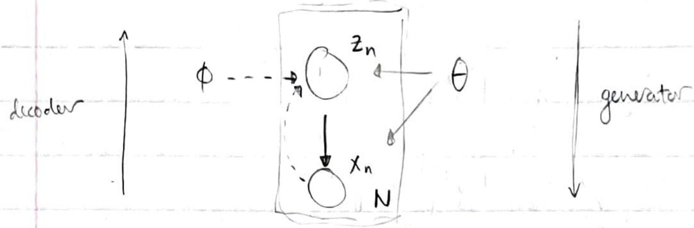
$$\frac{\partial \mathcal{E}(\theta)}{\partial K_i} = \frac{1}{2}\left(-K_i^{-1} + K_i^{-1} \langle x'_{i,:} x_{i,:} \rangle K_i^{-1} \right)$$

$$\frac{\partial K(t_1, t_2)}{\partial \tau_i} = \sigma_{f,i}^2 \frac{(t_1 - t_2)^2}{\tau_i^3} \exp\left(-\frac{(t_1 - t_2)^2}{2\tau_i^2} \right)$$

Session 8 : Lecture notes

1. Autoencoders

· "code" z generates observations x through $p_\theta(z) \to p_{\theta^*}(x|z)$



aim: given $p_\theta(z)$; $p_\theta(x|z)$, find θ that maximizes $p_\theta(x)$

problem: $p_\theta(x)$ is usually too difficult

solution : introduce approximate $q_\phi(z|x) \approx p_\theta(z|x) \propto p_\theta(x|z)\,p_\theta(z)$

$$x_n \xrightarrow{q_\phi(z_n|x_n)} z_n \xrightarrow{p_\theta(x_n|z_n)} x_n$$

2. Variational Bayesian Inference

· turn inference into optimization

· assume data X, latent variables Z, and parameters θ

approx. posterior $q(Z) \approx p_\theta(Z|X)$

$$\log p_\theta(X) = \underbrace{KL\left[q(Z) \,\|\, p_\theta(Z|X)\right]}_{\substack{KL(\cdot) \geq 0 \\ KL(\cdot) = 0 \text{ iff } q(Z) = p_\theta(Z|X)}} + \underbrace{\mathcal{L}\left[q(Z),\, p_\theta(X, Z)\right]}_{\text{lower bound on } \log p_\theta(X)}$$

maximize $\mathcal{L}[\cdot]$ w.r.t. q $\to$ find posterior dist. that minimizes $KL[\cdot]$