Probabilistic models for neural data: From single neurons to population dynamics

NEUROBIO 316QC

Jan Drugowitsch

jan_drugowitsch@hms.harvard.edu

Session 8: Variational autoencoders

Today

Q&A about previous session

Paper discussion (~1h)

Variational autoencoders (~25min)

Overview

Autoencoders

Variational Bayesian inference

Variational autoencoders & sequential variant

Overview

Autoencoders

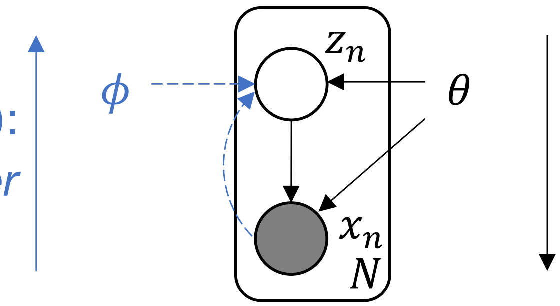
Variational Bayesian inference

Variational autoencoders & sequential variant

Autoencoders

“Code” $\boldsymbol{z}$ generates observations $\boldsymbol{x}$ through $p_{\theta^*}(\boldsymbol{z}) \rightarrow p_{\theta^*}(\boldsymbol{x}|\boldsymbol{z})$ (e.g., neural network)

recognition model $q_\phi(\boldsymbol{z}|\boldsymbol{x})$:
probabilistic *encoder*



generator $p_\theta(\boldsymbol{x}|\boldsymbol{z})$:
probabilistic *decoder*

Usually $\dim(\boldsymbol{z}) \ll \dim(\boldsymbol{x}) \rightarrow$ autoencoders perform non-linear dimensionality reduction

Aim: given $p_\theta(\boldsymbol{z})$ and $p_\theta(\boldsymbol{x}|\boldsymbol{z})$, find θ that maximizes $p_\theta(\boldsymbol{x}_{1:N}) = \prod_n \underbrace{\int p_\theta(\boldsymbol{x}_n|\boldsymbol{z}_n) p_\theta(\boldsymbol{z}_n) \mathrm{d}\boldsymbol{z}_n}_{p_\theta(\boldsymbol{x}_n)}$

Challenges: for sufficiently complex $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ (e.g., network), $p_\theta(\boldsymbol{x})$ intractable

Approach: introduce approximate $q_\phi(\boldsymbol{z}|\boldsymbol{x}) \approx p_\theta(\boldsymbol{z}|\boldsymbol{x}) \propto p_\theta(\boldsymbol{x}|\boldsymbol{z}) p_\theta(\boldsymbol{z})$

$$\boldsymbol{x}_n \xrightarrow{q_\phi(\boldsymbol{z}_n|\boldsymbol{x}_n)} \boldsymbol{z}_n \xrightarrow{p_\theta(\mathbf{x}_\mathrm{n}|\mathbf{z}_\mathrm{n})} \boldsymbol{x}_n$$

New aims: 1. optimize θ to maximize $p_\theta(\boldsymbol{x}_n|\boldsymbol{z}_n)$ (Wake phase)

2. optimize ϕ to align $q_\phi(\boldsymbol{z}_n|\boldsymbol{x}_n) \approx p_\theta(\boldsymbol{z}_n|\boldsymbol{x}_n)$ (Sleep phase)

Variational Bayesian inference

Turning inference into optimization

Assume data $\boldsymbol{X}$, latent variables $\boldsymbol{Z}$, and parameters θ

Approximate posterior $q(\boldsymbol{Z}) \approx p_\theta(\boldsymbol{Z}|\boldsymbol{X})$

$$\underbrace{\log p_\theta(\boldsymbol{X})}_{\text{independent of } q} = \underbrace{\mathrm{KL}[q(\boldsymbol{Z})||p_\theta(\boldsymbol{Z}|\boldsymbol{X})]}_{E_q\left[\log \frac{q(\boldsymbol{Z})}{p_\theta(\boldsymbol{Z}|\boldsymbol{X})}\right] \downarrow} + \underbrace{\mathcal{L}[q(\boldsymbol{Z}), p_\theta(\boldsymbol{X},\boldsymbol{Z})]}_{E_q\left[\log \frac{p_\theta(\boldsymbol{X},\boldsymbol{Z})}{q(\boldsymbol{Z})}\right] \uparrow} \geq \underbrace{\mathcal{L}[q(\boldsymbol{Z}), p_\theta(\boldsymbol{X},\boldsymbol{Z})]}_{\text{"lower bound"}}$$

$$\mathrm{KL}(\cdot) \geq 0$$

$$\mathrm{KL}(\cdot) = 0 \text{ iff } q(\boldsymbol{Z}) = p_\theta(\boldsymbol{Z}|\boldsymbol{X})$$

EM maximize $\mathcal{L}[\cdot]$ wrt. $\theta \rightarrow$ maximize lower bound on $p_\theta(\boldsymbol{X})$

Variational Bayes maximize $\mathcal{L}[\cdot]$ wrt. $q \rightarrow$ find posterior that minimizes $\mathrm{KL}[q(\boldsymbol{Z})||p_\theta(\boldsymbol{Z}|\boldsymbol{X})]$

Approaches assume factorization $q(\boldsymbol{Z}) = \prod_k q_k(\boldsymbol{z}_k)$ & non-parametric optimization

parametrize $q_\phi(\boldsymbol{Z})$ & optimize wrt. ϕ

Overview

Autoencoders

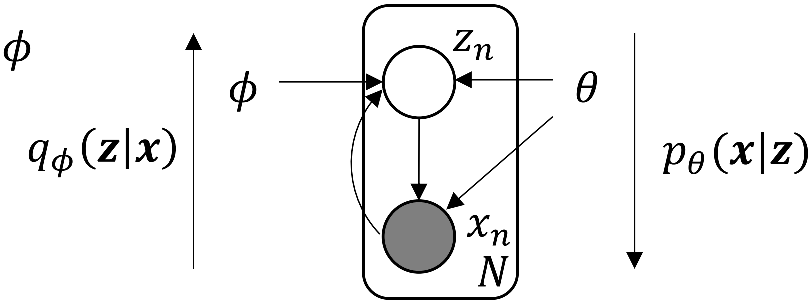
Variational Bayesian inference

Variational autoencoders & sequential variant

Variational autoencoders

Single variational objective optimized wrt. both θ and ϕ

Both p_θ and q_ϕ are (deep) neural networks



$$\mathcal{L}\left(q_\phi(\boldsymbol{Z}|\boldsymbol{X}), p_\theta(\boldsymbol{X},\boldsymbol{Z})\right) = E_{q_\phi}[\log p_\theta(\boldsymbol{X}|\boldsymbol{Z}) + \log p_\theta(\boldsymbol{Z})] - E_{q_\phi}[\log q_\phi(\boldsymbol{Z}|\boldsymbol{X})]$$

$$= \underbrace{E_{q_\phi}[\log p_\theta(\boldsymbol{X}|\boldsymbol{Z})]}_{\text{data likelihood under approx. posterior}} - \underbrace{\mathrm{KL}[q_\phi(\boldsymbol{Z}|\boldsymbol{X})||p_\theta(\boldsymbol{Z})]}_{\text{distance between prior and approx. posterior}}$$

Challenge: low-variance estimator of $\nabla\mathcal{L}(q_\phi, p_\theta)$, involving $E_{q_\phi}[\cdot]$ (usually very noisy)

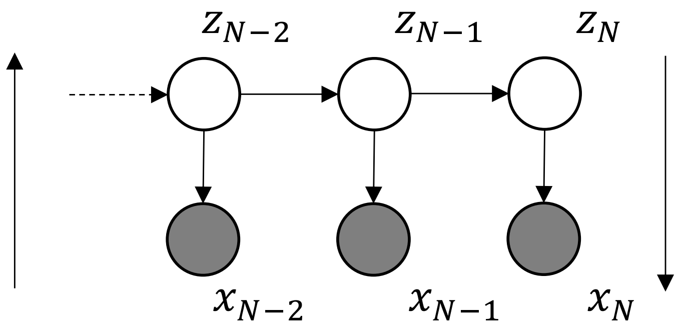
Approach: use "reparametrization trick" that separates noise from gradient

Training (alternate):

1. Draw minibatch from $\boldsymbol{X}$, samples $\boldsymbol{Z} \sim q_\phi(\boldsymbol{Z}|\boldsymbol{X})$ from minibatch
2. Compute gradients, using $\boldsymbol{Z}$ samples to approximate $\nabla_{\theta,\phi}\mathcal{L}(\cdot)$, follow gradient

Variational autoencoders for state space models

recognition model
$q_\phi(\boldsymbol{z}_{1:N}|\boldsymbol{x}_{1:N})$
(smoothing)



generator
$p_\theta(\boldsymbol{x}_{1:N}|\boldsymbol{z}_{1:N}) = \prod_n p_\theta(\boldsymbol{x}_n|\boldsymbol{z}_n)$:

$$\mathcal{L}\left(q_\phi(\boldsymbol{z}_{1:N}|\boldsymbol{x}_{1:N}), p_\theta(\boldsymbol{x}_{1:N}, \boldsymbol{z}_{1:N})\right) = E_{q_\phi}[\log p_\theta(\boldsymbol{x}_{1:N}|\boldsymbol{z}_{1:N})] - \mathrm{KL}[q_\phi(\boldsymbol{z}_{1:N}|\boldsymbol{x}_{1:N})||p_\theta(\boldsymbol{z}_{1:N})]$$

Data likelihood under approx. posterior

$$E_{q_\phi}[\log p_\theta(\boldsymbol{x}_{1:N}|\boldsymbol{z}_{1:N})] = \sum_n E_{q_\phi(\boldsymbol{z}_n)}[\log p_\theta(\boldsymbol{x}_n|\boldsymbol{z}_n)]$$

Distance between approx. posterior and prior

$$\mathrm{KL}\big[q_\phi(\boldsymbol{z}_{1:N}|\boldsymbol{x}_{1:N})||p_\theta(\boldsymbol{z}_{1:N})\big] = \mathrm{KL}\big[q_\phi(\boldsymbol{z}_1|\boldsymbol{x}_{1:N})||p_\theta(\boldsymbol{z}_1)\big] + \sum_{n=2}^{N} \mathrm{KL}\big[q_\phi(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}, \boldsymbol{x}_{1:N})||p_\theta(\boldsymbol{z}_n|\boldsymbol{z}_{n-1})\big]$$

Training (alternate):

1. Draw minibatch from $\boldsymbol{X}$, samples $\boldsymbol{Z} \sim q_\phi(\boldsymbol{Z}|\boldsymbol{X})$ from minibatch
2. Compute gradients, using $\boldsymbol{Z}$ samples to approximate $\nabla_{\theta,\phi}\mathcal{L}(\cdot)$, follow gradient

Overview

Autoencoders

Variational Bayesian inference

Variational autoencoders & sequential variant

Summary

Autoencoders: model $p(\boldsymbol{x})$ by assuming generation $\boldsymbol{z} \rightarrow \boldsymbol{x}$ from lower-d latent state $\boldsymbol{z}$

Tractable approach separates decoder $p_\theta(\boldsymbol{x}|\boldsymbol{z})$ and encoder $p_\phi(\boldsymbol{z}|\boldsymbol{x})$, jointly trains them

Variational inference: inference as optimization by maximizing lower bound on $\log p(\boldsymbol{x})$

Variational autoencoder: use same (variational) objective to jointly train encoder/decoder

Sequential variational autoencoder: variational autoencoder applied to state space models

Until next week

Complete GPFA exercise (see notes for Session 8)

Read paper and prepare presentation (see notes for Session 9)

Next session

Q & A for previous session

Discussing GPFA exercise (~20min)

Paper discussion (~1h)

Course wrap-up (~20min)