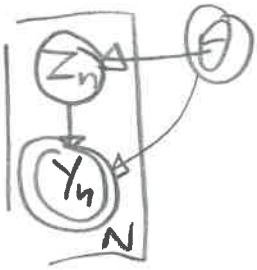
JOHNSON ET AL. "COMPOSING GMs w/ NNs ..." 11/19

1. MOTIVATION: How to model complicated data?

APPROACH: GMM



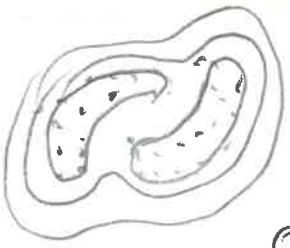
- Let $\theta \triangleq \{\pi, \mu_1, \Sigma_1, \ldots, \mu_K, \Sigma_K\}$
  - π: Dir
  - $\mu_1, \Sigma_1, \ldots, \mu_K, \Sigma_K$: $\sim$ NIW

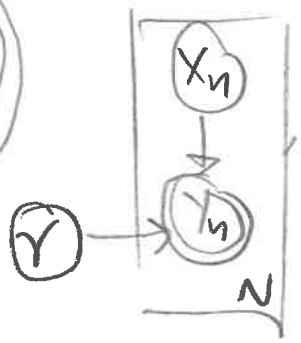
$$Z_n | \pi \sim \text{CAT}(\pi)$$

$$Y_n | z_n, \theta \sim N(\mu_{z_n}, \Sigma_{z_n})$$

✓ INTERPRETABLE / STRUCTURED ✗ RESTRICTIVE

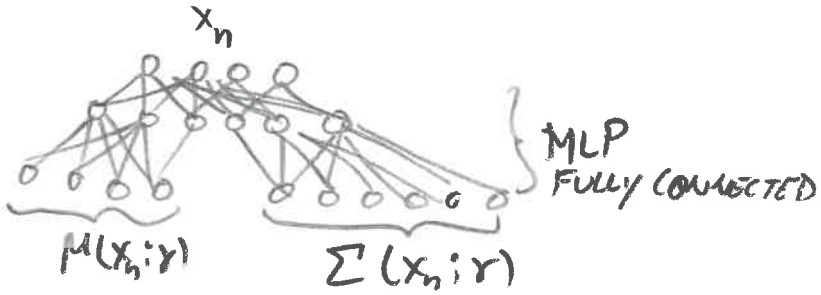
APPROACH: DENSITY NET (VAE)





$$\gamma \sim p(\gamma), \quad X_n \sim N(0, I)$$

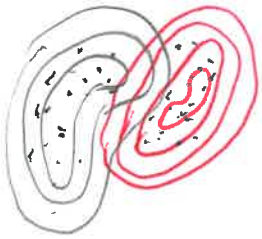
$$Y_n | x_n, \gamma \sim N(\mu(x_n; \gamma), \Sigma(x_n; \gamma))$$

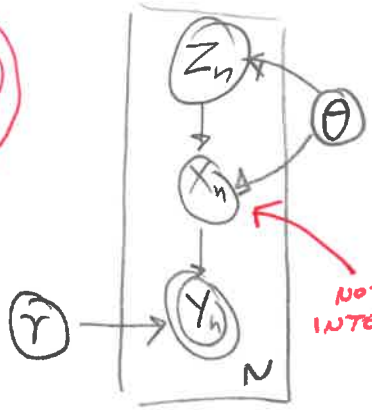


✓ FLEXIBLE

✗ NOT INTERPRETABLE

APPROACH: LATENT GMM (SVAE)





- θ: SAME AS ABOVE

$$\gamma \sim p, \quad X_n \sim N(\mu_{z_n}, \Sigma_{z_n}), \quad Z_n \sim \text{CAT}(\pi) \quad \leftarrow \text{GMM}$$

$$Y_n | x_n, \gamma \sim N(\mu(x_n; \gamma), \Sigma(x_n; \gamma))$$

μ, Σ: MLP

✓ FLEXIBLE ✓ STRUCTURED / INTERPRETABLE

☐ BUT HOW TO DO INFERENCE?

3: VAE RECAP

Y: DATA, X: LATENT, $\{y_n\}_{n=1}^N$: Training Data

$p_\gamma(y|x) \Rightarrow N(\mu(x;\gamma), \Sigma(x;\gamma))$: DECODER

$q_\phi(x|y) \Rightarrow N(\mu(y;\phi), \Sigma(y;\phi))$: ENCODER

MAX. LIK. ESTIMATION

$$\max_\gamma \sum_{n=1}^N \log p_\gamma(y_n) \geq \max_{\gamma,\phi} \sum_{n=1}^N - KL(q_\phi(X|y_i) \,\|\, p_\gamma(x, y_i)) \equiv \mathcal{L}(\phi, \gamma; y_i)$$

- ☐ AVOIDS ISSUES w/ CONJUGACY
- ☐ SGD optimization ☐ LACKS INTERPRETABLE X
- ☐ "Reparameterization Trick" ⇒ Lower variance gradient est.
- ☐ Ammortized inference $q_\phi(X|Y)$ learns mapping for any input y

START HERE

2: STRUCTURED MEAN FIELD VARIATIONAL

☐ Consider Conjugate linear-Gaussian dynamical system.

$x_t | x_{t-1} \sim N(Ax_{t-1}, \Sigma)$, $y_t | x_t \sim N(Bx_t, R)$

$\theta \triangleq \{A, B, \Sigma, R\} \sim$ CONJUGATE PRIOR $p(\theta)$

- VARIATIONAL LOWER BOUND:

$$\log p(Y) \geq \max_{q(\theta) q(x)} - KL(q(\theta) q(x) \,\|\, p(x, y, \theta))$$

$q(x) = q(x_1) \prod_{t=2}^T q(x_t | x_{t-1})$ — GAUSSIAN

- Complete conditionals are Conjugate
- Closed-form coordinate ascent updates

~~[illegible]~~

$$q(\theta) \propto \exp\{\mathbb{E}_{q(x)}[\log p(\theta | x, y)]\}$$

⇒ Same family as $p(\theta)$

⇒ Expfam w/ expected natural params

★ When $p(y|x)$ more general closed-form updates not possible!

4. Structured VAE

- Main Idea: Combine VAE encoder for portions of structured model so closed-form SVI updates possible via conditional conjugacy.
- Assume conjugate pair: $p(\theta)p(x|\theta)$
- Likelihood non-conjugate: e.g. $p(y|x,\gamma) = \mathcal{N}(\mu(x;\gamma), \Sigma(x;\gamma))$ (MLP)

≥ Variational Bound:

$$\log p(y) \geq \max_{q(\theta)q(\gamma)q(x)} -KL\left[q(\theta)q(\gamma)q(x) \,\|\, p(\theta)p(\gamma)p(x|\theta)p(y|x,\gamma)\right]$$

- Parametric variational bound:

$$\log p(y) \geq \max_{\eta_\theta,\eta_\gamma,\eta_x} -KL\left[q_{\eta_\theta}(\theta)\, q_{\eta_\gamma}(\gamma)\, q_{\eta_x}(x) \,\|\, p(\theta)p(\gamma)p(x|\theta)p(y|x,\gamma)\right]$$

$$\equiv \mathcal{L}(\eta_\theta, \eta_\gamma, \eta_x)$$

- One Issue: $p(x|\theta)$ not conjugate to $p(y|x,\gamma)$ so update not closed-form

⇒ Sol'n: Surrogate objective:

$$\hat{\mathcal{L}}(\eta_\theta, \eta_x, \phi) \triangleq -KL\left[q(\theta)q(x) \,\|\, p(\theta)p(x|\theta)\tilde{p}_\phi(y|x)\right]$$

where $\tilde{p}_\phi(y|x) \propto \exp\{\langle r(y;\phi), t_x(x)\rangle\}$ (MLP; soff. stats. of $p(x|\theta)$)

⇒ $p(x|\theta)$ conjugate to $\tilde{p}_\phi(y|x)$

⇒ Closed-form update:

$$\eta_x^*(\eta_\theta, \phi) = \operatorname*{argmin}_{\eta_x} \hat{\mathcal{L}}(\eta_\theta, \eta_x, \phi), \quad q^*(x) \propto \exp\{\langle \eta_x^*(\eta_\theta,\phi), t_x(x)\rangle\}$$

⇒ Closed-form updates of η_θ & η_γ minimizing $\mathcal{L}(\eta_\theta, \eta_\gamma, \eta_x^*(\eta_\theta,\phi))$

$\mathcal{L}_T \equiv \mathcal{L}^{SVAE}(\eta_\theta, \eta_\gamma, \phi)$

ASIDE: CONDITIONALLY CONJUGATE MF UPDATES

- Conjugate pair:

$$p(x|\theta) = \exp\{\eta(\theta)^T T(x) - A(\theta)\}$$

$$p(\theta) \propto \exp\left\{\begin{pmatrix}\tau \\ \nu\end{pmatrix}^T \begin{pmatrix}\eta(\theta) \\ -A(\theta)\end{pmatrix}\right\}$$

- Posterior conjugacy:

$$p(\theta|x) \propto p(x,\theta) \propto \exp\left\{\begin{pmatrix}\tau \\ \nu\end{pmatrix}^T \begin{pmatrix}\eta(\theta) \\ -A(\theta)\end{pmatrix} + \eta(\theta)^T T(x) - A(\theta)\right\}$$

$$= \exp\left\{\begin{pmatrix}\tau + T(x) \\ \nu + 1\end{pmatrix}^T \begin{pmatrix}\eta(\theta) \\ -A(\theta)\end{pmatrix}\right\} \Rightarrow \text{SAME EXPFAM AS } p(\theta)$$

- MF FIXED-POINT:

$$q(\theta) \propto \exp\{\mathbb{E}_{q_x}[\log p(\theta, x)]\}$$

$$= \exp\left\{\begin{pmatrix}\tau + \mathbb{E}_{q_x}[T(x)] \\ \nu + 1\end{pmatrix}^T \begin{pmatrix}\eta(\theta) \\ -A(\theta)\end{pmatrix}\right\} \Rightarrow \text{SAME EXPFAM AS } p(\theta)$$