

Derivations and Equations

1 Deriving the bound

The standard variational formulation is

$$\begin{aligned}\log p(y) &= \int q(z) \log \frac{p(y|z)p(z)}{q(z)} \frac{q(z)}{p(z|y)} dz \\ &= \underbrace{\mathbb{E}_{q(z)} [\log p(y|z)p(z) - \log q(z)]}_{\mathcal{L}_1} + \text{KL}[q(z) \| p(z|y)]\end{aligned}\quad (1)$$

The GP-LVM-CMF variational joint in the augmented latent space is:

$$q(z, f, u, X) = \prod_{i=1}^N \{q(z_i | f(X_i))q(f(X_i) | u, X_i)\} q(u)q(X) \quad (2)$$

which induces an intractable variational marginal

$$q(z) = \int \prod_{i=1}^N \{q(z_i | f(X_i))q(f(X_i) | u, X_i)\} q(u)q(X) df du dX. \quad (3)$$

However we can apply the variational formulation again, noting that $q(z) = \frac{q(z|f,u,X)q(f,u,X)}{q(f,u,X|z)}$,

and integrating 1 wrt the auxiliary variables, to obtain a further lower bound:

$$\log p(y) = \int q(z|f, u, X)q(f, u, X) \log \frac{p(y|z)p(z)r(f, u, X|z)}{q(z|f, u, X)q(f, u, X)} \frac{q(f, u, X|z)}{r(f, u, X|z)} \frac{q(z)}{p(z|y)} df du dX dz$$

$$= \mathbb{E}_{q(z,f,u,X)} [\log p(y|z)p(z)r(f, u, X|z) - \log q(z|f, X)q(f, u, X)] \quad (4)$$

$$+ \mathbb{E}_q(z) [\text{KL}[q(f, u, X|z) \| r(f, u, X|z)]] + \text{KL}[q(z) \| p(z|y)] \quad (5)$$

in which the new auxiliary lower bound, \mathcal{L}_{aux} is given by the expression 4, and where we have introduced the auxiliary distribution $r(f, u, X|z)$ which serves to approximate the variational posterior, $q(f, u, X|z)$, of the auxiliary variables conditioned on the latent variables.

We may re-express \mathcal{L}_{aux} in a way which makes use of the analytical expression for the K-L divergence between two Gaussians, $q(f, u, x)$ and $r(f, u, X|z)$ and, in the case that the prior of the generative model, $p(z)$, is also Gaussian distributed - as is the case for the continuous latent variable MLP model we'll consider first - then the bound contains a second Gaussian KL term:

$$\mathcal{L}_{aux} = \underbrace{\mathbb{E}_{q(z)} [\log p(y|z)]}_A - \underbrace{\mathbb{E}_{q(f,u,X)} [\text{KL}[q(z|f, X) \| p(z)]]}_B - \underbrace{\mathbb{E}_{q(z)} [\text{KL}[q(f, u, X) \| r(f, u, X|z)]]}_C. \quad (6)$$

2 Expressions

2.1 Generative model

$$p(z) = \mathcal{N}(z \mid 0, I) \quad (7)$$

2.1.1 Continuous data \implies Gaussian MLP likelihood

$$\log p(y|z) = \log \mathcal{N}(x \mid \mu, \sigma^2 I) \quad (8)$$

$$\text{where } \mu = W_2 h + b_2 \quad (9)$$

$$\log \sigma^2 = W_3 h + b_3 \quad (10)$$

$$h = \tanh(W_1 z + b_1) \quad (11)$$

Term B in (6) then becomes:

$$-\frac{1}{2} \sum_{i=1}^N \mathbb{E}_{q(f,u,X)} [1 + \log \sigma^2 - f(X_i)^2 - \sigma^2] \quad (12)$$

$$= -\frac{1}{2} \sum_{i=1}^N \mathbb{E}_{q(u)q(X)} [1 + \log \sigma^2 - (K_{f_i u} K_{uu}^{-1} u)^2 - k_{f_i f_i} + K_{f_i u} K_{uu}^{-1} K_{u f_i} - \sigma^2] \quad (13)$$

$$= -\frac{N}{2} (1 + \log \sigma^2 - \sigma^2) + \frac{1}{2} \sum_{i=1}^N \mathbb{E}_{q(X)} [K_{f_i u} K_{uu}^{-1} (\kappa \kappa^\top + K_{uu}) K_{uu}^{-1} K_{u f_i} + K_{f_i f_i} - K_{f_i u} K_{uu}^{-1} K_{u f_i}] \quad (14)$$

$$(15)$$

2.1.2 Discrete data \implies Bernoulli MLP likelihood

$$\log p(y|z) = \sum_{p=1}^P y_p \log \hat{z}_p + (1 - y_p) \log(1 - \hat{z}_p) \quad (16)$$

$$\text{where } \hat{z} = f_\sigma(W_2 h + b_2) \quad (17)$$

$$h = \tanh(W_1 z + b_1). \quad (18)$$

2.2 Variational model

$$q(X) = \prod_{i=1}^N \mathcal{N}(x_{:,n} \mid 0, I_R) = \prod_{n=1}^N \mathcal{N}(x_{n,:} \mid 0, 1) \quad (19)$$

$$q(u) = \prod_{m=1}^M \mathcal{N}(u_{:,m} \mid \kappa_{:,m}, K_{X_u X_u}) \quad (20)$$

$$q(f(X)|u) = \mathcal{N}(f(X) \mid K_{X_f X_u} K_{X_u X_u}^{-1} u, K_{X_f X_f} - K_{X_u X_f} K_{X_u X_u}^{-1} K_{X_u X_f}) \quad (21)$$

$$q(z|f(X)) = \prod_{i=1}^N \mathcal{N}(z_i \mid f(X_i), \sigma^2) \quad (22)$$

I think the variational distributions need to have variational parameters. Also there is a single gaussian process, right? If so $q(u)$ is just one gaussian, not a product of

gaussians

$$q(X) = \prod_{i=1}^N \mathcal{N}(X_i \mid \phi_i, \Phi_i) \quad (23)$$

$$q(u) = \mathcal{N}(u \mid \kappa, K_{X_u X_u}) \quad (24)$$

$$(25)$$

2.3 Auxiliary model

$$r(f, u, X|z) = q(f(X)|u, X)r(u, X|z) \quad (26)$$

$$r(u, X|z) = r(u|z) \prod_i r(X_i|z) = \mathcal{N}(u; v, \Upsilon) \prod_i \mathcal{N}(X_i, \tau_i, \mathcal{T}_i) \quad (27)$$

2.4

$$\mathbb{E}_q(z)\mathcal{L}(z) = \mathbb{E}_q(z, f, u, X)\mathcal{L}(z, f, u, X) \quad (28)$$

$$= \mathbb{E}_{\mathcal{N}(\eta; 0, 1)\mathcal{N}(\xi; 0, 1)\mathcal{N}(\alpha; 0, 1) \prod \mathcal{N}(\beta_i; 0, 1)} \mathcal{L}_1(\eta, \xi, \alpha, \beta) \quad (29)$$

$$\mathcal{L}_1(\eta, \xi, \alpha, \beta) = \mathcal{L}(z, f, u, X) \quad (30)$$

where $z = f + \mathbf{I}\sigma\eta$

where $\Sigma = K_{X_f X_f} - K_{X_f X_u} K_{X_u X_u}^{-1} K_{X_u X_f}$

where $\mu = K_{X_f X_u} K_{X_u X_u}^{-1} u$

where $f = \mu + \text{chol}(\Sigma)\xi$

where $u = v + \text{chol}(\Upsilon)\alpha$

where $X_i = \tau_i + \text{chol}(\mathcal{T}_i)\beta_i$.

Term C in (6) then becomes:

$$\mathbb{E}_{q(z)}[\text{KL}[q(u)||r(u|z)]] = \mathbb{E}_{q(z)}[\text{KL}[\prod_{i=1}^N q(X_i)||\prod_{i=1}^N r(X_i|z)]] \quad (31)$$

$$= \frac{1}{2}\mathbb{E}_{q(z)}[(v - \kappa)^\top \Upsilon^{-1}(v - \kappa) + \text{tr}(\Upsilon^{-1}K_{uu} + \log \Upsilon - \log K_{uu})] - \frac{MQ}{2} \quad (32)$$

$$\frac{1}{2}\sum_{i=1}^N \{\mathbb{E}_{q(z)}[(\phi_i - \tau_i)^\top \mathcal{T}_i^{-1}(\phi_i - \tau_i) + \text{tr}(\mathcal{T}_i^{-1}\Phi_i + \log \mathcal{T}_i - \log \Phi_i)]\} - \frac{NR}{2}. \quad (33)$$