Derivations and Equations

1 Deriving the bound

The standard variational formulation is

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

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

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

which induces an intractable variational marginal

$$q(z) = \int \prod_{i=1}^{N} \{q(z_i|f(X_i))q(f(X_i)|u)\}q(u)q(X)dfdudX. \qquad (3)$$

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:

$$\begin{aligned}\log p(y) &= \int q(z|f,u,X)q(f,u,X) \log \frac{p(y|z)p(z)r(f,u,X|z)}{q(z|f,u,X)q(f,u,X)} \frac{q(f,u,X|z)}{r(f,u,X|z)} \frac{q(z)}{p(z|y)} dfdudXdz \\ &= \mathbb{E}_{q(z,f,u,X)}\big[\log p(y|z)p(z)r(f,u,X|z) - \log q(z|f,X)q(f,u,X)\big] \qquad (4) \\ &\quad + \mathbb{E}_q(z)\big[\mathrm{KL}[q(f,u,X|z)\|r(f,u,X|z)]\big] + \mathrm{KL}\big[q(z)\|p(z|y)\big] \qquad (5)\end{aligned}$$

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:

$$\mathcal{L}_{aux} = \underbrace{\mathbb{E}_{q(z)}\big[\log p(y|z)\big]}_{A} - \underbrace{\mathbb{E}_{q(f,u,X)}\big[\mathrm{KL}[q(z|f, X)\|p(z)]\big]}_{B} - \underbrace{\mathbb{E}_{q(z)}\big[\mathrm{KL}[q(f, u, X)\|r(f, u, X|z)]\big]}_{C}. \tag{6}$$

2 Expressions

2.1 Generative model

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

2.1.1 Continuous data $\Longrightarrow$ Gaussian MLP likelihood

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

$$\text{where } \mu = W_2 h + b_2 \tag{9}$$

$$\log \sigma^2 = W_3 h + b_3 \tag{10}$$

$$h = \tanh(W_1 z + b_1) \tag{11}$$

Term B in (6) then becomes:

$$
\begin{aligned}
& -\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}_{q(f,u,X)}\big[1+\log\sigma^2 - f(X_i)^2 - \sigma^2\big] && (12)\\
&= -\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}_{q(u)q(X)}\big[1+\log\sigma^2 - (K_{f_iu}K_{uu}^{-1}u)^2 - k_{f_if_i} + K_{f_iu}K_{uu}^{-1}K_{uf_i} - \sigma^2\big] && (13)\\
&= -\frac{N}{2}(1+\log\sigma^2-\sigma^2) + \frac{1}{2}\sum_{i=1}^{N}\mathbb{E}_{q(X)}\big[K_{f_iu}K_{uu}^{-1}\big(\kappa\kappa^\top + K_{uu}\big)K_{uu}^{-1}K_{uf_i} + K_{f_if_i} - K_{f_iu}K_{uu}^{-1}K_{uf_i}\big] && (14)\\
& && (15)
\end{aligned}
$$

2.1.2 Discrete data $\implies$ Bernoulli MLP likelihood

$$
\begin{aligned}
\log p(y|z) &= \sum_{p=1}^{P} y_p \log \hat{z}_p + (1-y_p)\log(1-\hat{z}_p) && (16)\\
\text{where } \hat{z} &= f_\sigma(W_2 h + b2) && (17)\\
h &= \tanh(W_1 z + b_1). && (18)
\end{aligned}
$$

2.2 Variational model

$$
\begin{aligned}
q(X) &= \prod_{i=1}^{N}\mathcal{N}(x_{:,n}\mid 0, I_R) = \prod_{n=1}^{N}\mathcal{N}(x_{n,:}\mid 0,1) && (19)\\
q(u) &= \prod_{m=1}^{M}\mathcal{N}(u_{:,m}\mid \kappa_{:,m}, K_{X_uX_u}) && (20)\\
q(f(X)|u) &= \mathcal{N}(f(X)\mid K_{X_fX_u}K_{X_uX_u}^{-1}u, K_{X_fX_f} - K_{X_uX_f}K_{X_uX_u}^{-1}K_{X_uX_f}) && (21)\\
q(z|f(X)) &= \prod_{i=1}^{N}\mathcal{N}(z_i \mid f(X_i), \sigma^2) && (22)
\end{aligned}
$$

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

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

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

$$\tag{25}$$

2.3 Auxiliary model

$$r(f, u, X|z) = q(f(X)|u, X) r(u, X|z) \tag{26}$$

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

2.4

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

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

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

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

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

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

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

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

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

Term C in (6) then becomes:

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

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

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