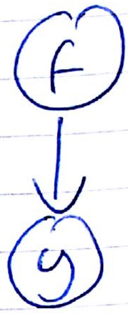
Danmarks Tekniske Universitet
Technical University of Denmark

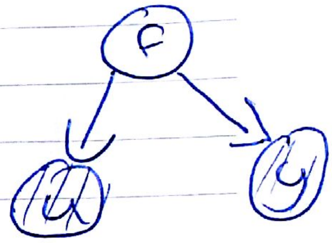
DTU

Thing One

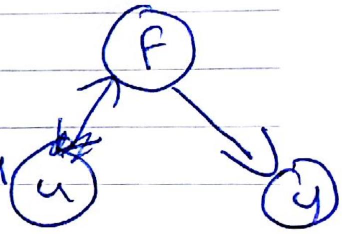
$$\log p(y) = \log \int p(y|f)\, p(f)\, df$$



$$\log p(y) = \log \int p(y|f)\, p(u|f)\, p(f)\, df\, du$$



$$\log p(y) = \log \int p(y|f)\, p(f|u)\, df\, p(u)\, du$$



$$\log p(y) \geq \int p(f|u) \log p(y|f)\, df\, p(u)\, du$$

If $p(y_i|f_i)$ factorizes $p(y|f) = \prod_{i=1}^{n} p(y_i|f_i)$

$$\log p(y) \geq \cancel{\sum_i} \log \int \prod_{i=1}^{n} \exp\left(\langle \log p(y_i|f_i) \rangle_{p(f_i|u)}\right) p(u)\, du$$

If $p(y_i|f_i)$ is Gaussian

Thing One

$$\log p(y) \geq \log \int \prod_{i=1}^{n} \mathcal{N}(y_i \mid \langle f_i \rangle_{f_i \mid u}, \sigma^2)$$

$$\times \exp\left(- \sum \frac{\mathrm{var}(f_i \mid u)}{2\sigma^2} \right)$$

$$\times \; p(u) \, du$$

Joint Gaussian over f & u

$$p(f) = \mathcal{N}(0, K)$$

$$p(u) = \mathcal{N}(0, C)$$

$$\cancel{\begin{bmatrix} f \\ u \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K & W \\ W & C \end{bmatrix} \right)}$$

$$\begin{bmatrix} f \\ u \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K & V \\ V^T & C \end{bmatrix} \right)$$

Danmarks Tekniske Universitet
Technical University of Denmark

DTU

Thing One

$$\Rightarrow p(\vec{f}\,|\,\vec{u}) = \mathcal{N}(\vec{f}\,|\,W C^{-1}\vec{u},\ K - W C^{-1} W^T)$$

must be ~~pos~~ non negative definite

$$\langle \vec{f} \rangle = V C^{-1} \vec{u}$$

$$\mathrm{var}(\vec{f}) = \mathrm{diag}(K - V C^{-1} V^T)$$

$$\log p(y) \geq \log \mathcal{N}(y\,|\,0, \sigma^2 I + V C^{-1} V^T) - \frac{1}{2}\mathrm{Tr}(K - V C^{-1} V^T)$$

If $W = V C^{-1/2}$ then

$$\log p(y) \geq \log \mathcal{N}(y\,|\,0, W W^T + \sigma^2 I) - \frac{1}{2}\mathrm{Tr}(K - W W^T)$$

Thing One

Special Cases ① Multivariate $y \Rightarrow Y \in \mathbb{R}^{n \times d}$ (or $\mathbb{R}^{d \times n}$)

K has maximum likelihood soln

$K = YY^T$ and solution for

W is the principal subspace of K

PCA (if noise non diagonal

FACTOR Analysis)

Special case ② K is defined as a Kernel, but data as yet unseen.

Integrate over y to mardinalize &

Minimize $Tr(K - WW^T)$

W is still principal subspace of K

Thing One

General Case 3

Y is known, K is constrained (e.g. Kernel or inverse of Graph Laplacian)

Now U is neither subspace of K nor is it subspace of yy^T.

Some form of compromise

$$\log|\sigma^2 I + WW^T| + \mathrm{Tr}\left(YY^T(I\sigma^2 + WW^T)^{-1}\right) + \mathrm{Tr}(K - WW^T) \quad \text{st} \quad K - WW^T \succcurlyeq 0$$

Thing Two

$$\log p(y) = \log \int p(y|f)\, p(f)\, df$$

$$p(y|f) = \prod_{i=1}^{n} p(y_i|f_i) \qquad \text{assume } q(f) = \prod_{i=1}^{n} q(f_i)$$

$$\text{assume } q(f_i) = \frac{1}{Z_i} \frac{p(y_i|f_i)}{p(u_i|f_i)} p(f_i|\vec{u})$$

$$p(\vec{u}|\vec{f}) = \prod_{i=1}^{n} p(u_i|f_i) \longrightarrow$$

where u are fake observations st $p(f_i|\vec{u})$ is tractable (for GPs Gaussian)

Aside:

$$p(f_i|\vec{u}) = \prod_{j=1}^{n} \frac{p(u_j|f_j)\, p(f)}{p(\vec{u})} \qquad \frac{p(f_i|\vec{u})}{p(u_i|f_i)} = \prod_{j \neq i} \frac{p(u_j|f_j)\, p(f)}{p(\vec{u})}$$

$$= \frac{p(f_i|\vec{u}_{\setminus i})\, p(\vec{u}_i)}{p(\vec{u})}$$

$$q(f_i) = \frac{1}{Z_i} p(y_i|f_i)\, \frac{p(f_i|\vec{u}_{\setminus i})\, p(\vec{u}_{\setminus i})}{p(\vec{u})}$$

Thing Two

$$q(f_i) = \frac{1}{Z_i} p(y_i | \vec{u}_{\setminus i})\, p(f_i | \vec{u}_{\setminus i}, y_i)\, \frac{p(\vec{u}_{\setminus i})}{p(\vec{u})}$$

$$\Rightarrow Z_i = \frac{p(y_i | u_{\setminus i})\, p(\vec{u}_{\setminus i})}{p(\vec{u})} = \frac{p(y_i | u_{\setminus i})}{p(u_i | u_{\setminus i})}$$

since it must be that

~~if~~ $q(f_i) = p(f_i | u_{\setminus i}, y_i)$

$$\log p(y) \geq \int \prod_{i=1}^{n} q(f_i) \log \frac{\prod_{i=1} p(y_i | f_i)\, p(f)}{\prod_{i=1}^{n} q(f_i)}$$

$$\mathcal{L}_1 = \int \prod_{i=1}^{n} q(f_i) \log \frac{\prod_{i=1}^{n} Z_i\, \cancel{p(y_i|f_i)}\, p(u_i|f_i)\, p(f)}{\prod_{i=1}^{n} \cancel{p(y_i|f_i)}\, p(f_i | \vec{u})}$$

$$= \sum_{i=1}^{n} \log Z_i + \int \prod_{i=1}^{n} q(f_i) \log \frac{\prod_{i=1} p(y_i | f_i)\, p(f)}{\prod_{i=1}^{n} p(f_i | \vec{u})}$$

Danmarks Tekniske Universitet
Technical University of Denmark

DTU

Thing Two

$$\mathcal{L}_r = \sum_{i=1}^{n} \log Z_i + \log p(\vec{u}) + \int \prod_{i=1}^{n} q(f_i) \log \frac{p(\vec{F}|\vec{u})}{\prod_i p(f_i|\vec{u})} dF$$

$$\log p(y) = \mathcal{L} + KL\left(\prod_{i=1}^{n} q(f_i) \,\|\, p(\vec{F}|\vec{y})\right)$$

$$= \underbrace{\sum_{i=1}^{n} \log Z_i + \log p(\vec{u})}_{\text{EP approximation}} + \int \prod_{i=1}^{n} q(f_i) \log \frac{p(\vec{F}|\vec{u})}{p(\vec{F}|\vec{y})} dF$$

$$+ \underbrace{\sum_{i=1}^{n} KL\left(q(f_i) \,\|\, p(f_i|\vec{u})\right)}_{\text{'EP KL divergence'}}$$

expected dissimilarity between $p(\vec{F}|\vec{u})$ and true posterior

EP approximation

Thing Two

EP objective is:

$$E = \sum_{i=1}^{n} \log z_i + \log p(\vec{u})$$

But $z_i = \dfrac{p(y_i|\vec{u}_{\setminus i})}{p(y_i|\vec{u}_{\setminus i})} = \dfrac{p(y_i|\vec{u}_{\setminus i})\,p(\vec{u}_{\setminus i})}{p(\vec{u})}$

So $E = \sum_{i=1}^{n} \log p(y_i|\vec{u}_{\setminus i})\,p(\vec{u}_{\setminus i})$

And TVB objective (a rigid lower bound) is

$$\mathcal{L} = E + \underbrace{\int \prod_{i=1}^{n} q(f_i) \log \frac{p(f|\vec{u})}{\prod_{i=1}^{n} p(f_i|\vec{u})} \, du}$$

$\boxed{\text{note this term is zero if } p(f|\vec{u}) \text{ factorizes.}}$

if moments of $q(f_i)$ are matched to $p(f_i|\vec{u})$

this term is negative for GP case