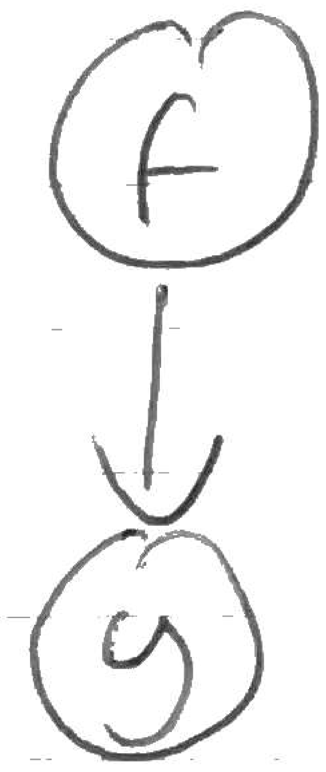
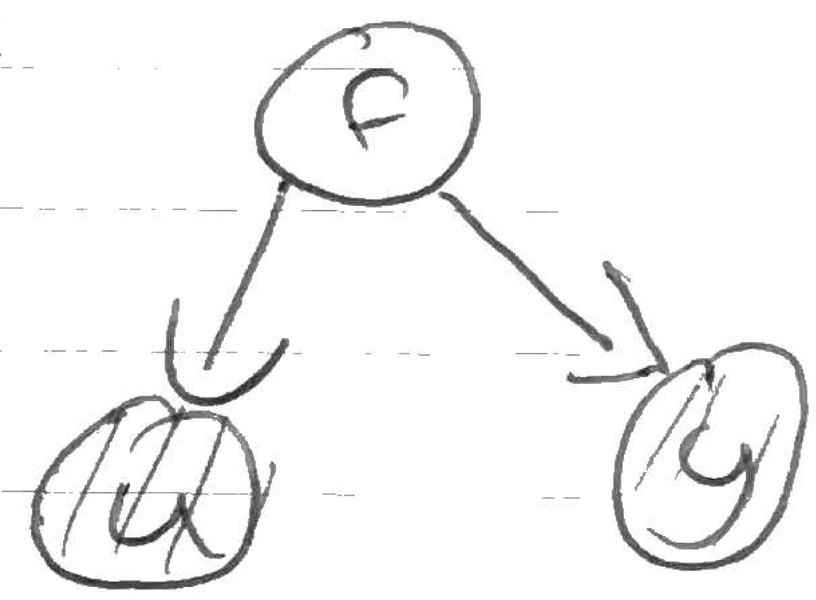
Thing One

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

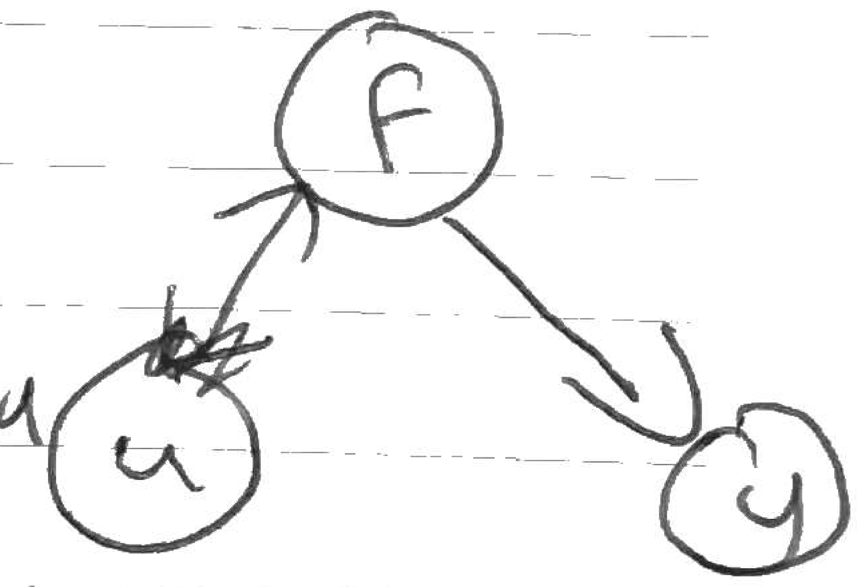


$$\log p(y) = \log \int p(y|f)\, p(u|f)\, p(f)\, df\, du$$



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

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



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

Danmarks Tekniske Universitet
Technical University of Denmark
DTU

Thing One

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

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

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

Joint Gaussian over f & u

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

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

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

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

Thing One

$$\Rightarrow \; p(\vec{f} \mid \vec{u}) = \mathcal{N}(\vec{f} \mid VC^{-1}\vec{u},\; K - VC^{-1}V^{T})$$

must be ~~pos~~ non negative definite

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

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

$$\log p(y) \geq \log \mathcal{N}(y \mid 0, \sigma^2 I + VC^{-1}V^{T})$$

$$- \tfrac{1}{2}\mathrm{Tr}(K - VC^{-1}V^{T})$$

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

$$\log p(y) \geq \log \mathcal{N}(y \mid 0, WW^{T} + \sigma^2 I)$$

$$- \tfrac{1}{2}\mathrm{Tr}(K - WW^{T})$$

(4)

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 marginalize &

Minimize $\mathrm{Tr}(K - WW^T)$

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

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