Toolboxes on Component Analysis

Matlab Toolbox for Dimensionality Reduction

http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html

Matlab 2013b has PPCA implemented

http://www.mathworks.co.uk/help/stats/ppca.html

Toolboxes on Component Analysis

ICA toolboxes for image and signal processing:

http://www.bsp.brain.riken.jp/ICALAB/

ICA for EEG Analysis:

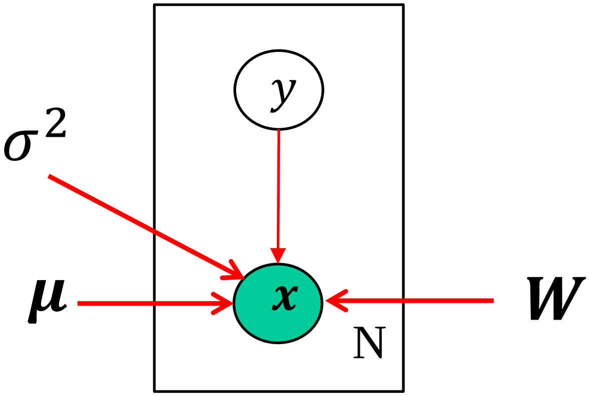
http://mialab.mrn.org/software/

FastICA

http://research.ics.aalto.fi/ica/fastica/

PPCA applications

In PPCA we have:



$$p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{W},\boldsymbol{\mu},\sigma^2) = \mathrm{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{y}+\boldsymbol{\mu},\sigma^2)$$

$$p(\boldsymbol{y}) = \mathrm{N}(\boldsymbol{y}|\boldsymbol{0},\boldsymbol{I})$$

$\Rightarrow$

$$p(\mathbf{x}|\theta) = N(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{D})$$

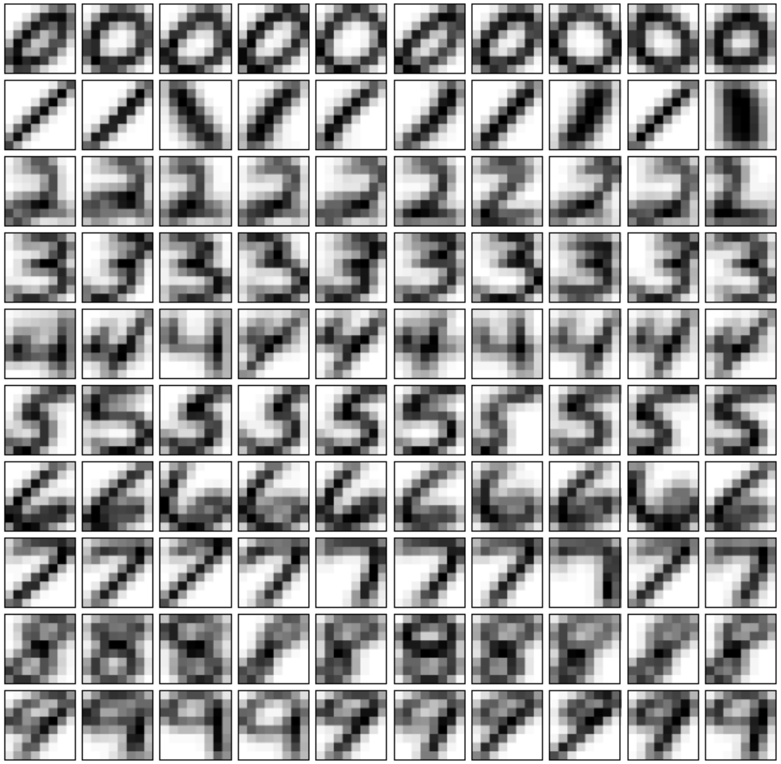
where $\boldsymbol{D} = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2\boldsymbol{I}$

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{W},\boldsymbol{\mu},\sigma^2) = N(\boldsymbol{y}|\boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x}-\boldsymbol{\mu}),\sigma^2\boldsymbol{M}^{-1})$$

where $\boldsymbol{M} = \sigma^2\boldsymbol{I} + \boldsymbol{W}^T\boldsymbol{W}$

PPCA classification

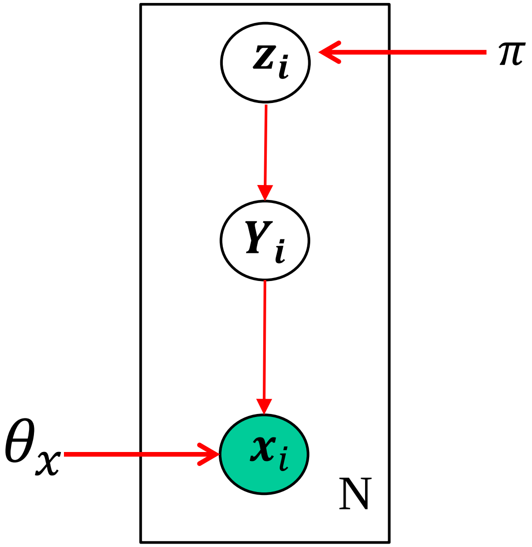
Learn a PPCA per class (9 PPCA's).



Classify to the class i with larger probability

$$p(\mathbf{x}|\theta) = N(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{D}_i)$$

Mixtures of PPCA clustering



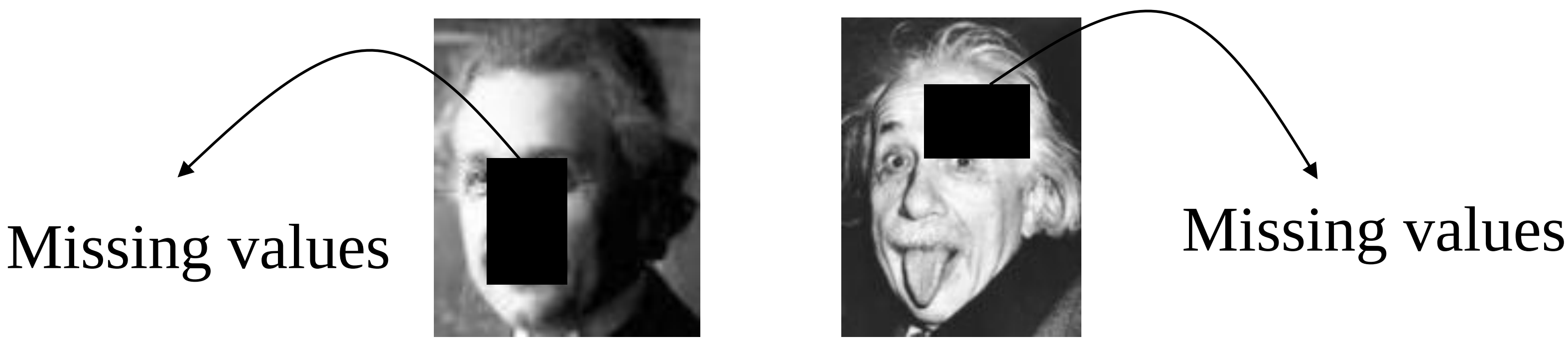
$$p(\mathbf{x}|\theta) = \sum_{i=1}^{C} \pi_i N(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{D}_i)$$

Assign to the cluster with the highest probability

$$N(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{D}_i)$$

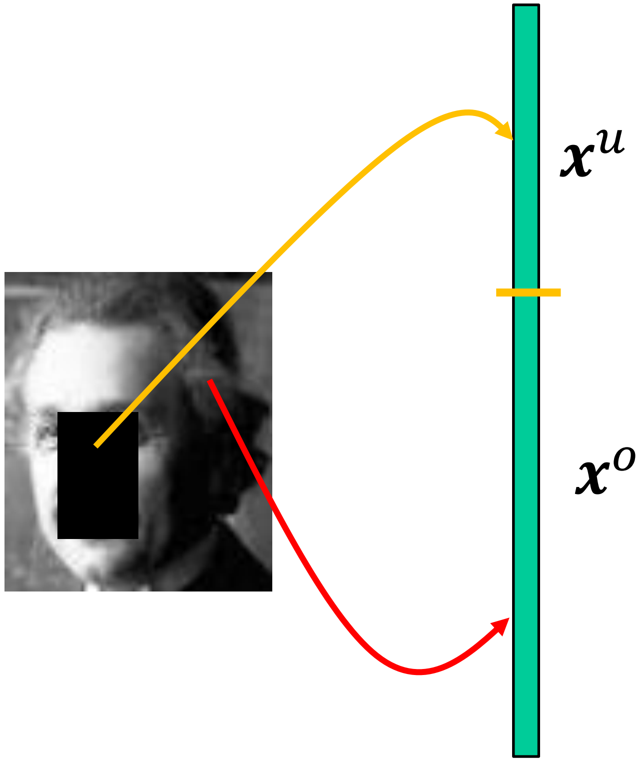
PPCA with missing data

- In many cases we have missing values in the dataset.



PPCA with missing data

$\mathbf{x}$



p($\boldsymbol{x}$)=p($\boldsymbol{x}^u, \boldsymbol{x}^o$) $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}^o \\ \boldsymbol{x}^u \end{bmatrix}$

unobserved observed

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^o \\ \boldsymbol{\mu}^u \end{bmatrix} \quad \boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{oo} & \boldsymbol{D}_{ou} \\ \boldsymbol{D}_{uo} & \boldsymbol{D}_{uu} \end{bmatrix}$$

$$p(\boldsymbol{x}^u | \boldsymbol{x}^o) = N(\boldsymbol{x}^u | \boldsymbol{\mu}^u + \boldsymbol{D}_{uo} \boldsymbol{D}_{oo}{}^{-1} (\boldsymbol{x}^o - \boldsymbol{\mu}^o), \boldsymbol{D}_{uu} - \boldsymbol{D}_{uo} \boldsymbol{D}_{oo}{}^{-1} \boldsymbol{D}_{ou})$$

PPCA with missing data

Exercise 1:
Prove that if

$$p(\mathbf{x}|\theta) = \mathrm{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{D})$$

then $p(\boldsymbol{x}^u|\boldsymbol{x}^o)$

$$= N(\boldsymbol{x}^u|\boldsymbol{\mu}^u + \boldsymbol{D}_{uo}\boldsymbol{D}_{oo}{}^{-1}(\boldsymbol{x}^o - \boldsymbol{\mu}^o), \boldsymbol{D}_{uu} - \boldsymbol{D}_{uo}\boldsymbol{D}_{oo}{}^{-1}\boldsymbol{D}_{ou})$$

Hint: $\mathrm{p}(\boldsymbol{x}^u|\boldsymbol{x}^o) = \frac{p(\boldsymbol{x})}{p(\boldsymbol{x}^o)}$

$$\mathrm{p}(\boldsymbol{x}^o) = \int_{\boldsymbol{x}^u} p(\boldsymbol{x}) d\boldsymbol{x}^u$$

and using

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix}$$

PPCA with missing data

$$p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o) = \mathrm{N}(\boldsymbol{x}_i|\boldsymbol{z}_i, \boldsymbol{Q})$$

$$\boldsymbol{z}_i = \begin{bmatrix} \boldsymbol{x}_i{}^o \\ \boldsymbol{\mu}^u + \boldsymbol{D}_{uo}\boldsymbol{D}_{oo}{}^{-1}(\boldsymbol{x}_i{}^o - \boldsymbol{\mu}^o) \end{bmatrix}$$

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{D}_{uu} - \boldsymbol{D}_{uo}\boldsymbol{D}_{oo}{}^{-1}\boldsymbol{D}_{ou} \end{bmatrix}$$

$$\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{x}_i\} = \int_{\boldsymbol{x}_i} \boldsymbol{x}_i p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o) d\boldsymbol{x}_i = \boldsymbol{z}_i$$

$$\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{x}_i\boldsymbol{x}_i{}^T\} = \boldsymbol{Q} + \boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{x}_i\}\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{x}_i\}^T$$

$$= \boldsymbol{Q} + \boldsymbol{z}_i\boldsymbol{z}_i{}^T$$

Expectations with missing data

Also for whatever $\boldsymbol{a}$ that does not depend on $\boldsymbol{x}_i$

$$\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^{o})}\{(\boldsymbol{x}_i - \boldsymbol{a})(\boldsymbol{x}_i - \boldsymbol{a})^T\} = \boldsymbol{Q} + (\boldsymbol{z}_i - \boldsymbol{a})(\boldsymbol{z}_i - \boldsymbol{a})^T$$

Exercise 2: Prove the above

Hint: Just expand !!

$$\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^{o})}\{(\boldsymbol{x}_i - \boldsymbol{a})(\boldsymbol{x}_i - \boldsymbol{a})^T\}$$

EM with missing data

$$\ln p(\boldsymbol{X},\boldsymbol{Y}|\theta) = -\frac{NF}{2}\ln 2\pi - NF\ln\sigma - N\ln 2\pi$$
$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[\mathrm{tr}((\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T) - 2\mathrm{tr}(\boldsymbol{y_i}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{W})$$
$$+\mathrm{tr}(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{y}_i\boldsymbol{y}_i{}^T)] - \sum_{i=1}^{N}\mathrm{tr}(\boldsymbol{y}_i\boldsymbol{y}_i{}^T)$$

In original PPCA we have to compute the posterior expectations with regards to $\boldsymbol{Y}|\boldsymbol{X}$ (since our $\boldsymbol{X}$ are fully observed)

Now we have to compute the expectations with regards to $\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{X}^o$

EM with missing data

$$\boldsymbol{E}_{p(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{X}^o)}\{\ln p(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{\vartheta})\} =$$

$$-\sum_{i=1}^{N}\frac{1}{2\sigma^2}\Big[\mathrm{tr}\Big[\boldsymbol{E}_{p(\boldsymbol{x}_i,\boldsymbol{y}_i|\boldsymbol{x}_i{}^o)}\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}\Big]$$

$$-2\mathrm{tr}\Big[\boldsymbol{E}_{p(\boldsymbol{x}_i,\boldsymbol{y}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{y}_i(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}\boldsymbol{W}\Big] + \mathrm{tr}[\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{E}\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}]$$

$$+\frac{1}{2}\mathrm{tr}[\boldsymbol{E}\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}] - \frac{NF}{2}\ln 2\pi - NF\ln\sigma - N\ln 2\pi$$

Let’s see how this is done in detail

EM with missing data

$$p(\boldsymbol{x}_i, \boldsymbol{y}_i | \boldsymbol{x}_i{}^o) = p(\boldsymbol{y}_i | \boldsymbol{x}_i \quad) p(\boldsymbol{x}_i | \boldsymbol{x}_i{}^o)$$

$$\boldsymbol{E}_{p(\boldsymbol{x}_i, \boldsymbol{y}_i | \boldsymbol{x}_i{}^o)} \{ (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T \}$$

$$= \int_{\boldsymbol{x}_i} \int_{\boldsymbol{y}_i} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{x}_i, \boldsymbol{y}_i | \boldsymbol{x}_i{}^o) d\boldsymbol{x}_i d\boldsymbol{y}_i$$

$$= \int_{\boldsymbol{x}_i} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{x}_i | \boldsymbol{x}_i{}^o) \underbrace{\int_{\boldsymbol{y}_i} p(\boldsymbol{y}_i | \boldsymbol{x}_i \quad) d\boldsymbol{y}_i}_{1} d\boldsymbol{x}_i$$

$$= \int_{\boldsymbol{x}_i} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{x}_i | \boldsymbol{x}_i{}^o) \, d\boldsymbol{x}_i$$

$$= \boldsymbol{Q} + (\boldsymbol{z}_i - \boldsymbol{\mu})(\boldsymbol{z}_i - \boldsymbol{\mu})^T$$

EM with missing data

$$\int_{\boldsymbol{y}_i} \boldsymbol{y}_i p(\boldsymbol{y}_i|\boldsymbol{x}_i) d\boldsymbol{y}_i = \boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x_i} - \boldsymbol{\mu})$$

$$\begin{aligned}
&\boldsymbol{E}_{p(\boldsymbol{y}_i,\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{y}_i(\boldsymbol{x}_i - \boldsymbol{\mu})^T\} \\
&= \int_{\boldsymbol{x}_i}\int_{\boldsymbol{y}_i} \boldsymbol{y}_i(\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{y}_i, \boldsymbol{x}_i|\boldsymbol{x}_i{}^o) d\boldsymbol{y}_i d\boldsymbol{x}_i \\
&= \int_{\boldsymbol{x}_i}\left[\int_{\boldsymbol{y}_i} \boldsymbol{y}_i p(\boldsymbol{y}_i|\boldsymbol{x}_i) d\boldsymbol{y}_i\right](\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o) d\boldsymbol{x}_i \\
&= \int_{\boldsymbol{x}_i} \boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o) d\boldsymbol{x}_i \\
&= \boldsymbol{M}^{-1}\boldsymbol{W}^T \int_{\boldsymbol{x}_i} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o) d\boldsymbol{x}_i \\
&= \boldsymbol{M}^{-1}\boldsymbol{W}^T \boldsymbol{E}\{(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\}
\end{aligned}$$

EM with missing data

$$\boldsymbol{E}_{p(\boldsymbol{y}_i,\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\} = \boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\left\{\boldsymbol{E}_{p(\boldsymbol{y}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}\right\}$$

$$= \boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{\sigma}^2\boldsymbol{M}^{-1} + \boldsymbol{E}\{\boldsymbol{y}_i\}\boldsymbol{E}\{\boldsymbol{y}_i\}^{\boldsymbol{T}}\}$$

$$= \boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\left\{\boldsymbol{\sigma}^2\boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{W}^T(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\boldsymbol{W}\boldsymbol{M}^{-\mathbf{1}}\right\}$$

$$= \boldsymbol{\sigma}^2\boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{W}^T\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\}\boldsymbol{W}\boldsymbol{M}^{-\mathbf{1}}$$

Summary

Expectation Step

$$\boldsymbol{E}_{p(\boldsymbol{x}_i,\boldsymbol{y}_i|\boldsymbol{x}_i{}^o)}\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}=\boldsymbol{Q}+(\boldsymbol{z}_i-\boldsymbol{\mu})(\boldsymbol{z}_i-\boldsymbol{\mu})^T$$

$$\boldsymbol{E}_{p(\boldsymbol{y}_i,\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{y}_i(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}=\boldsymbol{M}^{-1}\boldsymbol{W}^T\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}$$

$$\boldsymbol{E}_{p(\boldsymbol{y}_i,\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}$$
$$=\boldsymbol{\sigma}^2\boldsymbol{M}^{-1}+\boldsymbol{M}^{-1}\boldsymbol{W}^T\boldsymbol{E}_{p(\boldsymbol{x}_i|\boldsymbol{x}_i{}^o)}\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}\boldsymbol{W}\boldsymbol{M}^{-\mathbf{1}}$$

EM with missing data

Maximization step

$$\frac{\partial E\{\boldsymbol{\mu}\}}{\partial \boldsymbol{\mu}} = \mathbf{0} \Rightarrow \boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_i$$

$$\frac{\partial E\{\boldsymbol{W}\}}{\partial \boldsymbol{W}} = 0 \Rightarrow \sum_{i=1}^{N}\left(-\frac{1}{\sigma^2}E\{(\boldsymbol{x}_i - \boldsymbol{\mu})\boldsymbol{y}_i{}^T\} + \frac{2}{2\sigma^2}\boldsymbol{W}E\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}\right) = 0$$

$$\Rightarrow \boldsymbol{W} = \left[\sum_{i=1}^{N} E\{(\boldsymbol{x}_i - \boldsymbol{\mu})\boldsymbol{y}_i{}^T\}\right]\left[\sum_{i=1}^{N} E\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}\right]^{-1}$$

EM with missing data

$$\frac{\partial \boldsymbol{E}\{\sigma^2\}}{\partial \sigma} = 0 \Rightarrow$$

$$\sigma^{\mathbf{2}}$$

$$= \frac{1}{NF}\sum_{i=1}^{N}\{\mathrm{tr}[\boldsymbol{E}\{(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\}] - 2\mathrm{tr}[\boldsymbol{E}\{\boldsymbol{y}_i(\boldsymbol{x}_i - \boldsymbol{\mu})^T\}\boldsymbol{W}] +$$

$$+\mathrm{tr}[\boldsymbol{E}\{\mathbf{y}_i\mathbf{y}_i{}^T\}\boldsymbol{W}^T\boldsymbol{W}]\}$$

finally $\boldsymbol{D} = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2\boldsymbol{I}$

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{oo} & \boldsymbol{D}_{ou} \\ \boldsymbol{D}_{uo} & \boldsymbol{D}_{uu} \end{bmatrix}$$

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{D}_{uu} - \boldsymbol{D}_{uo}\boldsymbol{D}_{oo}{}^{-1}\boldsymbol{D}_{ou} \end{bmatrix}$$

Summary

- In PPCA the latent features are $\boldsymbol{y}_i$ hence we need to compute the expectations $E\{\boldsymbol{y}_i\}$ and $E\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}$ (with regards to $\boldsymbol{Y}|\boldsymbol{X}$

- In PPCA with missing data we assume that part of $\boldsymbol{x}_i$ is observed and part is missing (hidden). Hence we have to compute expectations $E\{(\boldsymbol{x}_i-\boldsymbol{\mu})\boldsymbol{y}_i{}^T\}, E\{\boldsymbol{y}_i\boldsymbol{y}_i{}^T\}, E\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\}$ with regards to $\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{X}^{\boldsymbol{o}}$