Análisis de factores

Lección 20

Dr. Pablo Alvarado Moya

CE5506 Introducción al reconocimiento de patrones
Área de Ingeniería en Computadores
Tecnológico de Costa Rica

II Semestre, 2019

Contenido

El caso de menos datos que dimensiones

- Hasta ahora hemos supuesto suficientes datos m para estimar n parámetros en mezcla de gaussianas.
- No tendremos ningún problema si $m \gg n$.
- En escenarios donde $n \gg m$ es incluso complicado modelar los datos con un solo gaussiano, así que una mezcla de gaussianos es una tarea imposible
- El problema principal es que el espacio engendrado por m datos tendrá a lo sumo $\min(m,n)$ dimensiones.

Singularidad de covarianza

- Si estimamos los parámetros de máxima verosimilitud

$$\underline{\boldsymbol{\mu}} = \frac{1}{m}\sum_{i=1}^{m}\underline{\mathbf{x}}^{(i)} \qquad \boldsymbol{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}(\underline{\mathbf{x}}^{(i)} - \underline{\boldsymbol{\mu}})(\underline{\mathbf{x}}^{(i)} - \underline{\boldsymbol{\mu}})^T$$

  con $n \gg m$ la matriz $\boldsymbol{\Sigma}$ resulta singular, lo que imposibilita la evaluación la densidad gaussiana.
- Este caso $n \gg m$ **confina** los datos a un subespacio de m dimensiones, lo que en principio hace improbable que un dato en n dimensiones *viva* en ese subespacio.
- Podemos imponer restricciones a $\boldsymbol{\Sigma}$ para evitar singularidad.
- Usar el modelo de análisis de factores ofrecerá una alternativa.

Restringiendo la matriz de covarianza (1)

- Podemos imponer alineamiento con los ejes.
- Esto implica que $\mathbf{\Sigma}$ es diagonal con

$$\Sigma_{jj} = \frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} - \mu_j)$$

- Si alguno de los elementos en la diagonal es cero, tampoco sirve esta matriz
- Una restricción más fuerte es definir $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ con

$$\sigma^2 = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}(x_j^{(i)} - \mu_j)^2$$

cuyas superficies equiprobables corresponden a esferas.

Restringiendo la matriz de covarianza (2)

- $\boldsymbol{\Sigma}$ completa requiere $m \geq n+1$ para una estimación (posiblemente) no singular.
- Con restricciones requerimos $m \geq 2$.

Distribuciones gaussianas multivariadas conjuntas

- Supongamos que tenemos una variable aleatoria vectorial

$$\underline{\mathbf{x}} = \begin{bmatrix} \underline{\mathbf{x}}_1 \\ \underline{\mathbf{x}}_2 \end{bmatrix}$$

con $\underline{\mathbf{x}}_1 \in \mathbb{R}^r$, $\underline{\mathbf{x}}_2 \in \mathbb{R}^s$, and $\underline{\mathbf{x}} \in \mathbb{R}^{r+s}$

- Supongamos que $\underline{\mathbf{x}} \sim \mathcal{N}(\underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$ con

$$\underline{\boldsymbol{\mu}} = \begin{bmatrix} \underline{\boldsymbol{\mu}}_1 \\ \underline{\boldsymbol{\mu}}_2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

con $\underline{\boldsymbol{\mu}}_1 \in \mathbb{R}^r$, $\underline{\boldsymbol{\mu}}_2 \in \mathbb{R}^s$, $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{r\times r}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T \in \mathbb{R}^{r\times s}$, $\boldsymbol{\Sigma}_{22} \in \mathbb{R}^{s\times s}$.

Distribución marginal

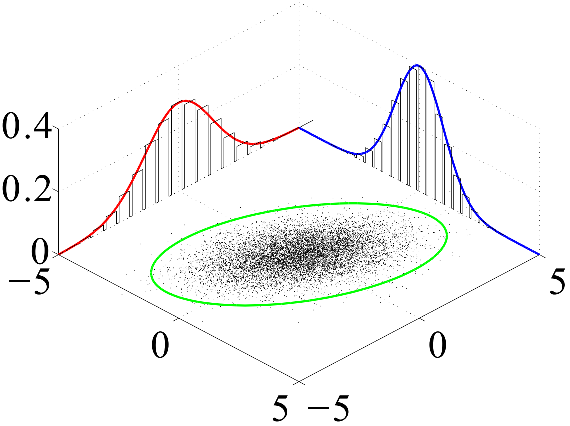
- Si $\underline{\mathbf{x}}_1$ y $\underline{\mathbf{x}}_2$ son conjuntamente gaussianos multivariados, ¿cuál es la distribución de $\underline{\mathbf{x}}_1$?
- Se cumple $\mathrm{E}[\underline{\mathbf{x}}_1] = \underline{\boldsymbol{\mu}}_1$
- La relación entre las covarianzas:

$$
\begin{aligned}
\mathrm{Cov}(\underline{\mathbf{x}}) = \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} &= \mathrm{E}[(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})^T] \\
&= \mathrm{E}\left[\begin{pmatrix} \underline{\mathbf{x}}_1 - \underline{\boldsymbol{\mu}}_1 \\ \underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2 \end{pmatrix} \begin{pmatrix} \underline{\mathbf{x}}_1 - \underline{\boldsymbol{\mu}}_1 \\ \underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2 \end{pmatrix}^T \right] \\
&= \mathrm{E}\begin{bmatrix} (\underline{\mathbf{x}}_1 - \underline{\boldsymbol{\mu}}_1)(\underline{\mathbf{x}}_1 - \underline{\boldsymbol{\mu}}_1)^T & (\underline{\mathbf{x}}_1 - \underline{\boldsymbol{\mu}}_1)(\underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2)^T \\ (\underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2)(\underline{\mathbf{x}}_1 - \underline{\boldsymbol{\mu}}_1)^T & (\underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2)(\underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2)^T \end{bmatrix}
\end{aligned}
$$

de donde se obtienen las equivalencias con $\mathbf{\Sigma}_{ij}$.

Marginales de una gaussiana

- Puesto que la distribución marginal de una gaussiana es a su vez gaussiana entonces $\underline{\mathbf{x}}_1 \sim \mathcal{N}(\underline{\boldsymbol{\mu}}_1, \boldsymbol{\Sigma}_{11})$



Distribución condicional

- ¿Cuál es la distribución condicional de $\underline{\mathbf{x}}_1$ dado $\underline{\mathbf{x}}_2$?
- Se puede demostrar con

$$p(\underline{\mathbf{x}}_1|\underline{\mathbf{x}}_2) = \frac{p(\underline{\mathbf{x}}_1,\underline{\mathbf{x}}_2)}{p(\underline{\mathbf{x}}_2)} \begin{array}{l} \leftarrow \mathcal{N}(\underline{\boldsymbol{\mu}},\boldsymbol{\Sigma}) \\ \leftarrow \mathcal{N}(\underline{\boldsymbol{\mu}}_2,\boldsymbol{\Sigma}_{22}) \end{array}$$

que $\underline{\mathbf{x}}_1|\underline{\mathbf{x}}_2 \sim \mathcal{N}(\underline{\boldsymbol{\mu}}_{1|2},\boldsymbol{\Sigma}_{1|2})$, con

$$\begin{aligned} \underline{\boldsymbol{\mu}}_{1|2} &= \underline{\boldsymbol{\mu}}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{aligned}$$

Modelo de análisis de factores

- En el modelo de análisis de factores postulamos la distribución conjunta en $(\underline{\mathbf{x}},\underline{\mathbf{z}})$ con $\underline{\mathbf{x}} \in \mathbb{R}^n$, la variable aleatoria latente $\underline{\mathbf{z}} \in \mathbb{R}^k$, $k < n$, como

$$\underline{\mathbf{z}} \sim \mathcal{N}(\underline{\mathbf{0}},\mathbf{I})$$
$$\underline{\mathbf{x}}|\underline{\mathbf{z}} \sim \mathcal{N}(\underline{\boldsymbol{\mu}} + \boldsymbol{\Lambda}\underline{\mathbf{z}},\boldsymbol{\Psi})$$

- Parámetros: $\underline{\boldsymbol{\mu}} \in \mathbb{R}^n$, $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times k}$, $\boldsymbol{\Psi} = \operatorname{diag}(\underline{\psi}) \in \mathbb{R}^{n\times n}$, eligiendo $k < n$.
- Note que la variable latente $\underline{\mathbf{z}}$ es ahora continua

Interpretación

- Cada punto $\underline{\mathbf{x}}^{(i)}$ se genera en varios pasos
- Primero se muestrea $\underline{\mathbf{z}}^{(i)} \in \mathbb{R}^k$ de gaussiano multivariado
- Luego $\underline{\mathbf{z}}^{(i)}$ se mapea de forma afín con $\underline{\boldsymbol{\mu}} + \mathbf{\Lambda}\underline{\mathbf{z}}^{(i)}$
- Finalmente $\underline{\mathbf{x}}^{(i)}$ se genera sumando ruido gaussiano de covarianza $\mathbf{\Psi}$

$$\underline{\mathbf{z}} \sim \mathcal{N}(\underline{\mathbf{0}}, \mathbf{I}) \qquad \underline{\varepsilon} \sim \mathcal{N}(\underline{\mathbf{0}}, \mathbf{\Psi})$$

$$\underline{\mathbf{x}} = \underline{\boldsymbol{\mu}} + \mathbf{\Lambda}\underline{\mathbf{z}} + \underline{\varepsilon}$$

  con $\underline{\varepsilon}$ y $\underline{\mathbf{z}}$ independientes.
- Las variables aleatorias $\underline{\mathbf{z}}$ y $\underline{\mathbf{x}}$ tienen una distribución gaussiana conjunta

$$\begin{bmatrix} \underline{\mathbf{z}} \\ \underline{\mathbf{x}} \end{bmatrix} \sim \mathcal{N}(\underline{\boldsymbol{\mu}}_{zx}, \mathbf{\Sigma})$$

- Las variables aleatorias $\underline{\mathbf{z}}$ y $\underline{\mathbf{x}}$ tienen una distribución gaussiana conjunta

$$\begin{bmatrix} \underline{\mathbf{z}} \\ \underline{\mathbf{x}} \end{bmatrix} \sim \mathcal{N}(\underline{\boldsymbol{\mu}}_{zx}, \boldsymbol{\Sigma})$$

- Necesitamos encontrar $\underline{\boldsymbol{\mu}}_{zx}$ y $\boldsymbol{\Sigma}$.
- Puesto que $\underline{\mathbf{z}} \sim \mathcal{N}(\underline{\mathbf{0}}, \underline{\mathbf{I}})$ entonces $\mathsf{E}[\underline{\mathbf{z}}] = \underline{\mathbf{0}}$
- La esperanza de $\underline{\mathbf{x}}$

$$\mathsf{E}[\underline{\mathbf{x}}] = \mathsf{E}[\underline{\boldsymbol{\mu}} + \boldsymbol{\Lambda}\underline{\mathbf{z}} + \underline{\varepsilon}] = \underline{\boldsymbol{\mu}} + \boldsymbol{\Lambda}\,\mathsf{E}[\underline{\mathbf{z}}] + \mathsf{E}[\underline{\varepsilon}] = \underline{\boldsymbol{\mu}}$$

- Por esta razón:

$$\underline{\boldsymbol{\mu}}_{zx} = \begin{bmatrix} \underline{\mathbf{0}} \\ \underline{\boldsymbol{\mu}} \end{bmatrix}$$

Covarianza por bloques

- Para encontrar $\boldsymbol{\Sigma}$ necesitamos calcular

$$
\begin{aligned}
\boldsymbol{\Sigma}_{zz} &= \mathrm{E}[(\underline{\mathbf{z}} - \mathrm{E}[\underline{\mathbf{z}}])(\underline{\mathbf{z}} - \mathrm{E}[\underline{\mathbf{z}}])^T] && \text{superior izquierdo} \\
\boldsymbol{\Sigma}_{zx} &= \mathrm{E}[(\underline{\mathbf{z}} - \mathrm{E}[\underline{\mathbf{z}}])(\underline{\mathbf{x}} - \mathrm{E}[\underline{\mathbf{x}}])^T] && \text{superior derecho} \\
\boldsymbol{\Sigma}_{xx} &= \mathrm{E}[(\underline{\mathbf{x}} - \mathrm{E}[\underline{\mathbf{x}}])(\underline{\mathbf{x}} - \mathrm{E}[\underline{\mathbf{x}}])^T] && \text{inferior derecho}
\end{aligned}
$$

- Puesto que $\underline{\mathbf{z}} \sim \mathcal{N}(\underline{\mathbf{0}}, \mathbf{I})$ entonces $\boldsymbol{\Sigma}_{zz} = \mathrm{Cov}(\underline{\mathbf{z}}) = \mathbf{I}$
- Para $\boldsymbol{\Sigma}_{zx}$, con $\mathrm{Cov}(\underline{\mathbf{z}}) = \mathrm{E}[\underline{\mathbf{z}}\underline{\mathbf{z}}^T] = \mathbf{I}$ porque $\mathrm{E}[\underline{\mathbf{z}}] = \underline{\mathbf{0}}$

$$
\begin{aligned}
\mathrm{E}[(\underline{\mathbf{z}} - \mathrm{E}[\underline{\mathbf{z}}])(\underline{\mathbf{x}} - \mathrm{E}[\underline{\mathbf{x}}])^T] &= \mathrm{E}[\underline{\mathbf{z}}(\underline{\boldsymbol{\mu}} + \boldsymbol{\Lambda}\underline{\mathbf{z}} + \underline{\boldsymbol{\varepsilon}} - \underline{\boldsymbol{\mu}})^T] \\
&= \underbrace{\mathrm{E}[\underline{\mathbf{z}}\underline{\mathbf{z}}^T]}_{\mathbf{I}}\boldsymbol{\Lambda}^T + \underbrace{\mathrm{E}[\underline{\mathbf{z}}\underline{\boldsymbol{\varepsilon}}^T]}_{\text{0 por independencia}} = \boldsymbol{\Lambda}^T
\end{aligned}
$$

Bloque $\mathbf{\Sigma}_{xx}$

- Para el último término $\mathbf{\Sigma}_{xx}$

$$
\begin{aligned}
\mathbf{\Sigma}_{xx} &= \mathrm{E}\left[(\underline{\mathbf{x}} - \mathrm{E}[\underline{\mathbf{x}}])(\underline{\mathbf{x}} - \mathrm{E}[\underline{\mathbf{x}}])^T\right] \\
&= \mathrm{E}\left[(\underline{\boldsymbol{\mu}} + \mathbf{\Lambda}\underline{\mathbf{z}} + \underline{\boldsymbol{\varepsilon}} - \underline{\boldsymbol{\mu}})(\underline{\boldsymbol{\mu}} + \mathbf{\Lambda}\underline{\mathbf{z}} + \underline{\boldsymbol{\varepsilon}} - \underline{\boldsymbol{\mu}})^T\right] \\
&= \mathrm{E}\left[\mathbf{\Lambda}\underline{\mathbf{z}}\underline{\mathbf{z}}^T\mathbf{\Lambda}^T + \underline{\boldsymbol{\varepsilon}}\underline{\mathbf{z}}^T\mathbf{\Lambda}^T + \mathbf{\Lambda}\underline{\mathbf{z}}\underline{\boldsymbol{\varepsilon}}^T + \underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}^T\right] \\
&= \mathbf{\Lambda}\,\mathrm{E}[\underline{\mathbf{z}}\underline{\mathbf{z}}^T]\mathbf{\Lambda}^T + \mathrm{E}[\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}^T] \\
&= \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}
\end{aligned}
$$

Distribución conjunta y marginal

- Agrupando todos los resultados anteriores:

$$\begin{bmatrix} \underline{\mathbf{z}} \\ \underline{\mathbf{x}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \underline{\mathbf{0}} \\ \underline{\boldsymbol{\mu}} \end{bmatrix}, \begin{bmatrix} \mathbf{I} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi} \end{bmatrix} \right)$$

- La distribución marginal de $\underline{\mathbf{x}} \sim \mathcal{N}(\underline{\boldsymbol{\mu}}, \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi})$
- Dado un conjunto de entrenamiento $\{\underline{\mathbf{x}}^{(i)}; i = 1, \ldots, m\}$ la verosimilitud logarítmica es:

$$\ell(\underline{\boldsymbol{\mu}}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}) = \ln \prod_{i=1}^{m} \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{xx}|^{1/2}} e^{\left(-\frac{1}{2} (\underline{\mathbf{x}}^{(i)} - \underline{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}_{xx}^{-1} (\underline{\mathbf{x}}^{(i)} - \underline{\boldsymbol{\mu}}) \right)}$$

- La estimación de máxima verosimilitud maximiza lo anterior respecto a sus parámetros, pero no se puede hacer fácilmente, por lo que requerimos de nuevo el algoritmo EM.

EM para análisis de factores

- Paso E
  Encuentre $Q_i(\underline{\mathbf{z}}^{(i)}) = p(\underline{\mathbf{z}}^{(i)}|\underline{\mathbf{x}}^{(i)};\underline{\boldsymbol{\theta}})$
- Paso M

$$\underline{\boldsymbol{\theta}} = \arg\max_{\underline{\boldsymbol{\theta}}} \sum_i \int_{\underline{\mathbf{z}}^{(i)}} Q_i(\underline{\mathbf{z}}^{(i)}) \ln\left(\frac{p(\underline{\mathbf{x}}^{(i)},\underline{\mathbf{z}}^{(i)};\underline{\boldsymbol{\theta}})}{Q_i(\underline{\mathbf{z}}^{(i)})}\right) d\underline{\mathbf{z}^{(\mathbf{i})}}$$

Paso E (1)

- Paso E es relativamente simple de derivar.
- Necesitamos $Q_i(\underline{\mathbf{z}}^{(i)}) = p(\underline{\mathbf{z}}^{(i)}|\underline{\mathbf{x}}^{(i)};\underline{\boldsymbol{\mu}},\mathbf{\Lambda},\mathbf{\Psi})$
- Considerando que

$$\begin{bmatrix}\underline{\mathbf{z}}\\ \underline{\mathbf{x}}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\underline{\mathbf{0}}\\ \underline{\boldsymbol{\mu}}\end{bmatrix}, \begin{bmatrix}\mathbf{I} & \mathbf{\Lambda}^T\\ \mathbf{\Lambda} & \mathbf{\Lambda}\mathbf{\Lambda}^T+\mathbf{\Psi}\end{bmatrix}\right)$$

y además que $\begin{bmatrix}\underline{\mathbf{x}}_1\\ \underline{\mathbf{x}}_2\end{bmatrix}$ tiene $\underline{\mathbf{x}}_1|\underline{\mathbf{x}}_2 \sim \mathcal{N}(\underline{\boldsymbol{\mu}}_{1|2},\mathbf{\Sigma}_{1|2})$, con

$$\underline{\boldsymbol{\mu}}_{1|2} = \underline{\boldsymbol{\mu}}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\underline{\mathbf{x}}_2 - \underline{\boldsymbol{\mu}}_2)$$
$$\mathbf{\Sigma}_{1|2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$$

Paso E (2)

entonces

$$\underline{\mathbf{z}}^{(i)}|\underline{\mathbf{x}}^{(i)};\underline{\boldsymbol{\mu}},\boldsymbol{\Lambda},\boldsymbol{\Psi}\sim\mathcal{N}(\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}},\boldsymbol{\Sigma}_{z^{(i)}|x^{(i)}})$$

donde

$$\begin{aligned}\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}} &= \boldsymbol{\Lambda}^T(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T+\boldsymbol{\Psi})^{-1}(\underline{\mathbf{x}}^{(i)}-\underline{\boldsymbol{\mu}})\\ \boldsymbol{\Sigma}_{z^{(i)}|x^{(i)}} &= \mathbf{I}-\boldsymbol{\Lambda}^T(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T+\boldsymbol{\Psi})^{-1}\boldsymbol{\Lambda}\end{aligned}$$

con lo que

$$Q_i(\underline{\mathbf{z}}^{(i)})=\frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}_{z^{(i)}|x^{(i)}}|}}e^{\left(-\frac{1}{2}(\underline{\mathbf{z}}^{(i)}-\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}})^T\boldsymbol{\Sigma}^{-1}_{z^{(i)}|x^{(i)}}(\underline{\mathbf{z}}^{(i)}-\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}})\right)}$$

Paso M (1)

- El paso M requiere maximizar

$$
\begin{aligned}
&\sum_{i=1}^{m} \int_{\underline{\mathbf{z}}^{(i)}} Q_i(\underline{\mathbf{z}}^{(i)}) \ln \frac{p(\underline{\mathbf{x}}^{(i)},\underline{\mathbf{z}}^{(i)};\underline{\boldsymbol{\mu}},\boldsymbol{\Lambda},\boldsymbol{\Psi})}{Q_i(\underline{\mathbf{z}}^{(i)})}\, d\underline{\mathbf{z}}^{(i)} \\
&= \sum_{i=1}^{m} \mathrm{E}_{\underline{\mathbf{z}}^{(i)}\sim Q_i}\left[\ln \frac{p(\underline{\mathbf{x}}^{(i)},\underline{\mathbf{z}}^{(i)};\underline{\boldsymbol{\mu}},\boldsymbol{\Lambda},\boldsymbol{\Psi})}{Q_i(\underline{\mathbf{z}}^{(i)})}\right] \\
&= \sum_{i=1}^{m} \mathrm{E}_{\underline{\mathbf{z}}^{(i)}\sim Q_i}\left[\ln p(\underline{\mathbf{x}}^{(i)}|\underline{\mathbf{z}}^{(i)};\underline{\boldsymbol{\mu}},\boldsymbol{\Lambda},\boldsymbol{\Psi})\right] + \mathrm{E}_{\underline{\mathbf{z}}^{(i)}\sim Q_i}\left[\ln \frac{p(\underline{\mathbf{z}}^{(i)};\underline{\boldsymbol{\mu}},\boldsymbol{\Lambda},\boldsymbol{\Psi})}{Q_i(\underline{\mathbf{z}}^{(i)})}\right]
\end{aligned}
$$

con respecto a $\underline{\boldsymbol{\mu}},\boldsymbol{\Lambda},\boldsymbol{\Psi}$

Paso M (2)

- Después de la manipulación algebraica se maximiza en:

$$
\begin{aligned}
\underline{\boldsymbol{\mu}} &= \frac{1}{m}\sum_{i=1}^{m}\underline{\mathbf{x}}^{(i)} \\
\mathbf{\Phi} &= \frac{1}{m}\sum_{i=1}^{m}\underline{\mathbf{x}}^{(i)}\underline{\mathbf{x}}^{(i)\,T} - \underline{\mathbf{x}}^{(i)}\underline{\boldsymbol{\mu}}^{T}_{z^{(i)}|x^{(i)}}\mathbf{\Lambda}^{T} - \mathbf{\Lambda}\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}}\underline{\mathbf{x}}^{(i)\,T} \\
&\quad + \mathbf{\Lambda}(\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}}\underline{\boldsymbol{\mu}}^{T}_{z^{(i)}|x^{(i)}} + \mathbf{\Sigma}_{z^{(i)}|x^{(i)}})\mathbf{\Lambda}^{T} \\
\mathbf{\Psi} &= \operatorname{diag}(\mathbf{\Phi}) \\
\mathbf{\Lambda} &= \left(\sum_{i=1}^{m}(\underline{\mathbf{x}}^{(i)} - \underline{\boldsymbol{\mu}})\underline{\boldsymbol{\mu}}^{T}_{z^{(i)}|x^{(i)}}\right)\left(\sum_{i=1}^{m}\underline{\boldsymbol{\mu}}_{z^{(i)}|x^{(i)}}\underline{\boldsymbol{\mu}}^{T}_{z^{(i)}|x^{(i)}} + \mathbf{\Sigma}_{z^{(i)}|x^{(i)}}\right)^{-1}
\end{aligned}
$$

Resumen

1. Matrices de covarianza singulares
2. Condicionales y marginales de gaussianas
3. Modelo de análisis de factores
4. EM para análisis de factores

Este documento ha sido elaborado con software libre incluyendo LaTeX, Beamer, GNUPlot, GNU/Octave, XFig, Inkscape, GNU-Make y Subversion en GNU/Linux