Optimización analítica: otro ejemplo

- Estimar los parámetros de una gaussiana multivariada (en $\mathbb{R}^D$), con Σ dada:

$$p(\boldsymbol{x}; \boldsymbol{\Theta}) = (2\pi)^{-D/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

donde $\boldsymbol{\Theta} \equiv (\boldsymbol{\mu}, \Sigma)$. Si Σ está prefijada, entonces $\boldsymbol{\Theta} \equiv \boldsymbol{\mu} \in \mathbb{R}^D$

- Logaritmo de la verosimilitud de una muestra $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$:

$$\begin{aligned} q_S(\boldsymbol{\Theta}) \equiv L_S(\boldsymbol{\Theta}) &= \log \prod_{n=1}^{N} p(\boldsymbol{x}_n; \boldsymbol{\Theta}) = \sum_{n=1}^{N} \log p(\boldsymbol{x}_n; \boldsymbol{\Theta}) \\ &= N \log\left((2\pi)^{-D/2} |\Sigma|^{-1/2}\right) - \frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}) \end{aligned}$$

- Para una estimación de máxima verosimilitud basta hacer $\boldsymbol{\nabla} L_S(\boldsymbol{\Theta}) = \mathbf{0}$. Si Σ está prefijada, se obtiene (*ejercicio*):

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n$$

3.3. (p.3.10, 0.25 puntos) Desarrollar con detalle todos los pasos necesarios para obtener el estimador de máxima verosimiltud del vector media de una de una gaussiana multivariada cuya matriz de covarianza es fija y conocida, a partir de una muestra de de vectores x_1,x_2,...,x_N.

1

$$P(x;\theta) = (2\pi)^{-\frac{D}{2}} \, |\Sigma|^{-\frac{1}{2}} \cdot e^{\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^t \Sigma^{-1} (\vec{x}-\vec{\mu})\right)}$$

Logaritmo de la verosimilitud

$$L_S(\theta) = \log\left(\prod_{n=1}^{N} \left(\underbrace{2\pi^{-\frac{D}{2}}}_{A} \, \underbrace{|\Sigma|^{-\frac{1}{2}}}_{B} \cdot e^{\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^t \Sigma^{-1}(\vec{x}-\vec{\mu})\right)}\right)\right) = \log\left(\prod_{n=1}^{N} (A \cdot B \cdot e^{C})\right)$$

$$= \log\left(A^N \cdot B^N \cdot e^{\left(\sum_{n=1}^{N} \left(-\frac{1}{2}(\vec{x}_n-\vec{\mu})^t \Sigma^{-1}(\vec{x}_n-\vec{\mu})\right)\right)}\right)$$

$$= \log A^N + \log B^N + \sum_{n=1}^{N} \left(-\frac{1}{2}(\vec{x}_n-\vec{\mu})^t \Sigma^{-1}(\vec{x}_n-\vec{\mu})\right)$$

$$= \frac{-ND}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} \left((\vec{x}_n-\vec{\mu})^t \Sigma^{-1}(\vec{x}_n-\vec{\mu})\right)$$

②

$$L_S(\theta) = -\frac{ND}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}\left((\vec{x}_n-\vec{\mu})^t\,\Sigma^{-1}(\vec{x}_n-\vec{\mu})\right)$$

$\boxed{\text{Derivamos en función de } \vec{\mu}}$

$$\frac{\partial}{\partial \mu} = 0 - 0 - \frac{1}{2}\sum_{n=1}^{N}\left(2\cdot\Sigma^{-1}(\vec{x}_n-\vec{\mu})\right)$$

$$\frac{\partial}{\partial \mu} = -\sum_{n=1}^{N}\left(\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu})\right)$$

Dado que Σ es simétrica, Σ^{-1} también lo es y como $(\vec{x}-\vec{\mu})$ no depende de Σ^{-1}, podemos aplicar la fórmula para el gradiente de una forma cuadrática

$$\frac{\partial\, \vec{z}^t A \vec{z}}{\partial \vec{z}} = 2A\vec{z}$$

$\boxed{\text{Igualamos a } 0}$

$$\sum_{n=1}^{N}\left(\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu})\right)=0 \Rightarrow \Sigma^{-1}\sum_{n=1}^{N}(\vec{x}_n-\vec{\mu})=0$$

Dado que Σ está prefijada, la expresion solo se cumple si: $\sum_{n=1}^{N}(\vec{x}_n-\vec{\mu})=0$

$$\Rightarrow \sum_{n=1}^{N}(\vec{x}_n-\vec{\mu})=0 \Rightarrow \sum_{n=1}^{N}\vec{x}_n - \sum_{n=1}^{N}\vec{\mu}=0 \Rightarrow \sum_{n=1}^{N}\vec{x}_n - N\vec{\mu}$$

$$\Rightarrow \vec{\mu} = \frac{\sum_{n=1}^{N}\vec{x}_n}{N}$$