Introducción a la Inteligencia Artificial
Clase 7

.UBA fiuba
FACULTAD DE INGENIERÍA

Índice Clase 7.

1. Continuación PCA

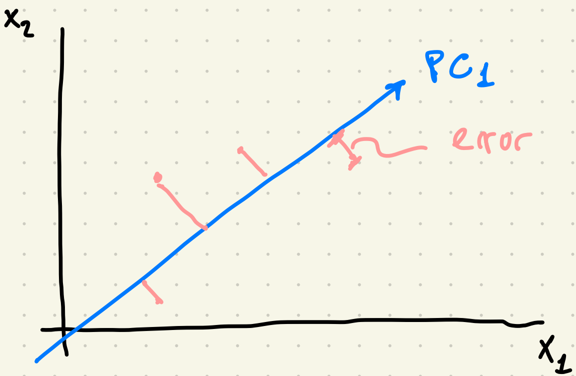
   Minimización error de representación

2. Modelos de mixtura

   - Mixtura Gaussiana

Segunda forma de derivación de PCA:

Minimización del error de reducción:



Si proyectamos un pto. sobre una dirección de la reconstrucción vamos a poder definir la posición del pto $\tilde{p}$

p_i : pto. original
$\tilde{p}_i$: pto. proyectado

Vamos a minimizar $\|p_i - \tilde{p}_i\|$

Vamos a suponer $\exists$ B base ortonormal de $\mathbb{R}^D$ con esto decimos que $\forall x \in \mathbb{R}^D$ podemos asegurar:

$$\vec{x} = \sum_{d=1}^{D} \alpha_d \cdot \bar{b}_d = \underbrace{\sum_{i=1}^{M} \alpha_i \bar{b}_i}_{\tilde{x}} + \sum_{d=M+1}^{D} \alpha_d \bar{b}_d$$

queremos encontrar $\tilde{x} = \sum_{i=1}^{m} \alpha_i \cdot \bar{b}_i \in U \subseteq \mathbb{R}^D$ tal que podamos maximizar la similaridad entre $\tilde{x}$ y x:

$$\tilde{x} = \sum_{i=1}^{m} \tilde{z}_i b_i = \bar{\bar{B}} \cdot \bar{\tilde{z}}_i$$

buscamos minimizar el MSE $\| x - \tilde{x} \|^2$:

①. optimizar $\tilde{z}_n$ para una dada base.

②. encontrar una base óptima.

$$J_M = \frac{1}{N} \sum_{i=1}^{N} \| x_n - \tilde{x}_n \|^2$$

$$\partial_{z_{in}} J_M = \frac{\partial J_M}{\partial \tilde{x}_n} \cdot \frac{\partial \tilde{x}_n}{\partial z_{in}}$$

$$= \left(-\frac{2}{N} \cdot (x_n - \tilde{x}_n)^t \right) \cdot \left(\partial_{z_{in}} \left(\sum_{i=1}^{M} z_{in} \cdot b \right) \right) = -\frac{2}{N} (x_n - \tilde{x}_n)^t \cdot b_i$$

$$\partial_{z_{in}} J_M = -\frac{2}{N}\left(x_n - \sum_{i=1}^{M} z_{in} \cdot b_i\right)^t \cdot b_i = -\frac{2}{N}\left(x_n^t \cdot b_i - z_{in} \cdot \underbrace{b_i^t \cdot b_i}_{=1}\right)$$

$$= -\frac{2}{N}\left(x_n^t b_i - z_{in}\right)$$

$$\partial_{z_{in}} J_M = 0 \Rightarrow -\frac{2}{N}\left(x_n^t b_i - z_{in}\right) = 0 \rightarrow \boxed{z_{in} = b_i \, x_n^t}$$

los z_{in} son las coord. que vamos a encontrar para cada vector $(b_i \, x_n^t)$

. buscamos ahora la base óptima:

$$\tilde{x}_n = \sum_{n=1}^{M} z_{mn} \cdot b_n = \sum_{n=1}^{M} (x_n^t b_n) \, b_n = \sum_{n=1}^{M} (b_n b_n^t) . x_n$$

$$x = \left(\sum_{n=1}^{M} b_n b_n^t\right) . x_n = \left(\sum_{j=M+1}^{D} b_j b_j^t\right) . x_j$$

distancia (error) : $x_n - \tilde{x}_n = \sum_{j=M+1}^{D} b_j b_j^t \cdot x_j = \sum_{j=M+1}^{D} (x_n^t \cdot b_j) \cdot b_j$

$$J_M = \frac{1}{N} \sum_n \|x_n - \tilde{x}_n\|^2 = \frac{1}{N} \cdot \sum_{n=1}^{N} \left\| \sum_{j=M+1}^{D} (x_n^t \cdot b_j) b_j \right\|^2$$

$$= \sum_{j=M+1}^{D} b_j^t \underbrace{\left(\frac{1}{N} \sum_{n=1}^{N} x_n \cdot x_n^t \right)}_{S} \cdot b_j$$

$$= \sum_{j=M+1}^{D} b_j^t S \, b_j = \sum_{j=M+1}^{D} tr(b_j^t S b_j) = \sum_{j=M+1}^{D} tr(S \, b_j^t \cdot b_j)$$

$$= tr\left(\underbrace{\sum_{j=M+1}^{D} b_j b_j^t}_{\text{Matriz de proyección}} \cdot S \right)$$

$\rightarrow$ el error de reconstrucción se puede pensar cómo la matriz S proyectada sobre el complemento ortogonal de U.

Al reducir dimensiones:

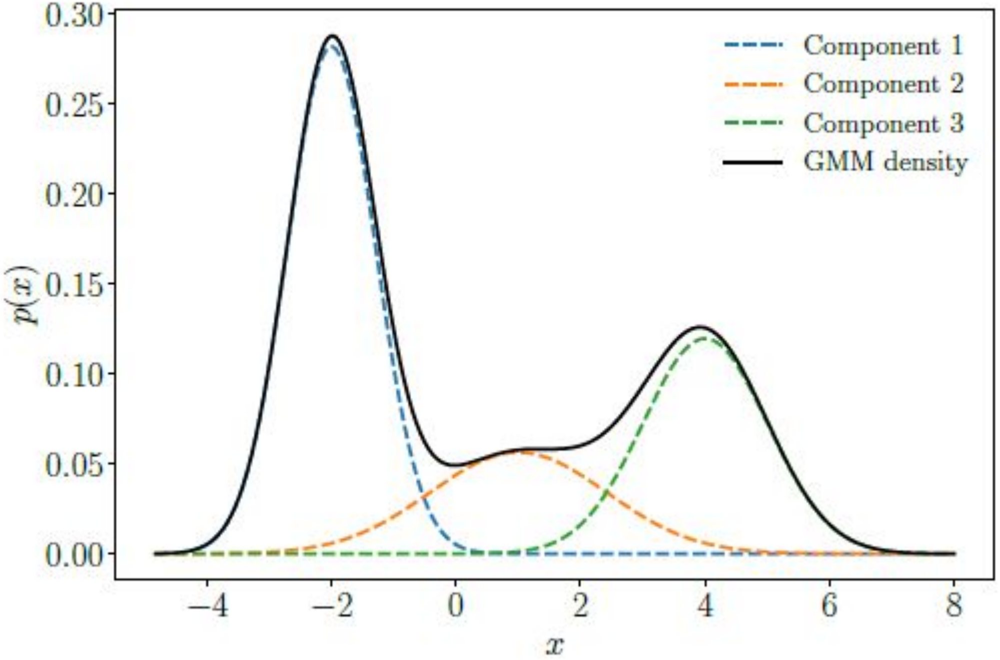
$$\mathbb{R}^D \longmapsto \underbrace{U}_{\text{subespacio explicado}} + \overbrace{\mathbb{I}}^{\text{subespacio no explicado}}$$

para calcular PCA, nosotros podemos elegir maximizar U o minimizar $\mathbb{I}$.

Gaussian Mixture Models

Formulación

$$p(x / \bar{u}, \bar{\Sigma}) = \sum_i \pi_i \mathcal{N}(x / u_i, \sigma_i^2)$$

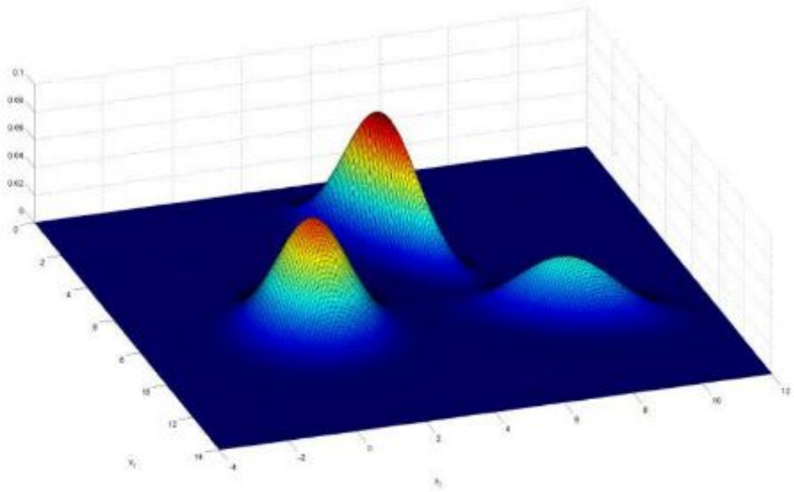


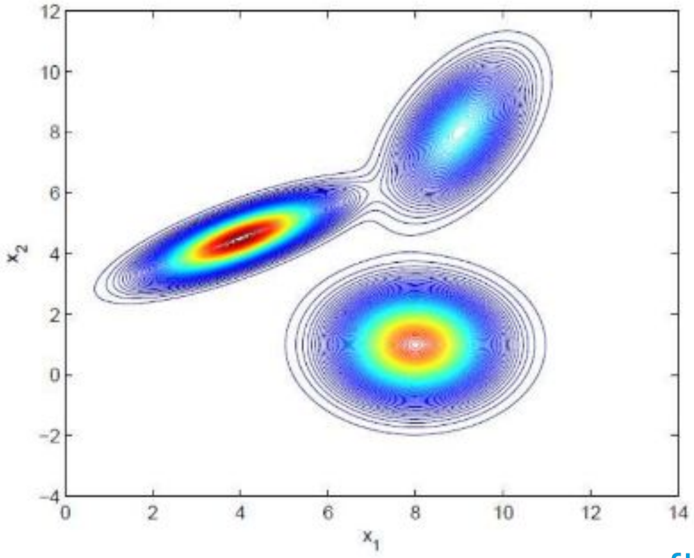
$$p(x \mid \boldsymbol{\theta}) = 0.5\mathcal{N}\left(x \mid -2, \tfrac{1}{2}\right) + 0.2\mathcal{N}\left(x \mid 1, 2\right) + 0.3\mathcal{N}\left(x \mid 4, 1\right)$$

Gaussian Mixture Models

Formulación

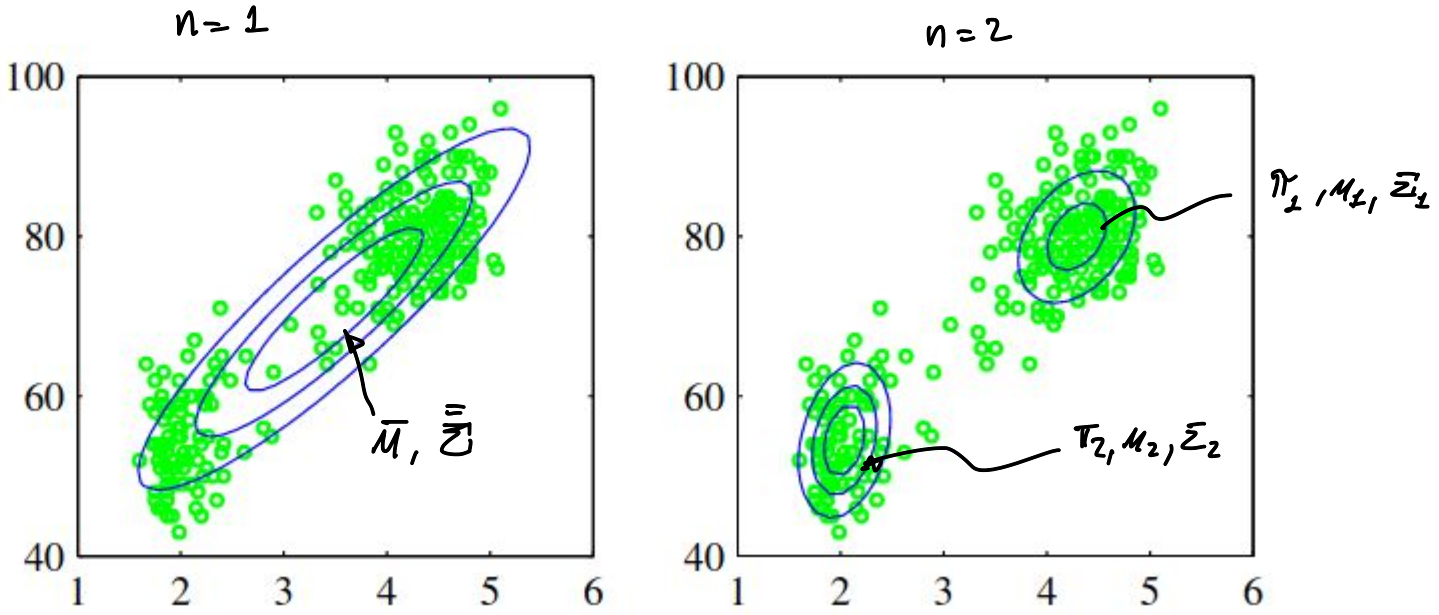
$$p(x) = \underbrace{0.3}_{\pi_1} \mathcal{N}\left(x \mid \underbrace{\begin{pmatrix} 4 \\ 4.5 \end{pmatrix}}_{\mu_1}, \underbrace{\begin{pmatrix} 1.2 & 0.6 \\ 0.6 & 0.5 \end{pmatrix}}_{\Sigma_1}\right) + \underbrace{0.5}_{\pi_2} \mathcal{N}\left(x \mid \underbrace{\begin{pmatrix} 8 \\ 1 \end{pmatrix}}_{\mu_2}, \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{\Sigma_2}\right) + \underbrace{0.2}_{\pi_3} \mathcal{N}\left(x \mid \underbrace{\begin{pmatrix} 9 \\ 8 \end{pmatrix}}_{\mu_3}, \underbrace{\begin{pmatrix} 0.6 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}}_{\Sigma_3}\right)$$





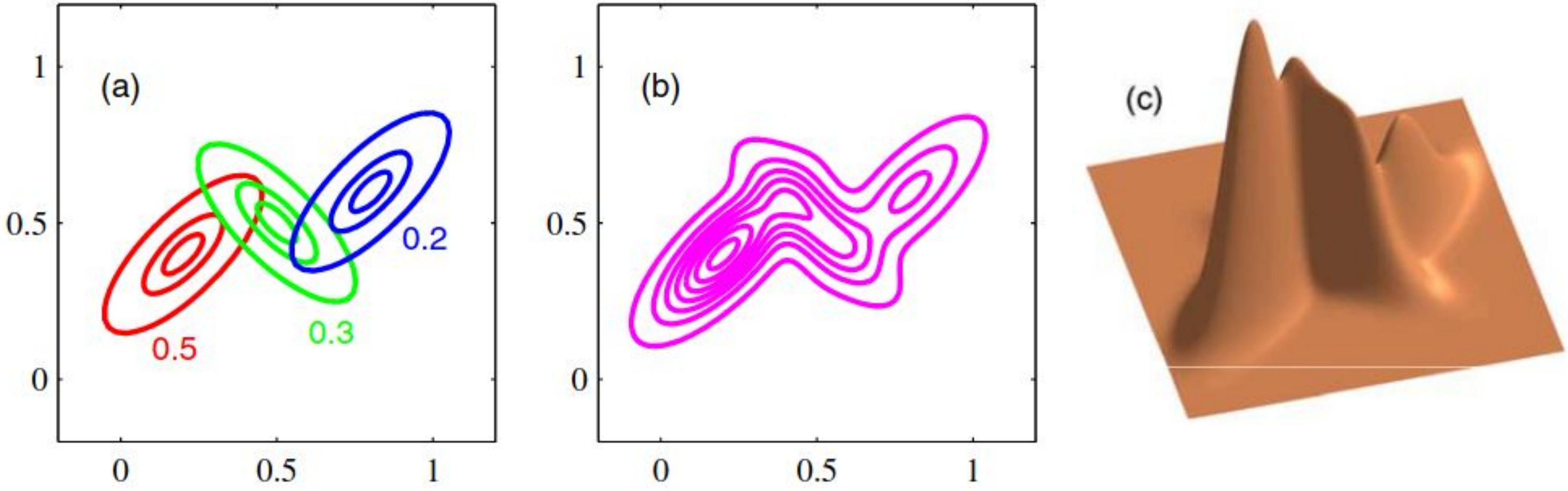
.UBAfiuba FACULTAD DE INGENIERÍA

Gaussian Mixture Models: Estudio de fenómenos naturales



"Old Faithful" dataset. 272 mediciones de erupciones del "Old Faithful" geyser en el Parque Nacional Yellowstone. El eje horizontal representa la duración de una erupción (medida en minutos) y el vertical el tiempo hasta la próxima erupción.

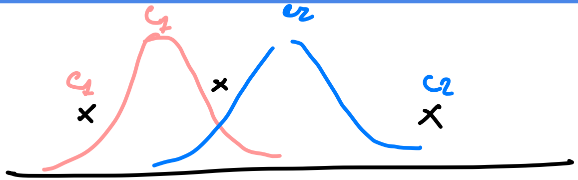
Gaussian Mixture Models: Clustering

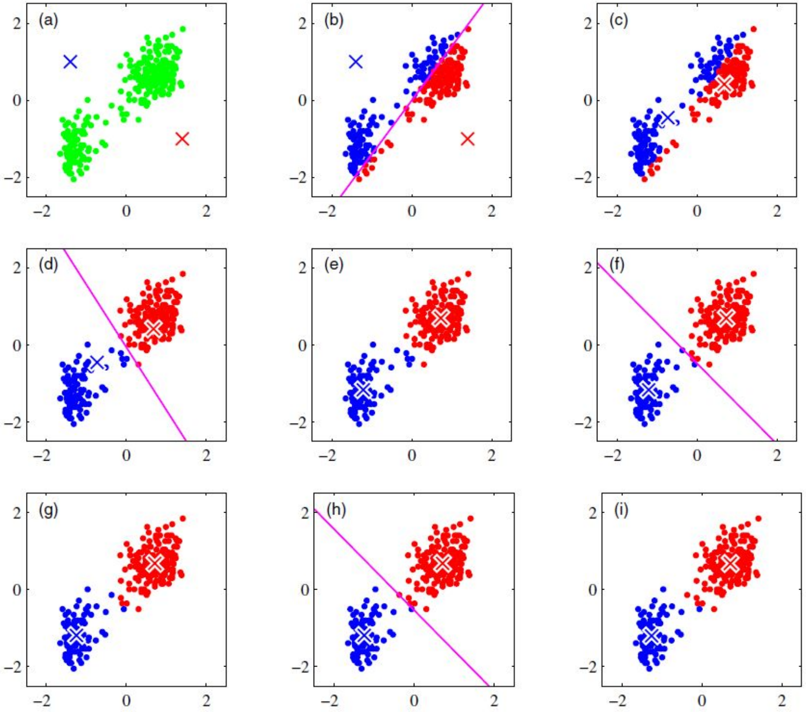


Ejemplo de Gaussian Mixture. En la imagen (a) se muestran las tres distribuciones subyacentes indicando con colores sus variables latentes. En la imagen (b) las curvas de nivel de la distribución conjunta y en la (c) la densidad.

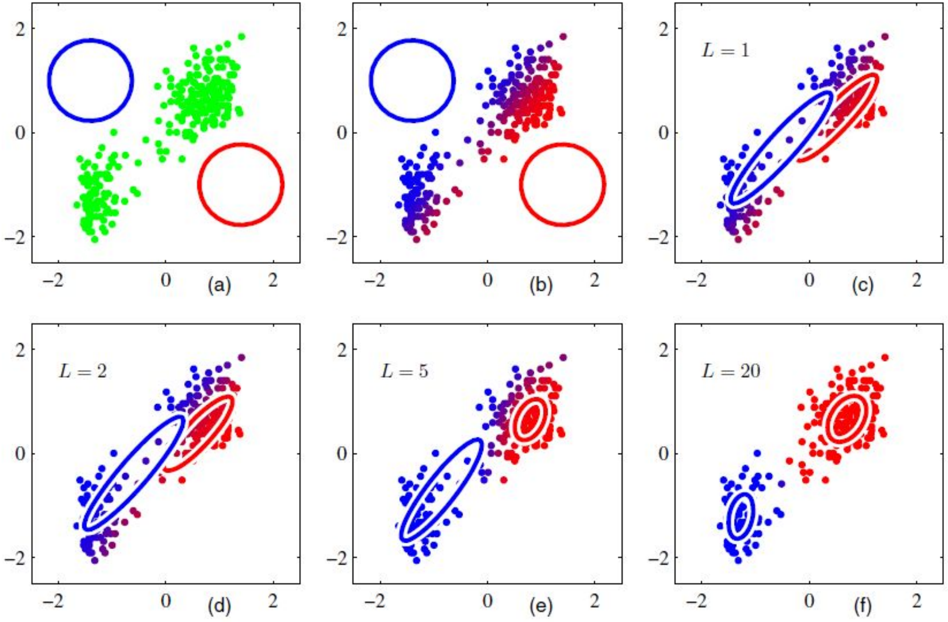
.UBAfiuba FACULTAD DE INGENIERÍA

Gaussian Mixture Models: Clustering



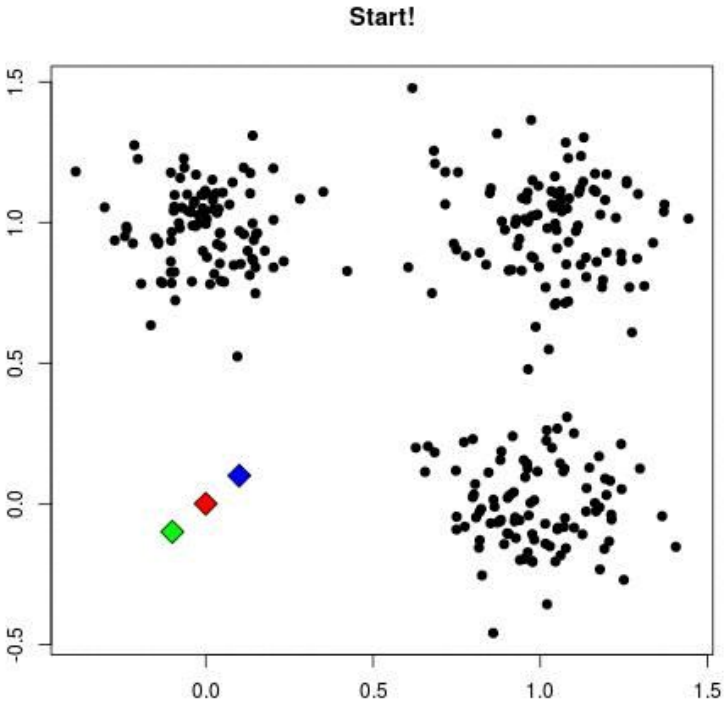


KMeans

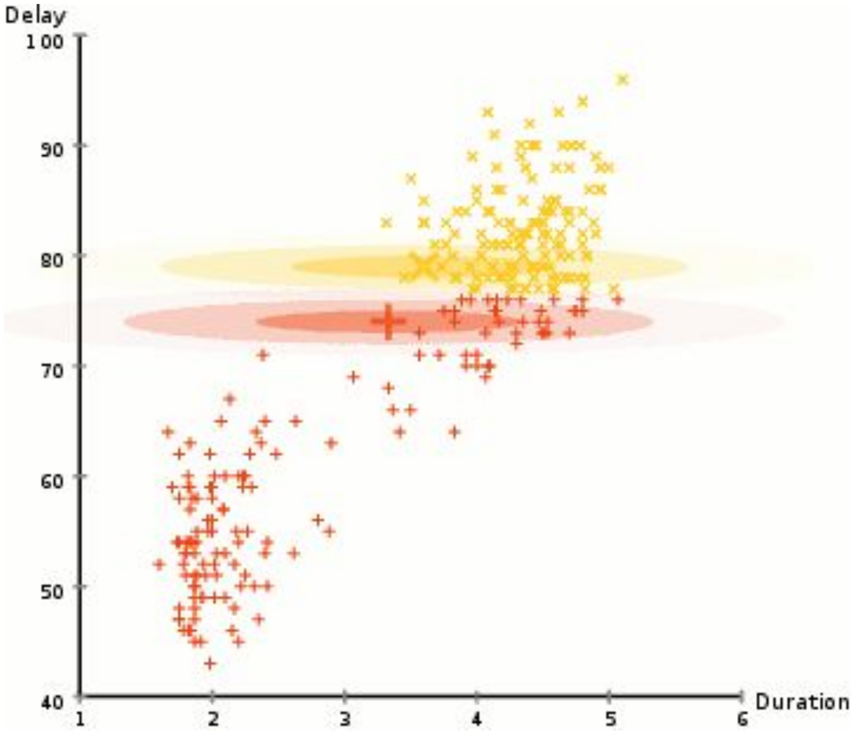


GMM

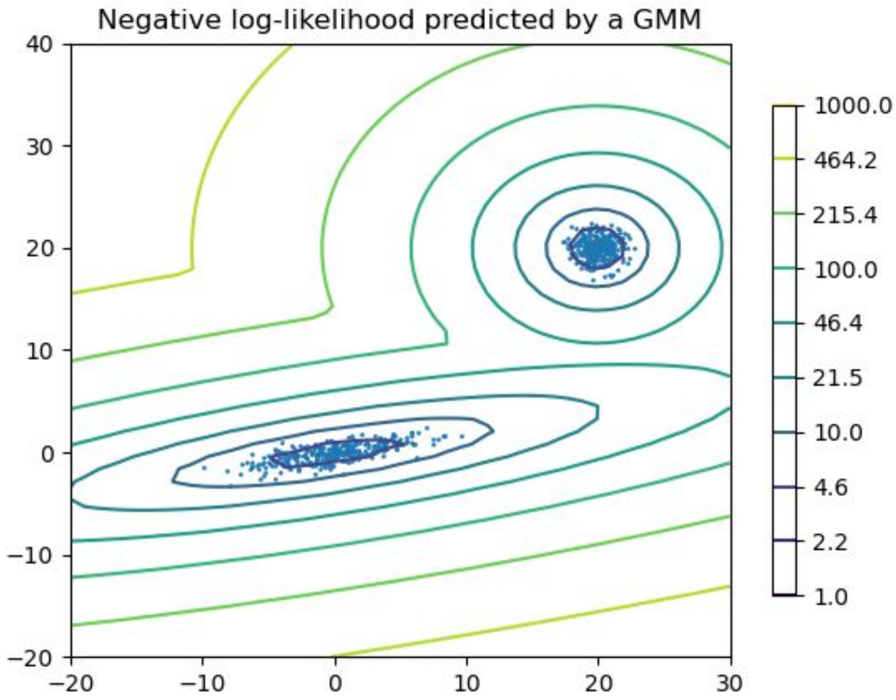
Gaussian Mixture Models - kMeans



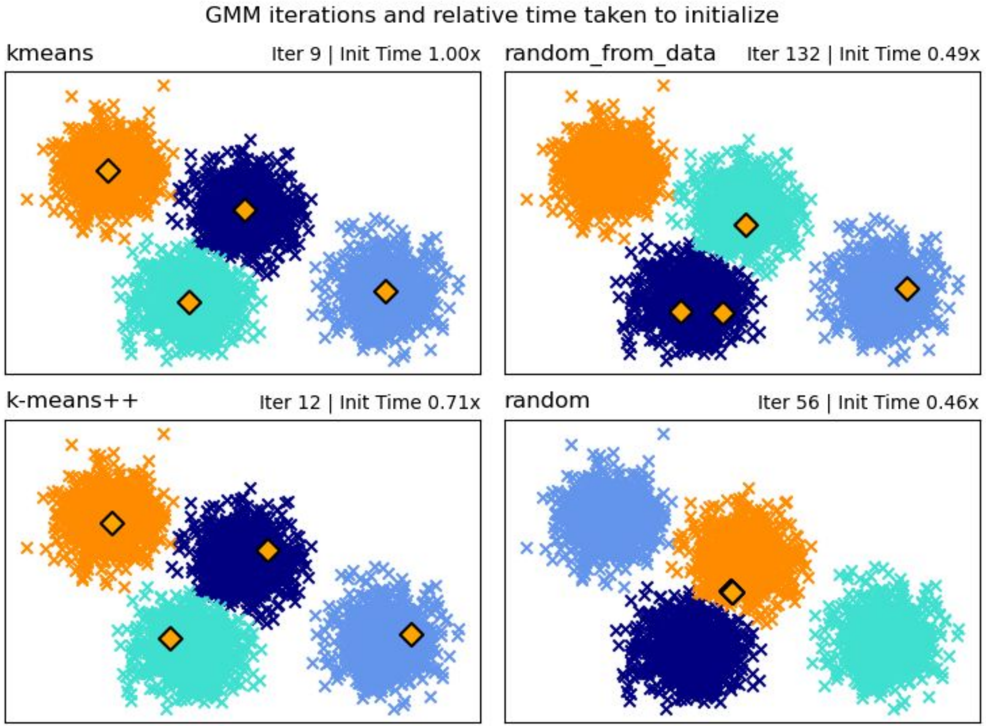
Gaussian Mixture Models: Clustering



Gaussian Mixture Models: Detección de anomalías



Gaussian Mixture Models: Inicialización



Formulación

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k p_k(\boldsymbol{x})$$

$$0 \leqslant \pi_k \leqslant 1\,, \quad \sum_{k=1}^{K} \pi_k = 1\,,$$

Mixture Models - General

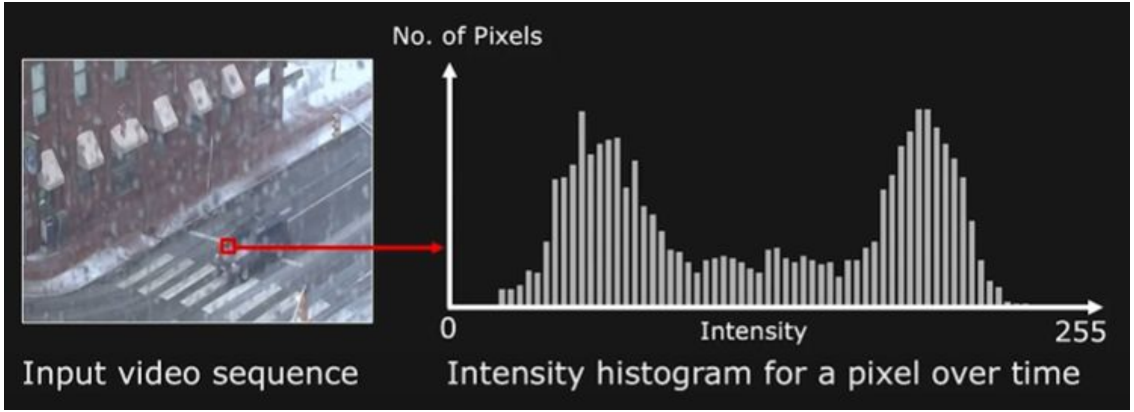
$$p(\boldsymbol{x}\,|\,\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big)$$

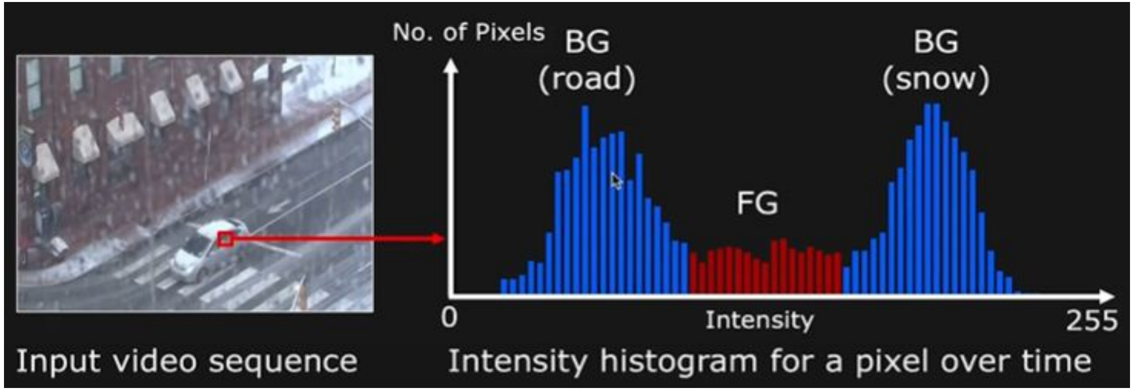
$$0 \leqslant \pi_k \leqslant 1\,, \quad \sum_{k=1}^{K} \pi_k = 1\,,$$

$$\boldsymbol{\theta} := \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k : \ k = 1, \ldots, K\}$$

Gaussian Mixture Models

Gaussian Mixture Models - Object Tracking





Gaussian Mixture Models - Object Tracking

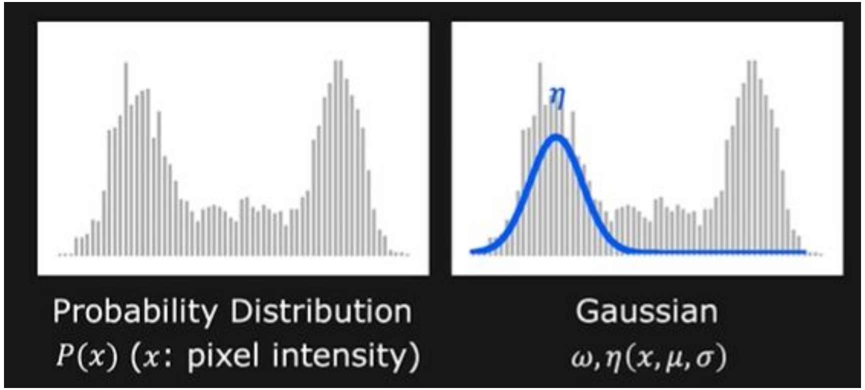
$$\boxed{P(\mathbf{X}) \cong \sum_{k=1}^{K} \omega_k \eta_k(\mathbf{X}, \boldsymbol{\mu}_k, \Sigma_k)} \quad \text{such that} \quad \sum_{k=1}^{K} \omega_k = 1$$

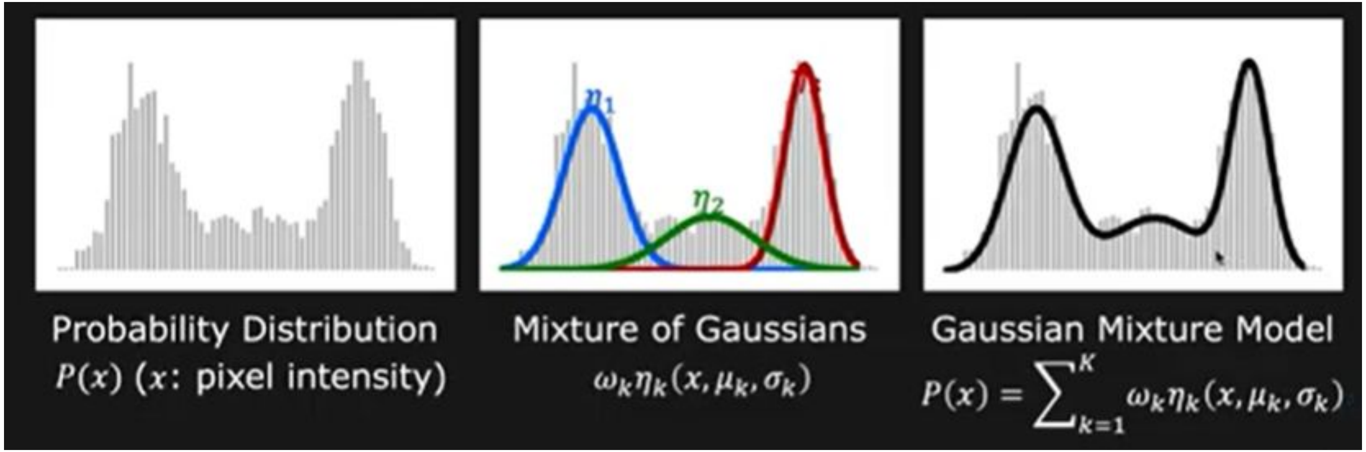
$$\text{where: } \boxed{\eta(\mathbf{X}, \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T(\Sigma)^{-1}(\mathbf{X}-\boldsymbol{\mu})}}$$

$$\text{Mean } \boldsymbol{\mu} = \begin{bmatrix} \mu_r \\ \mu_g \\ \mu_b \end{bmatrix} \quad \text{Covariance matrix } \Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

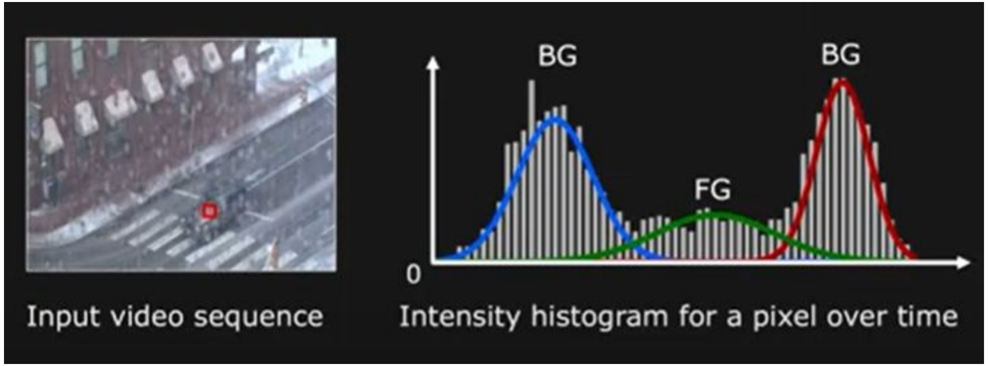
(can be a full matrix)

Gaussian Mixture Models - Object Tracking





Gaussian Mixture Models - Object Tracking



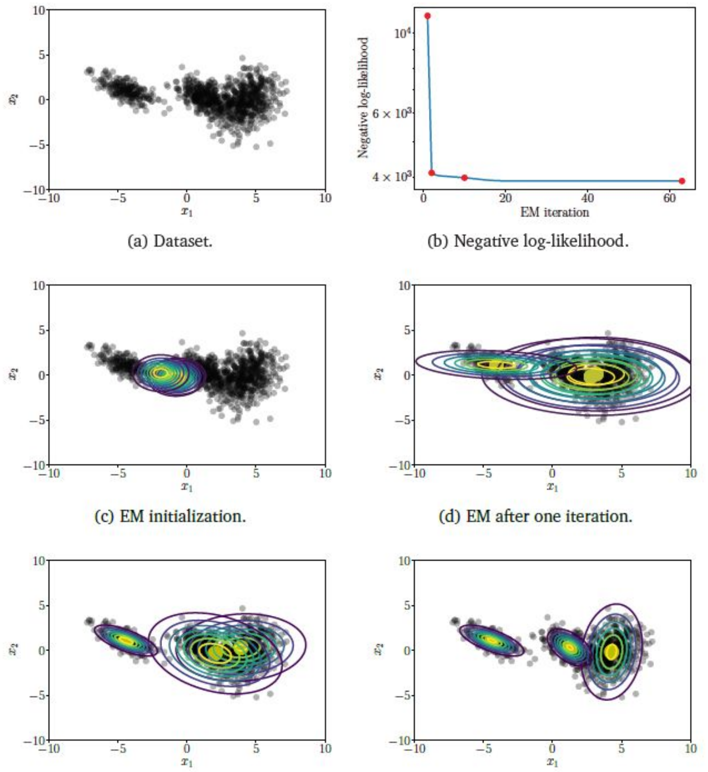
$\uparrow \frac{\omega}{\sigma}$ **Background** $\frac{\pi}{\|\Sigma\|^2}$

$\downarrow \frac{\omega}{\sigma}$ **Foreground**

.UBAfiuba FACULTAD DE INGENIERÍA

GMM y EM - JAMBOARD

Gaussian Mixture Models - Teoría



(a) Dataset.

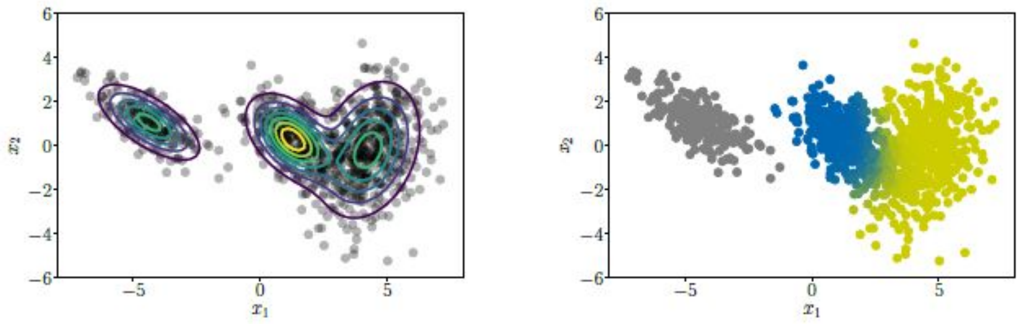
(b) Negative log-likelihood.

(c) EM initialization.

(d) EM after one iteration.

(e) EM after 10 iterations.

(f) EM after 62 iterations.



(a) GMM fit after 62 iterations.

(b) Dataset colored according to the responsibilities of the mixture components.

Notebooks

Bibliografía

- Mathematics for Machine Learning | Deisenroth, Faisal, Ong
- Pattern Recognition and Machine Learning | Bishop
- Gaussian Mixture Model | John McGonagle, Geoff Pilling, Andrei Dobre
- Expectation-Maximization Algorithms | Stanford CS229: Machine Learning
- First Principles of Computer Vision| Computer Science Department, School of Engineering and Applied Sciences, Columbia University