Introducción a la Inteligencia Artificial
Clase 7

.UBA fiuba
FACULTAD DE INGENIERÍA

Índice

Clase 7

1. Motivación
   a. Aprendizaje No supervisado
   b. Aplicaciones
2. Gaussian Mixture Models
   a. Aplicaciones
   b. Formulación

Aprendizaje no supervisado

Machine Learning Supervisado	Machine Learning no Supervisado
Proceso aleatorio $\bar{X}, y$	Proceso aleatorio $\bar{X}$
¿$f_{y/\bar{x}}(y\|\bar{x})$? ⟶ Bayes y M.V.	¿$f_{\bar{x}}(\bar{x})$? ⟶ Bayes y M.V.
Inferencias, predicciones	Clusterización, Reducción Dimensionalidad

$\bar{X}$ datos

y label (etiqueta)

en supervisado podemos definir un concepto de error $e = y - \hat{y}$

métricas de NS ⇒ métricas de relajación/equilibrio

en K-means $U = \sigma_{ig} / \sigma_{eg}$

$U_{11} \sim U_{12} \sim U_{13} \sim U_{21} \sim \dots$

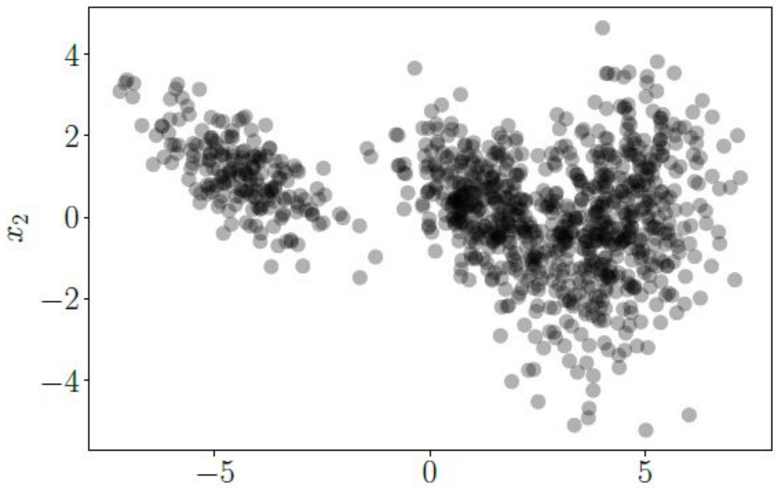
Aprendizaje no supervisado

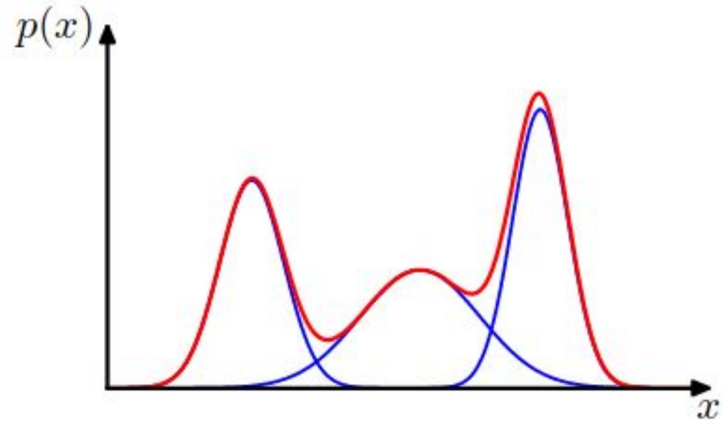
Aplicaciones Generales

- Data Mining
- Pattern Recognition
- Statistical Analysis

Aplicaciones Específicas

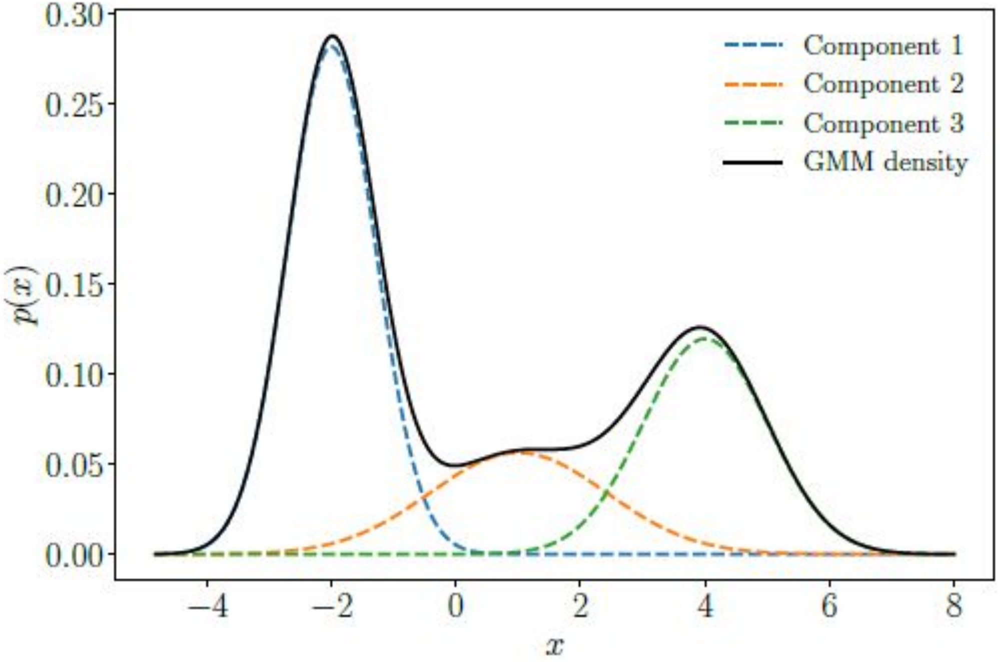
- Density Estimation
- Clustering
- Anomaly Detection
- Object Tracking
- Speech Feature Extraction





Formulación

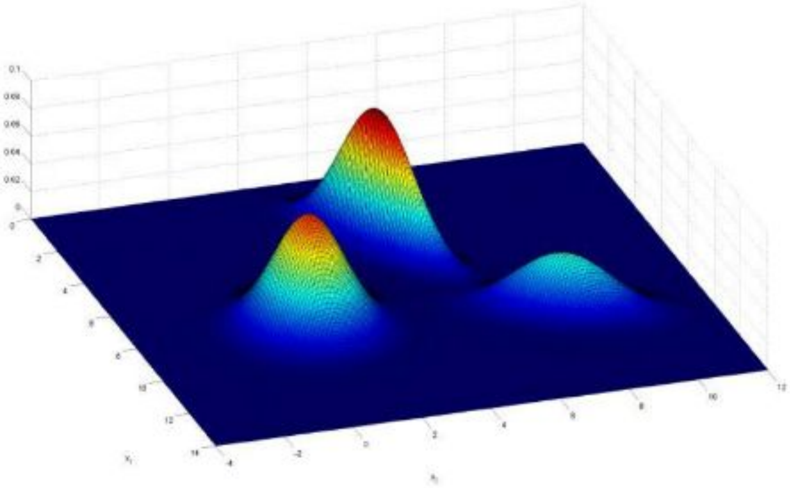
$$p(x/\hat{\theta}) = \sum_{i=1}^{N} \lambda_i \mathcal{N}(x/\mu_i, \sigma_i^2)$$

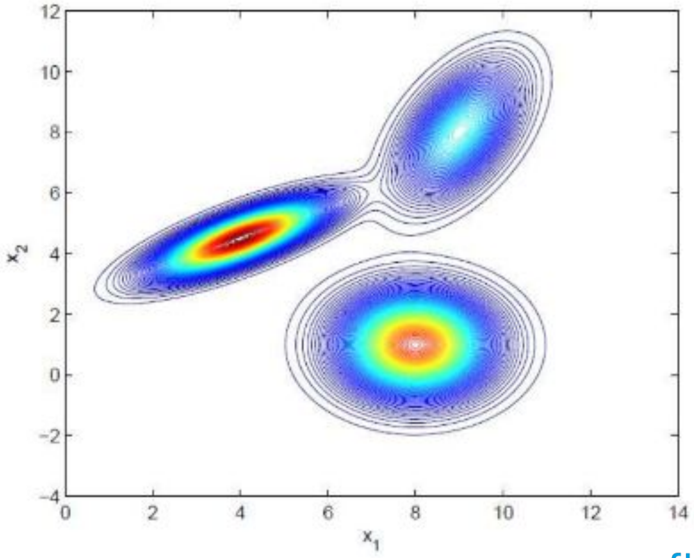


$$p(x \mid \boldsymbol{\theta}) = 0.5\mathcal{N}(x \mid -2, \tfrac{1}{2}) + 0.2\mathcal{N}(x \mid 1, 2) + 0.3\mathcal{N}(x \mid 4, 1)$$

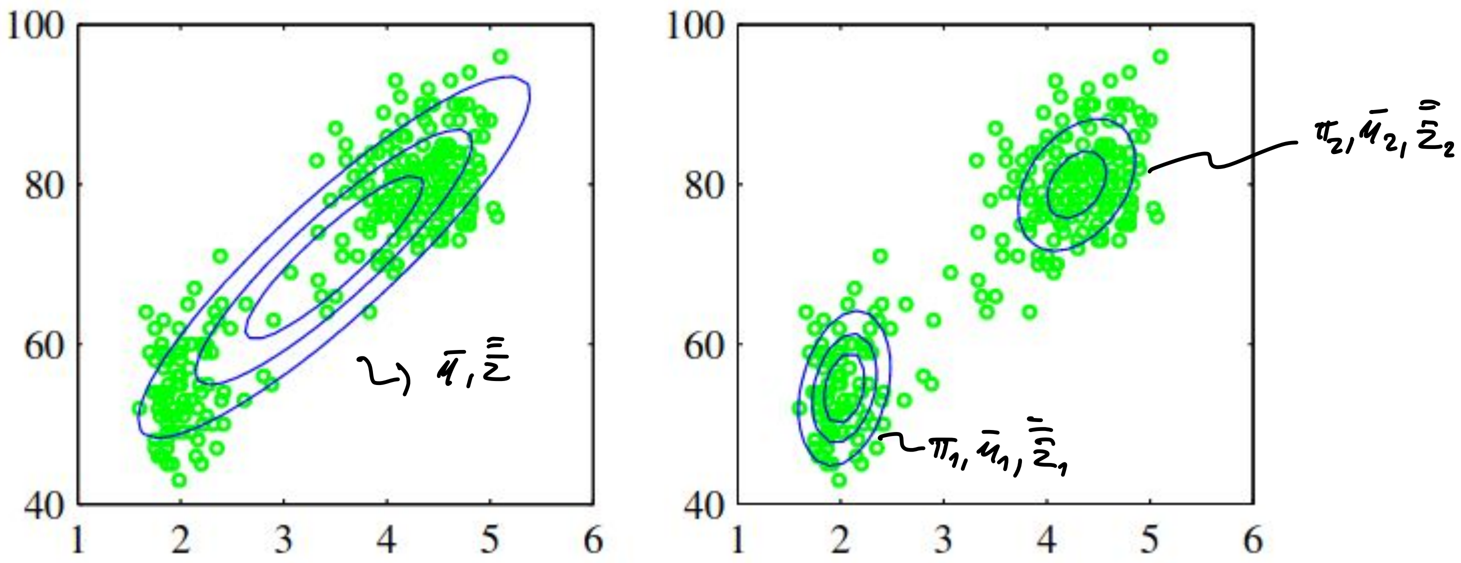
Formulación

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



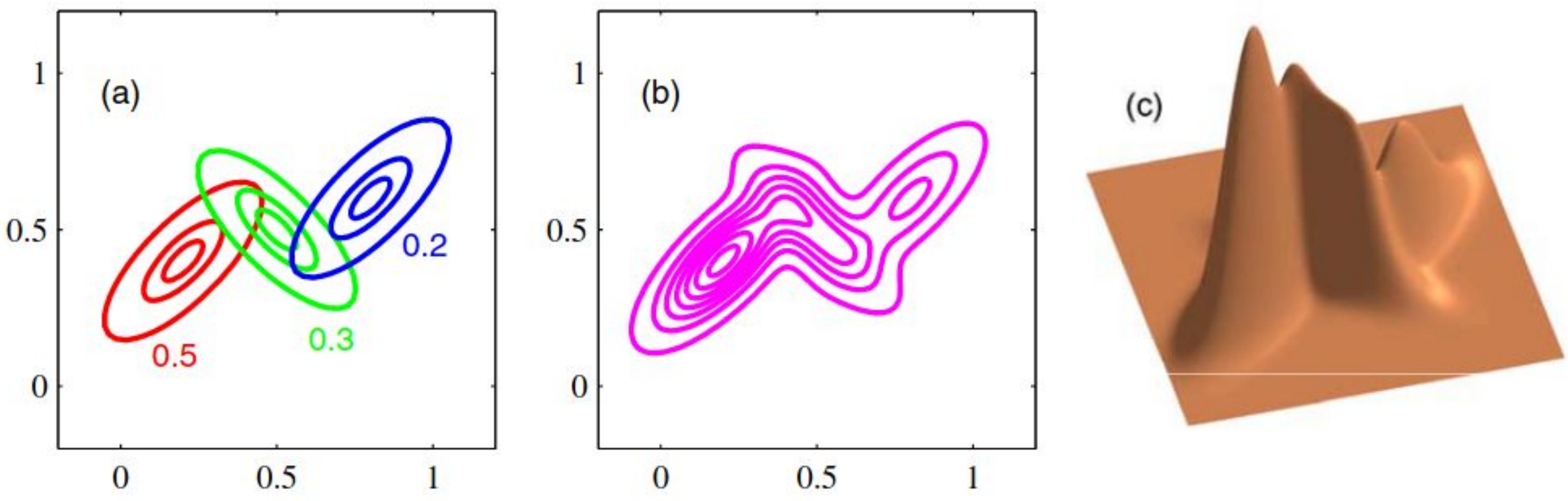


Gaussian Mixture Models: Estudio de fenómenos naturales



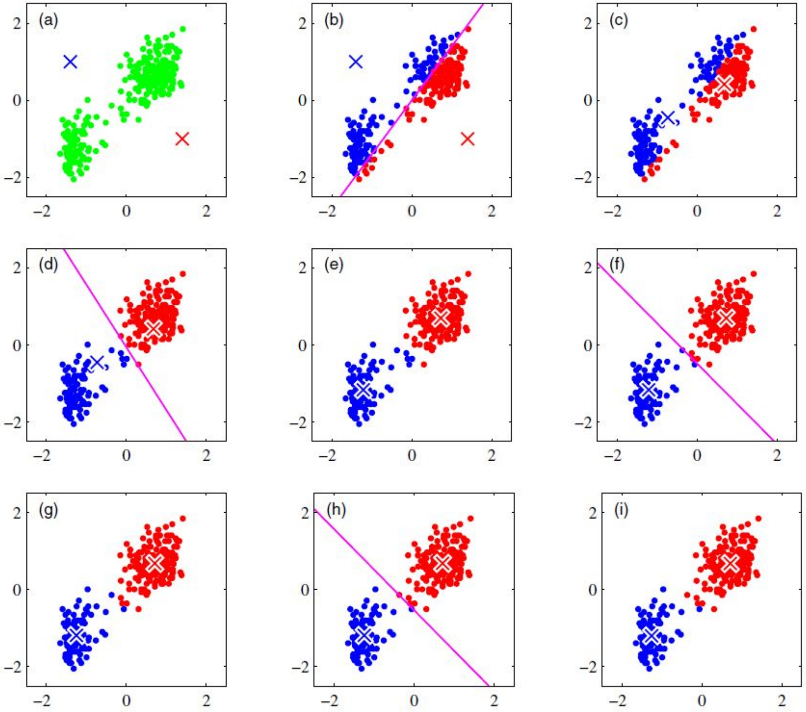
“Old Faithful” dataset. 272 mediciones de erupciones del “Old Faithful” geyser en el Parque Nacional Yellowstone. El eje horizontal representa la duración de una erupción (medida en minutos) y el vertical el tiempo hasta la próxima erupción.

Gaussian Mixture Models: Clustering

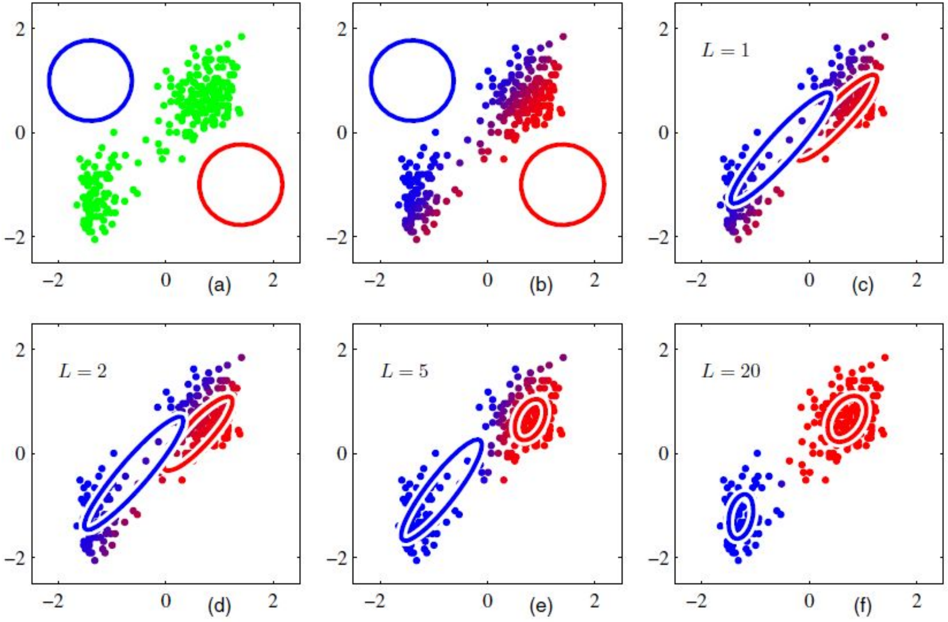


Ejemplo de Gaussian Mixture. En la imagen (a) se muestran las tres distribuciones subyacentes indicando con colores sus variables latentes. En la imagen (b) las curvas de nivel de la distribución conjunta y en la (c) la densidad.

Gaussian Mixture Models: Clustering

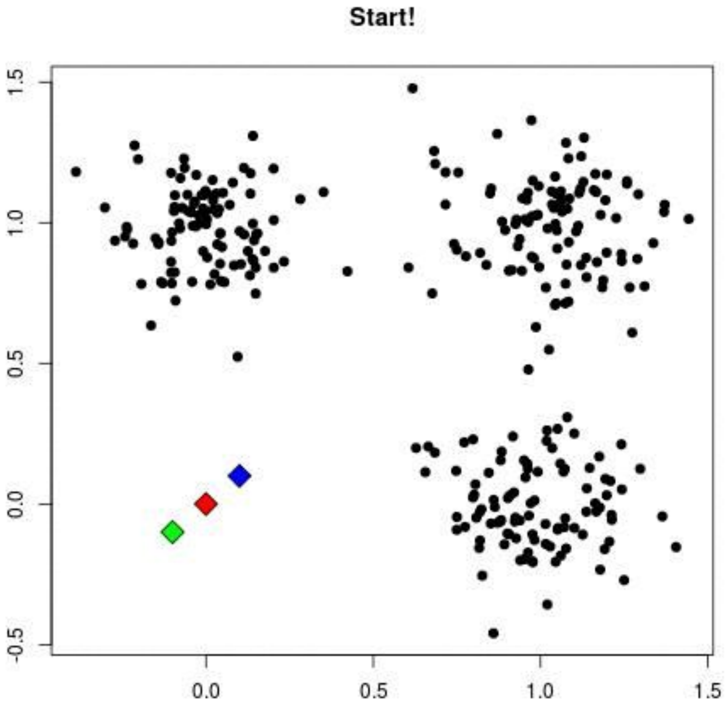


KMeans

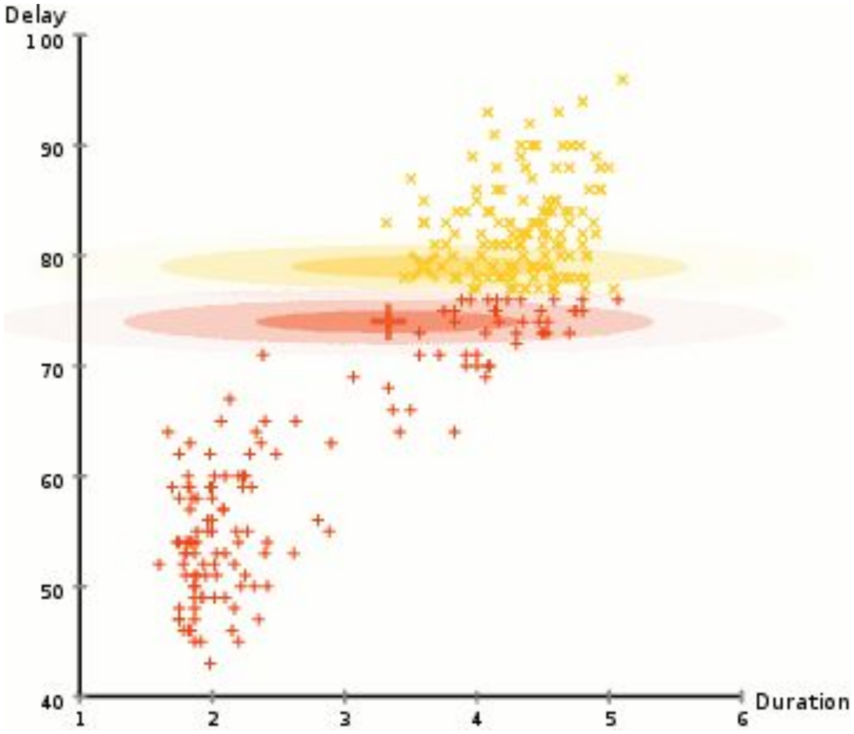


GMM

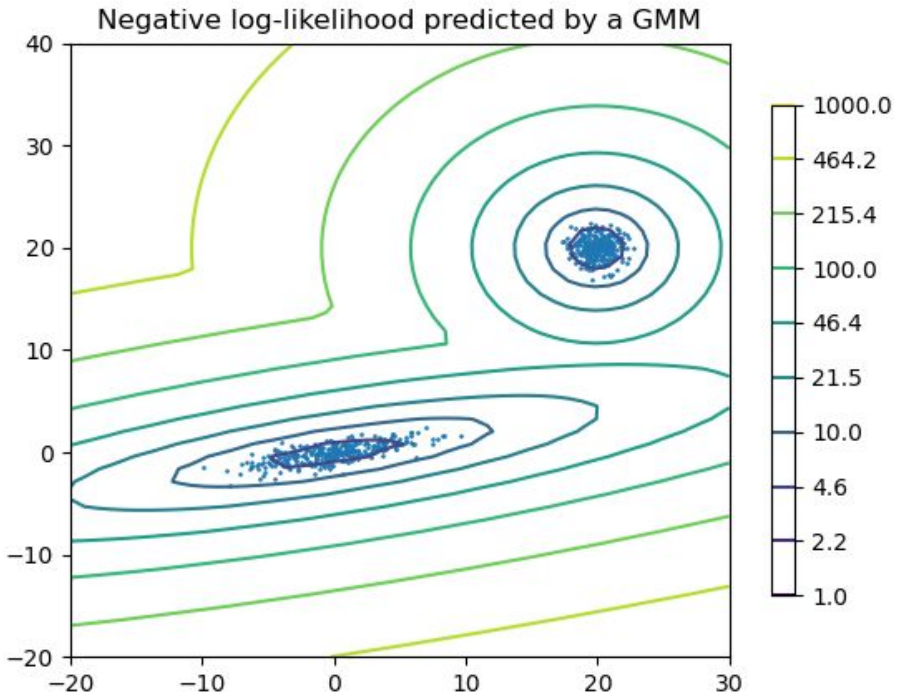
Gaussian Mixture Models - kMeans



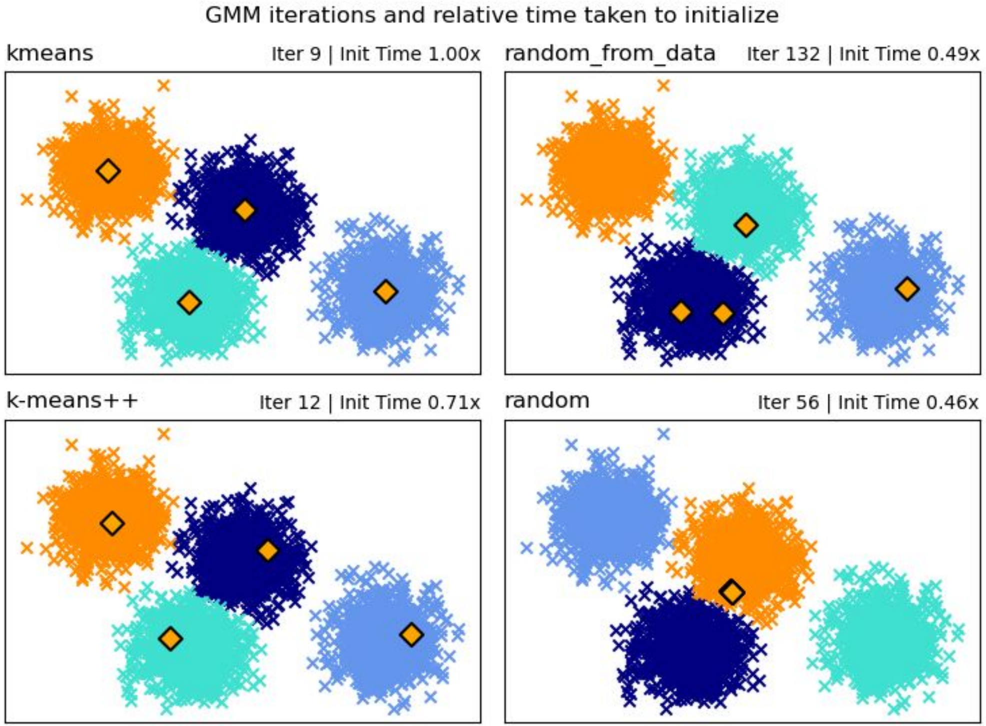
Gaussian Mixture Models: Clustering



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

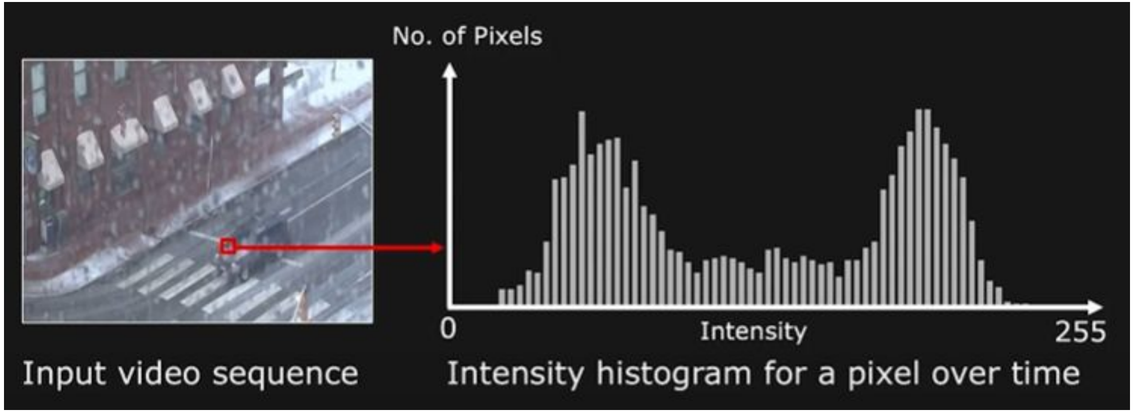
$$p(\boldsymbol{x} \,|\, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\big(\boldsymbol{x} \,|\, \boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big)$$

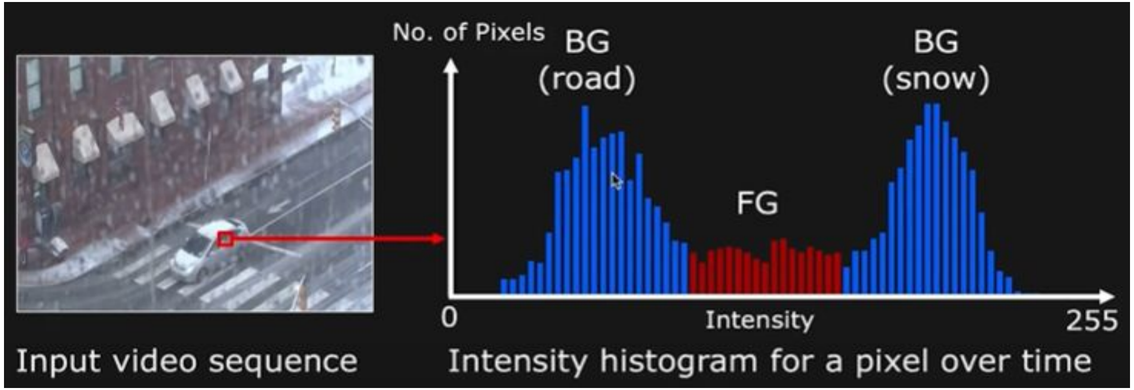
$$0 \leqslant \pi_k \leqslant 1\,, \quad \sum_{k=1}^{K} \pi_k = 1\,,$$

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

Gaussian Mixture Models

Gaussian Mixture Models - Object Tracking





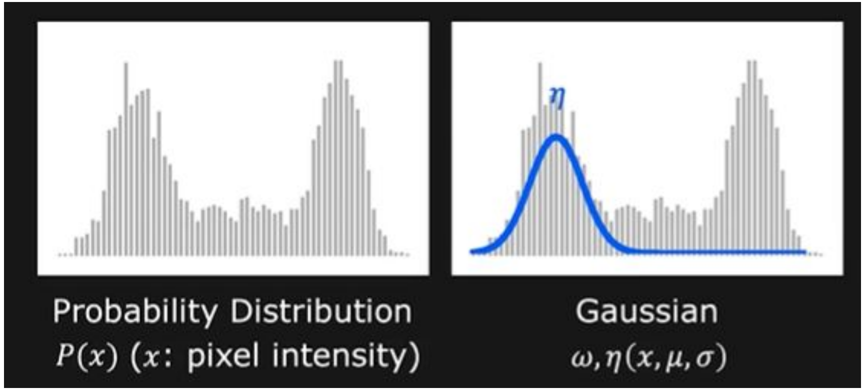
Gaussian Mixture Models - Object Tracking

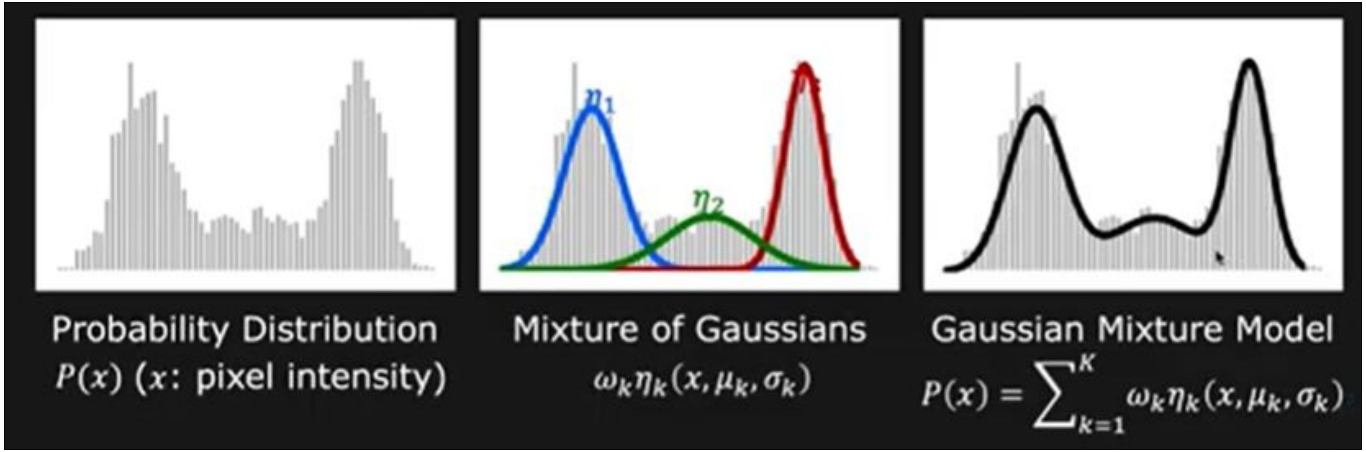
$$\boxed{P(\mathbf{X}) \cong \sum_{k=1}^{K} \omega_k \eta_k(\mathbf{X}, \boldsymbol{\mu}_k, \Sigma_k)} \quad \text{such that} \quad \sum_{k=1}^{K} \omega_k = 1$$

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

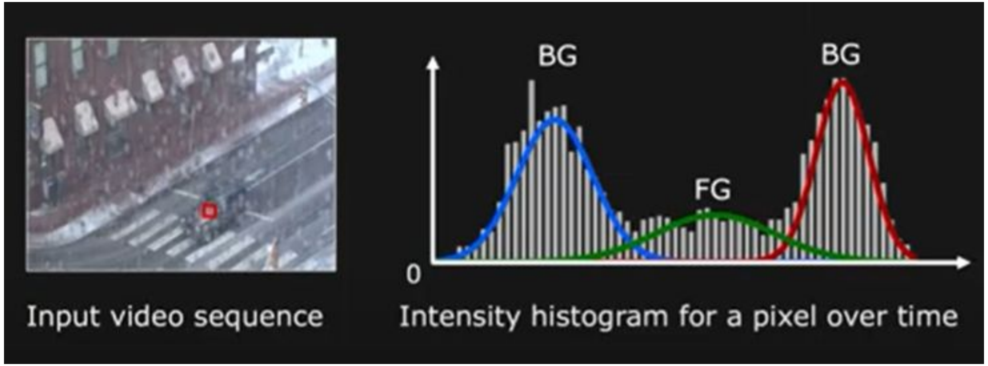
Mean $\boldsymbol{\mu} = \begin{bmatrix} \mu_r \\ \mu_g \\ \mu_b \end{bmatrix}$ Covariance matrix $\Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$ (can be a full matrix)

Gaussian Mixture Models - Object Tracking





Gaussian Mixture Models - Object Tracking



$\uparrow \frac{\omega}{\sigma}$ **Background**

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

GMM y EM - JAMBOARD

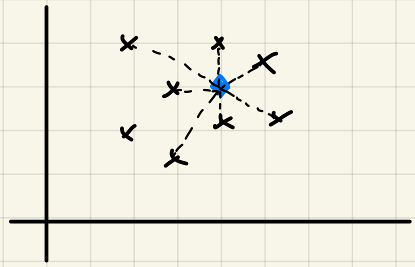
K means:

Tenemos de un dataset $D=\{X_1,\dots,X_n\}$ $X_i\in\mathbb{R}$ $\rightarrow$ Queremos encontrar K clusteres que modelen a D.

μ_K, $K=\{1,2,\dots,K\}$ que los llamo centroides. Encontrar μ_K y asignar a x_i minimizando la perdida:

$$r_{nK}=\begin{cases}1 & \text{si } x_n\in K\\ 0 & \text{o.w.}\end{cases} \Rightarrow \text{mi matriz de "labels"} \rightarrow \begin{array}{cc} \overbrace{\hphantom{1\;\;0\;\;\cdots}}^{\mu_K} & \\ \left[\begin{matrix}1 & 0 & \cdots\\ 1 & 0 & \cdots\\ 0 & 1 & \cdots\\ \vdots & \vdots & \cdots\end{matrix}\right] & \Big\} x_i\end{array}$$

$$J=\sum\sum r_{nK}\,\|x_n-\mu_K\|^2$$



Pasos a seguir:

1. inicializamos centroides.
2. Asignar los puntos según distancia $\Rightarrow$ minimizar J respecto a r_{nK} manteniendo μ_K fijos.
3. minimizar J respecto a μ_K, fijando r_{nK} (Actualización de centroides)

$$\mu_K=\frac{1}{\sum_n r_{nK}}\cdot\sum_K r_{nK}\,x_K=\sum_{n\mid x_n\in K} x_n\cdot\frac{1}{N_K}$$

Primera variación $\rightarrow$ dist euclidea $\quad \tilde{J}=\sum_n\sum_K r_{nK}\cdot N(x_n,\mu_K)$

K-meloids ; K-proto

Gaussian Mixture Model (GMM)

$$p(x)\sim\sum\pi_K\cdot\mathcal{N}(x/\mu_K,\sigma_K)\;;\;\sum_K\pi_K=1$$

Consideramos Z VA. $\quad Z_K\in\{0,1\}\wedge\sum_K z_K=1$

Definimos la conjunta $P(x,z)=P(x|z)\,P(z)$

$P(z_K=1)=\pi_K \rightarrow$ prob. del centro K $\Rightarrow \sum\pi_K=1$

$P(x/z_K) \sim \mathcal{N}(\mu_K, \Sigma_K) \rightsquigarrow$ según K la distrib. cambia.

la distrib. observada $\Rightarrow p(x) = \sum \pi_K \cdot \mathcal{N}(\mu_K, \Sigma_K)$

tenemos de dato $\to$ conjunta $P(x,z) = P(x|z)\, p(z)$

$\to$ marginal de z $\quad P(z) = [\pi_1, \pi_2, \ldots, \pi_K]$

$\to$ ① cond. $\quad P(x|z) \sim \mathcal{N}$

$\to$ marg de x $\quad P(x) = \sum \ldots$

Partimos de la cond. ①:

$$\gamma(z_K) = p(z_K = 1/x) = \frac{P(z_K=1) \cdot P(x|z_K=1)}{\sum_{j=1}^{K} p(z_j=1) \cdot P(x|z_j=1)} = \frac{\pi_K\, N(\mu_K, \Sigma_K)}{\sum_{j=1}^{K} \pi_j\, N(\mu_j, \Sigma_j)}$$

log likelihood: $\to \ln p(x|\pi,\mu,\Sigma) = \sum_n \ln \left\{ \sum_K \pi_K\, N(\mu_K, \Sigma_K) \right\}$

max likelihood $\rightsquigarrow$ no tiene solución cerrada

$\rightsquigarrow$ $K!$ soluciones.

Para resolver esto utilizamos Expectation Maximization

Derivar llk

$\partial_\mu \mathcal{L} = 0 \Rightarrow \mu_K^* = \frac{1}{N_K} \sum^{N} \gamma(z_{nK})\, x_n$ ⓐ cant. de puntos efectivos en K. $N_K = \sum_n \gamma(z_{nK})$

$\partial_\Sigma \mathcal{L} = 0 \Rightarrow$ ⓑ

$\partial_\pi \mathcal{L} = 0 \Rightarrow$ ⓒ

ⓑ $\to \Sigma^* = \frac{1}{N_K} \sum_n \gamma(z_{nK})\, (x_n - \mu_K)(x_n - \mu_K)^t$

ⓒ $\to \pi_K^* = N_K / N$

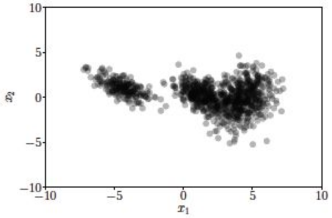
1 – inicializa μ_K, Σ_K, π_K

2 – evalua $\gamma(z_{nK})$ (construye el "posterior")

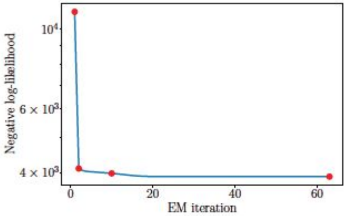
3 – actualizar los parametros en $\gamma(z)$

4 – calculamos el llk

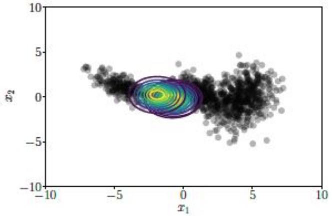
Gaussian Mixture Models - Teoría



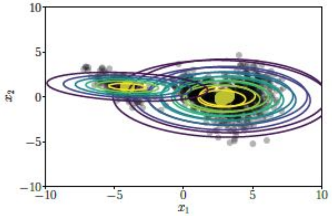
(a) Dataset.



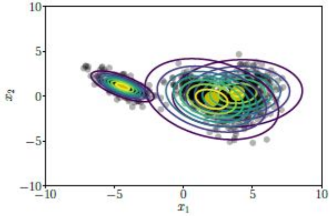
(b) Negative log-likelihood.



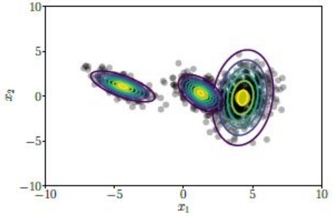
(c) EM initialization.



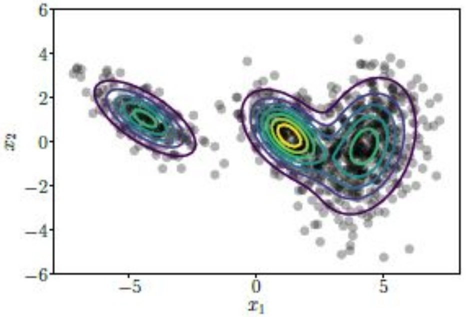
(d) EM after one iteration.



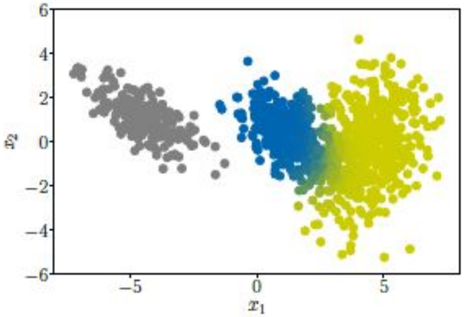
(e) EM after 10 iterations.



(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