Introducción a la Inteligencia Artificial
Clase 7

.UBAfiuba FACULTAD DE INGENIERÍA

Índice

Clase 7

1. Motivación
   a. Aprendizaje No supervisado
   b. Aplicaciones
2. Gaussian Mixture Models (Además ya vimos otros solo que no le pusimos etiqueta)
   a. Aplicaciones
   b. Formulación

.UBAfiuba FACULTAD DE INGENIERÍA

Aprendizaje no supervisado

clusterización – K means
red. dim – PCA

Acá puedo definir un concepto de error (error de inferencia)
✗

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

segmentación

$\bar{x}$ datos

y label/regresada

$\mathcal{L}(y, \hat{y})$

$y \in \mathbb{R} \to$ regresión

$y \in \mathbb{K} \to$ clasif.

tenemos métricas de relajación/equilibrio, métricas de desigualdad, etc

en Kmeans $U = \tilde{\sigma}_{ij} / \tilde{\sigma}_{eq}$

$U_1, U_2, U_3, \ldots, U_n$

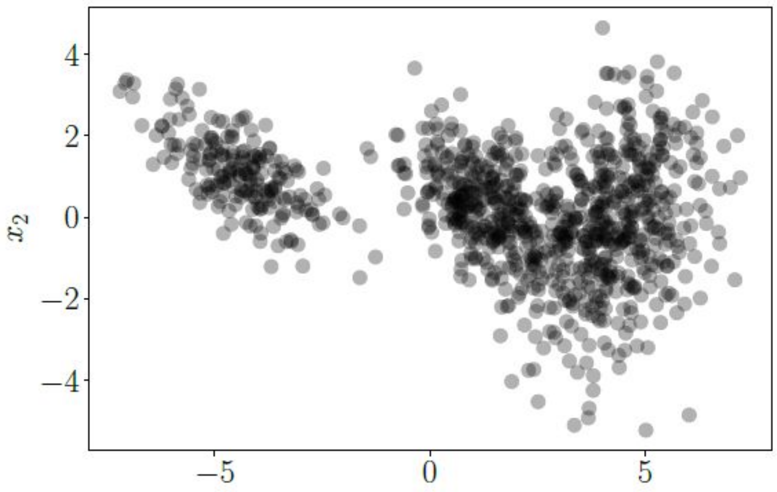
Aprendizaje no supervisado

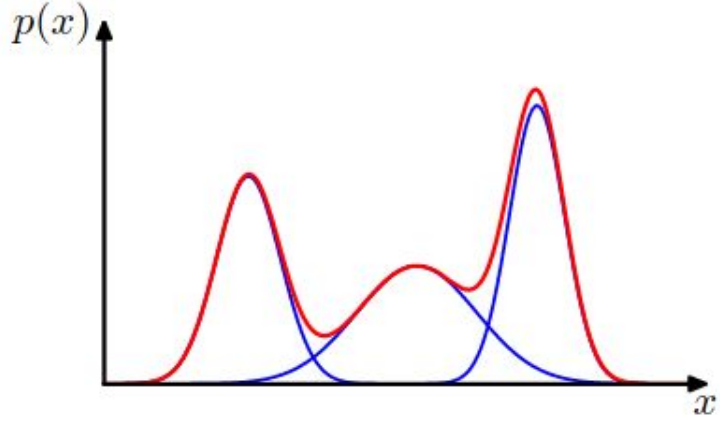
Aplicaciones Generales

- Data Mining
- Pattern Recognition
- Statistical Analysis

Aplicaciones Específicas

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

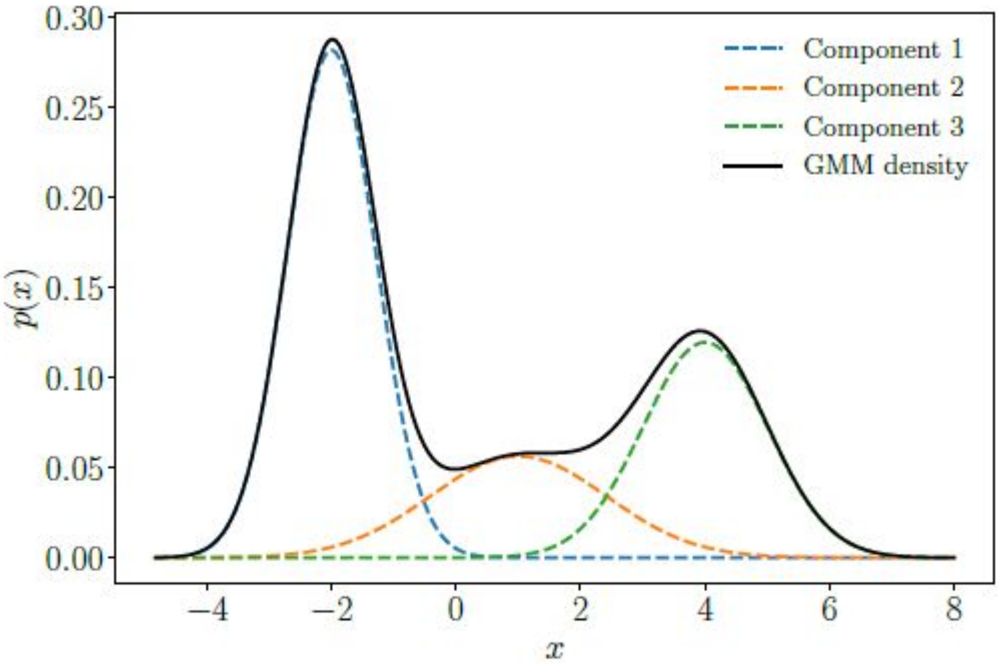




Gaussian Mixture Models

Formulación

$$p(x/\hat{\mu},\hat{\sigma}) = \sum_i \lambda_i \mathcal{N}(x/\mu_i,\sigma_i^2)$$

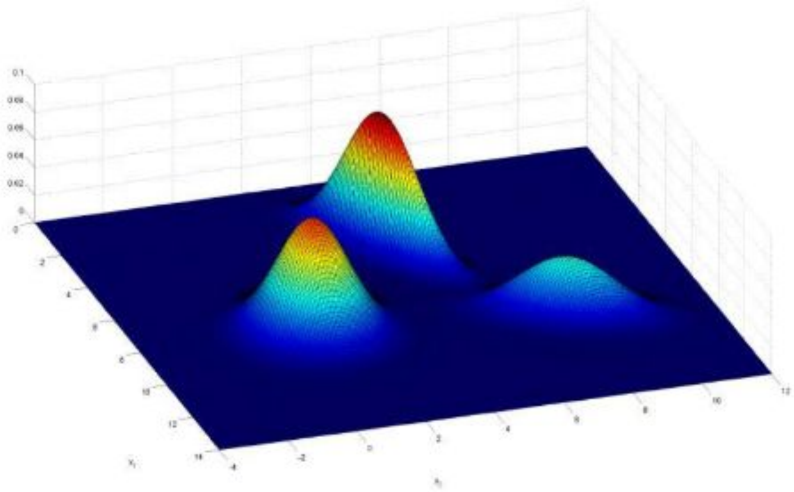


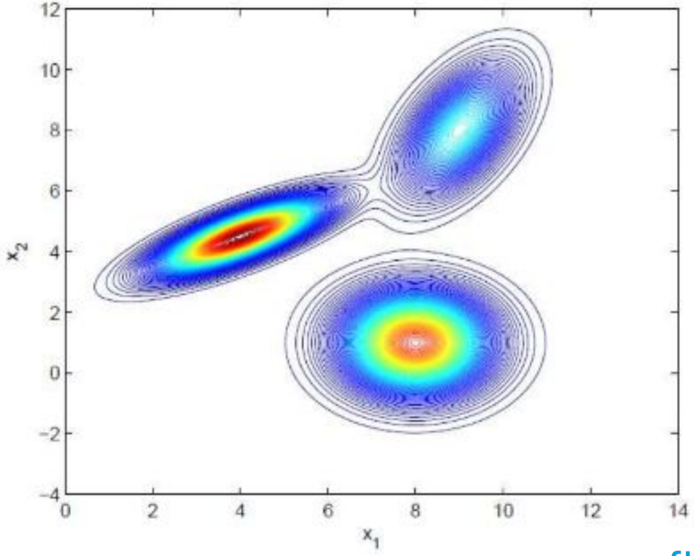
$$p(x\,|\,\boldsymbol{\theta}) = 0.5\mathcal{N}\left(x\,|\,-2,\,\tfrac{1}{2}\right) + 0.2\mathcal{N}\left(x\,|\,1,\,2\right) + 0.3\mathcal{N}\left(x\,|\,4,\,1\right)$$

.UBAfiuba FACULTAD DE INGENIERÍA

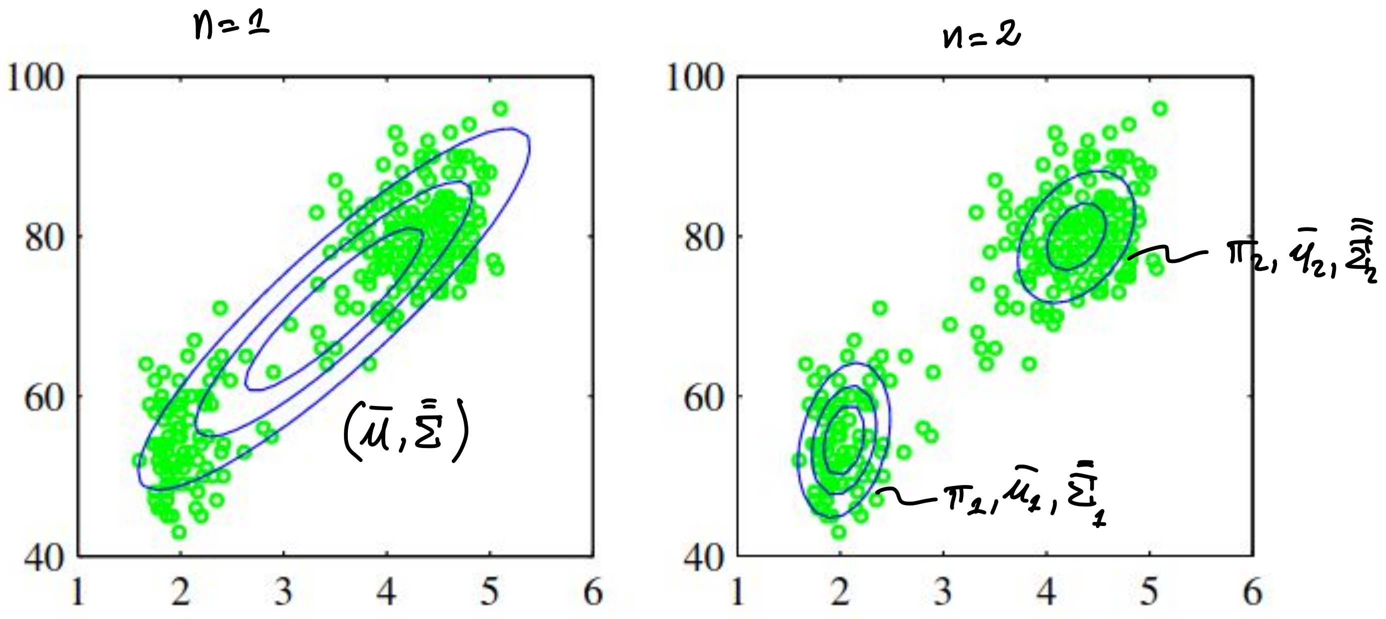
Formulación

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



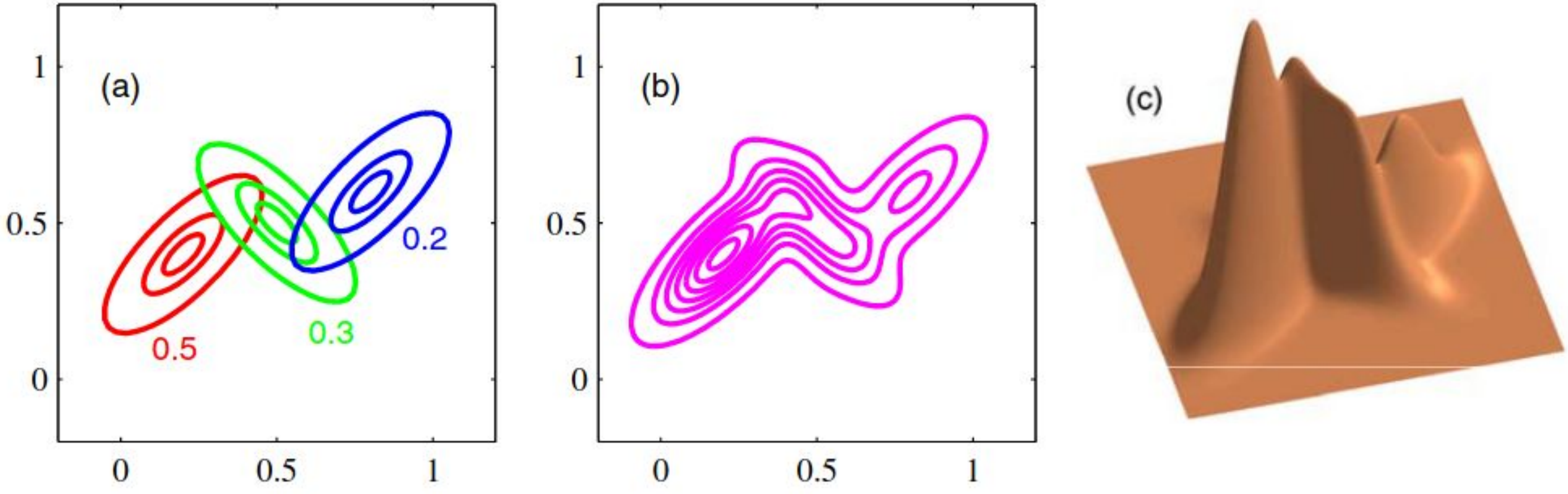


Gaussian Mixture Models: Estudio de fenómenos naturales



“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



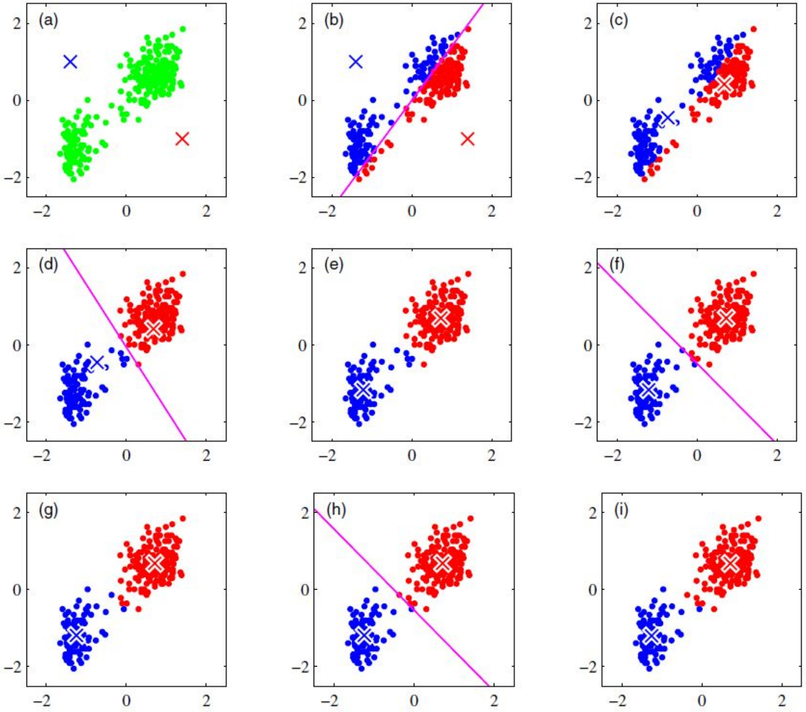
Ejemplo de Gaussian Mixture. En la imagen (a) se muestran las tres distribuciones subyacentes indicando con colores sus variables latentes. En la imagen (b) las curvas de nivel de la distribución conjunta y en la (c) la densidad.

Gaussian Mixture Models: Clustering

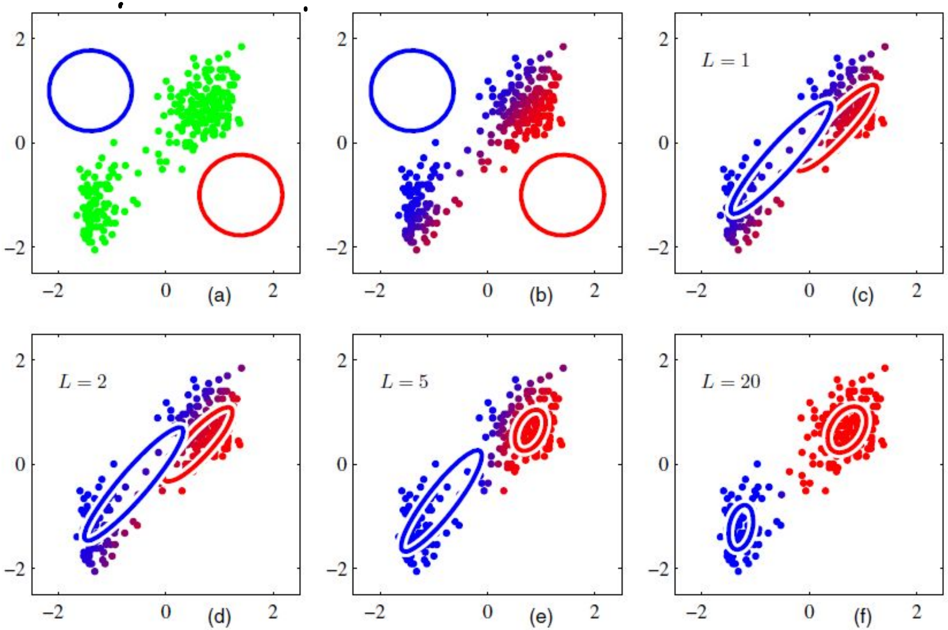
μ_1, μ_2

- $\mu_1^1 \; \mu_2^1 \; (\sigma_1^1, \sigma_2^1)$
- $\mu_1^2 \; \mu_2^2 \; (\sigma_1^2, \sigma_2^2)$

$\vdots \qquad \vdots$

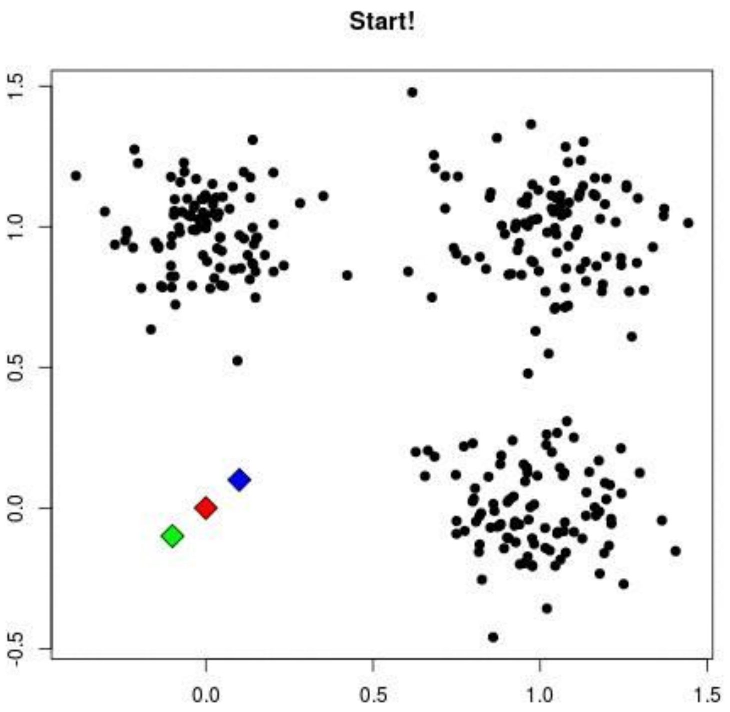


KMeans

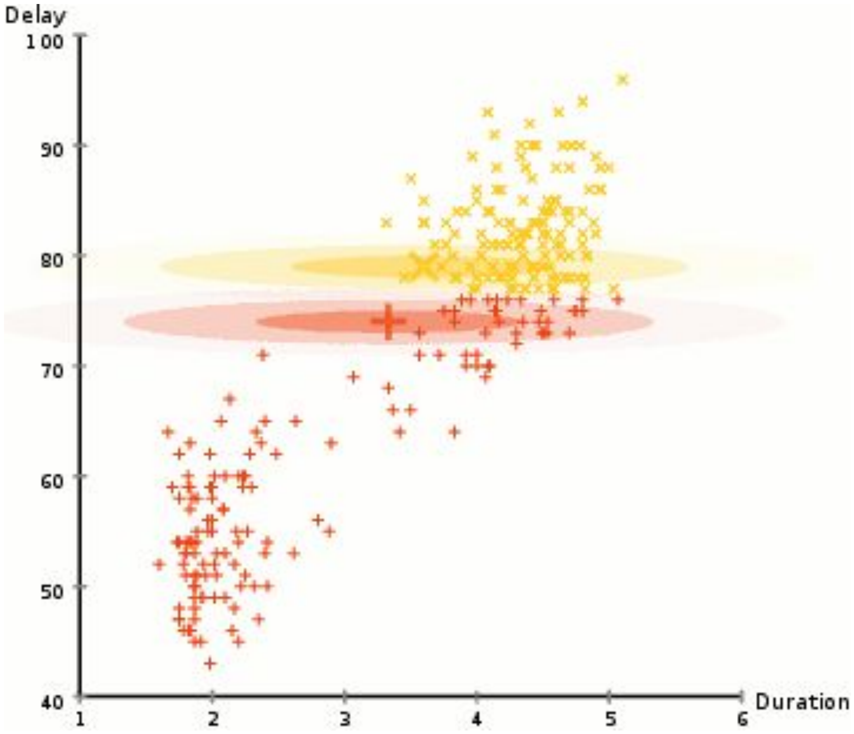


GMM

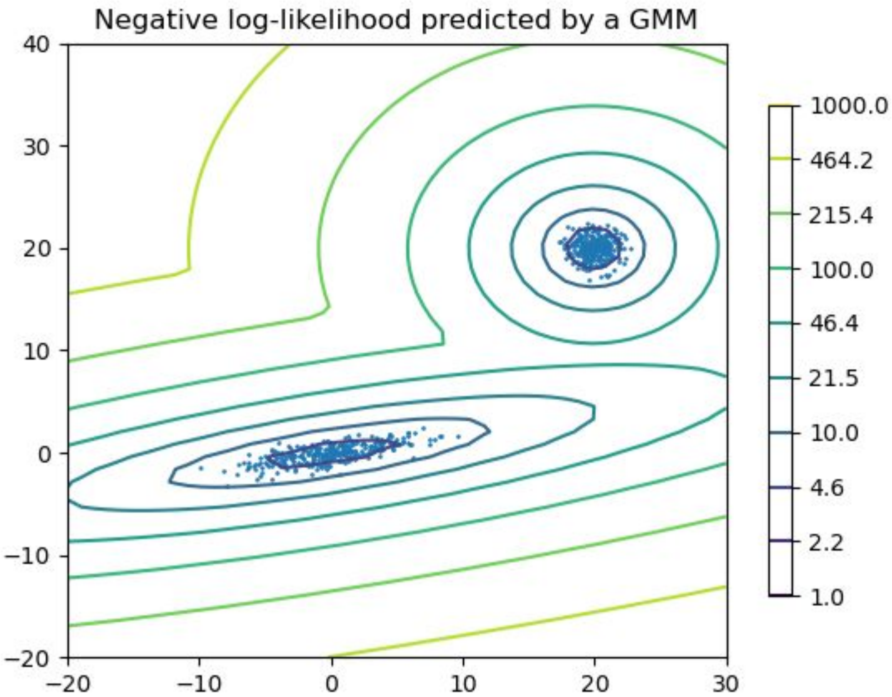
Gaussian Mixture Models - kMeans



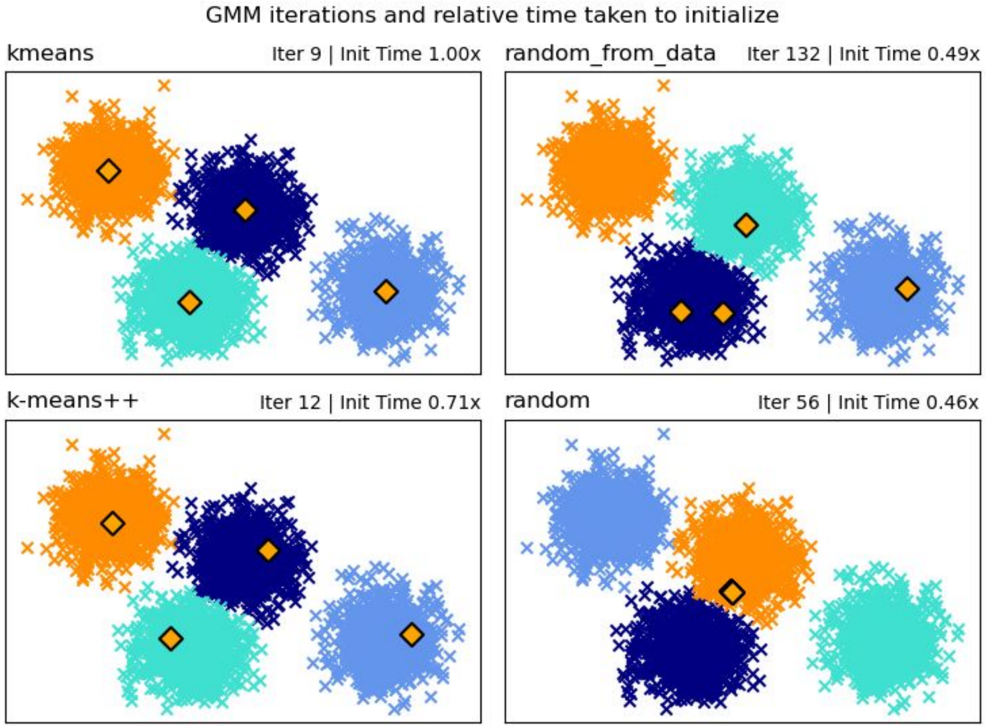
Gaussian Mixture Models: Clustering



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

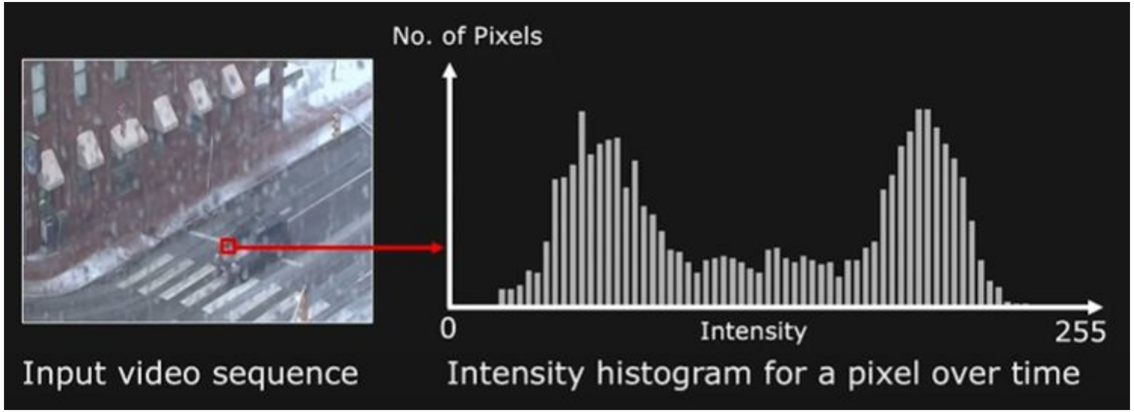
$$p(\boldsymbol{x}\,|\,\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{\mu}_k,\, \boldsymbol{\Sigma}_k\big)$$

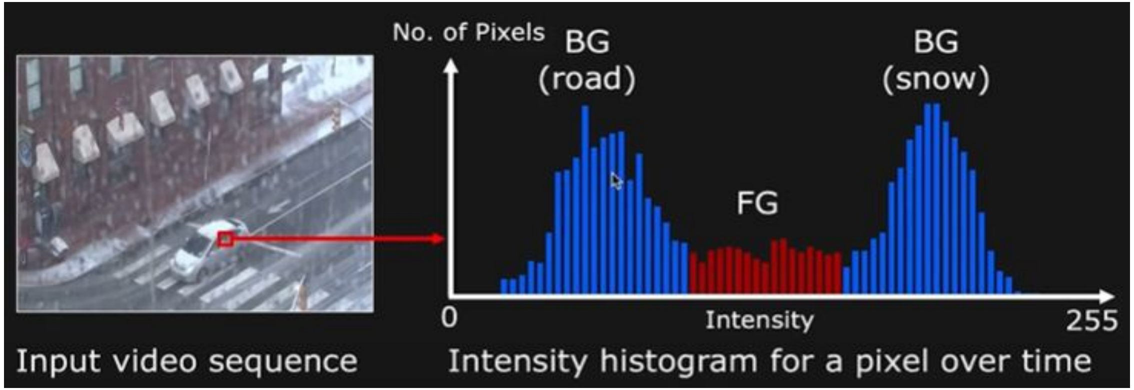
$$0 \leqslant \pi_k \leqslant 1\,, \quad \sum_{k=1}^{K} \pi_k = 1\,,$$

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

Gaussian Mixture Models

Gaussian Mixture Models - Object Tracking





.UBAfiuba FACULTAD DE INGENIERÍA

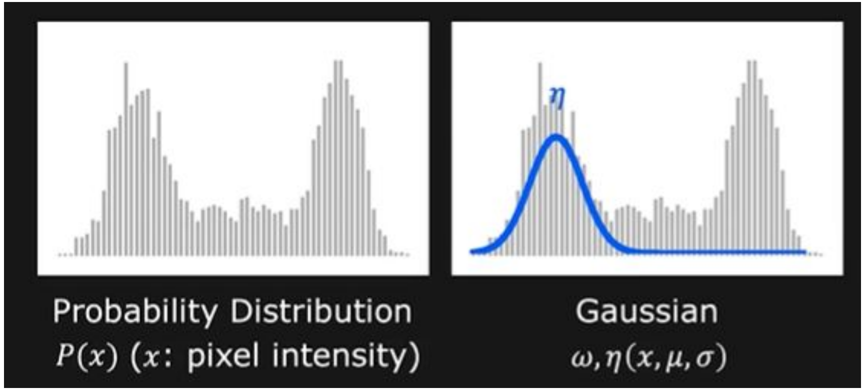
Gaussian Mixture Models - Object Tracking

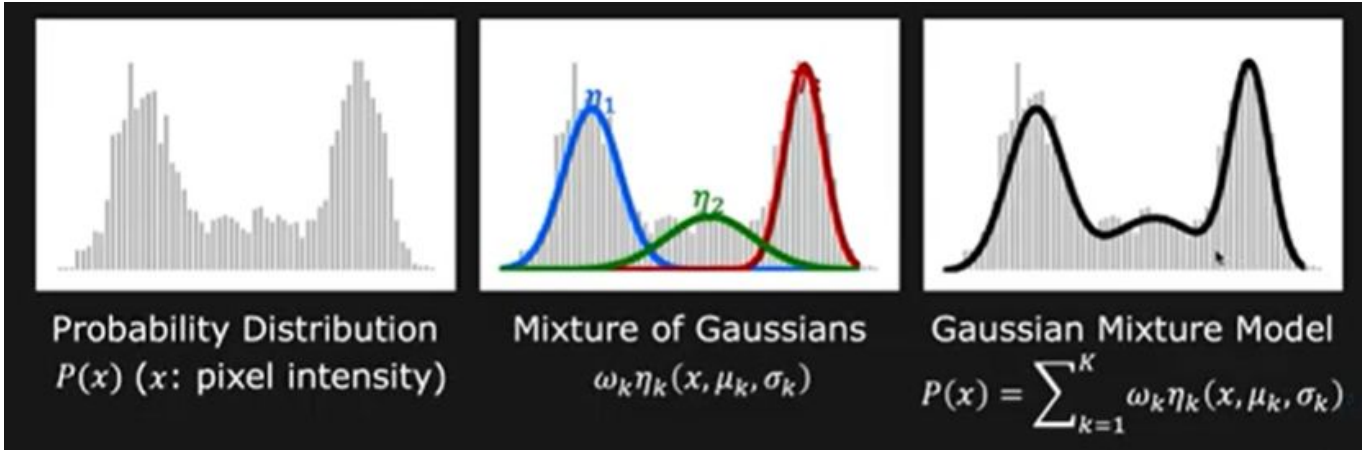
$$\boxed{P(\mathbf{X}) \cong \sum_{k=1}^{K} \omega_k \eta_k(\mathbf{X}, \boldsymbol{\mu}_k, \Sigma_k)} \quad \text{such that} \quad \sum_{k=1}^{K} \omega_k = 1$$

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

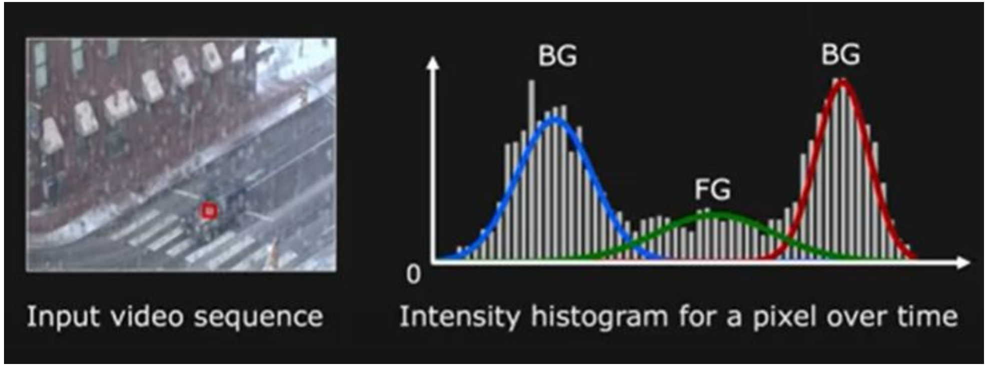
Mean $\boldsymbol{\mu} = \begin{bmatrix} \mu_r \\ \mu_g \\ \mu_b \end{bmatrix}$ Covariance matrix $\Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$ (can be a full matrix)

Gaussian Mixture Models - Object Tracking





Gaussian Mixture Models - Object Tracking



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

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

GMM y EM - JAMBOARD

K - Means:

Partimos de un dataset $D = \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^n$, con esto queremos encontrar K clusteres que modelen a D.

Vamos a considerar μ_K centros con $K \in [1, \dots, K]$ llamados centroides que van a ser nuestros parámetros a estimar. ¿Como los estimamos?

Vamos a tomar un μ_K y asignar mi dato x_i de manera de minimizar una función de perdida:

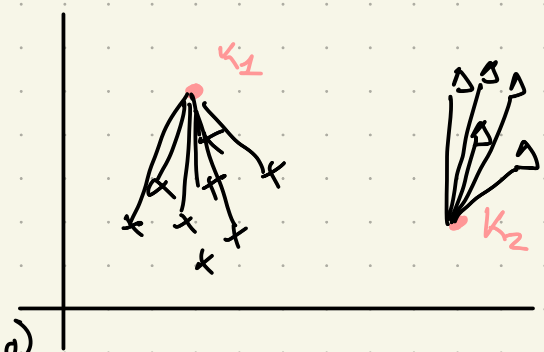
$$r_{nK} = \begin{cases} 1 & \text{si } x_n \in K \\ 0 & \text{o.w.} \end{cases} \longrightarrow$$

Esto me 'labeliza' mis datos

$$\underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}}_{\mu_K} \Bigg\} x_i$$

Con esto definimos: $$\boxed{J = \sum_n \sum_K r_{nK} \|x_n - \mu_K\|^2}$$

Pasos a seguir:



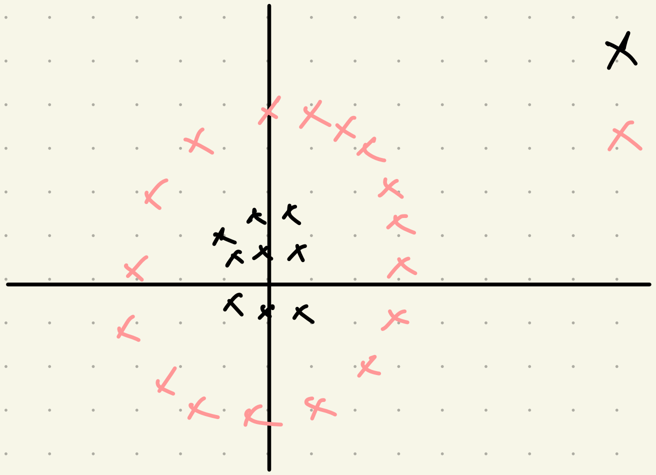
1. inicializamos μ_K (criterio)
2. Asignamos 'labels' según distancia $\Rightarrow \min_{r_{nK}} J(n,K)$ (μ_K fijo)
3. minimizar J respecto a μ_K, fijando r_{nK} (Actualización de centroides)

$$\min_{\mu_k} J(r_{n,k}, \mu_k) \rightsquigarrow \mu_k = \frac{1}{\sum_n r_{nk}} \cdot \sum_n r_{nk} x_k = \sum_{n/x_n \in k} x_n \cdot \frac{1}{N_k}$$

K means → dist euclidea $\tilde{J} = \sum \sum r_{nk} \, N(\mu_k, x_n)$

si yo vario esto → K-proto

K-meloids



x label 1 (real)

x label 2 (real)

← caso atipico en kmeans.

GMM

$$p(x) = \sum_{k} \pi_k \, \mathcal{N}(x \mid \bar{\mu}_k, \bar{\bar{\Sigma}}_k) \; ; \quad \sum_{k} \pi_i = 1$$

Consideramos z V.A. $\in \{0, 1\} \wedge \sum_{k} z_k = 1$.

Definimos $P(x, z) = P(x \mid z) \, P(z) \to P(z_k = 1) = \pi_k \to$ prob del centro k

$P(x \mid z_k) \sim \mathcal{N}(\bar{\mu}_k, \bar{\bar{\Sigma}}_k)$ (cada k tiene par. distintos)

con esto $p(x) = \sum \pi_k \cdot P(x \mid z_k)$

con esto tenemos de datos $\to$ conjunta $P(x, z) = P(x \mid z) \, P(z)$

$\to$ marginales $P(z) = [\pi_1, \dots, \pi_k]$

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

$\to$ marg $P(x) = \sum_{k} (\dots)$

partimos de ①:

$$\gamma(z_k) = p(z_k=1/X) = \frac{p(z_k=1)\; P(X \mid z_k=1)}{\sum_{j=1}^{K} p(z_j=1)\, P(X/z_j=1)} = \frac{\pi_k \; \mathcal{N}(\bar{\mu}_k, \bar{\bar{\Sigma}}_k)}{\sum_{j=1}^{K} \pi_j \, \mathcal{N}(\bar{\mu}_j, \bar{\bar{\Sigma}}_j)}$$

↑ distrib. a posteriori

log likelihood $\rightsquigarrow$ $\ln P(X \mid \bar{\pi}, \bar{\mu}, \bar{\bar{\Sigma}}) = \sum_n \ln \left\{ \sum_k \pi_k \, N(\bar{\mu}_k, \bar{\bar{\Sigma}}_k) \right\}$

$\max\limits_{\pi, \mu, \Sigma} \ell$ $\rightsquigarrow$ no tiene sol. cerrada

$\rightsquigarrow$ tenemos $K!$ soluciones.

para resolver esto vamos a usar Expectation Maximization.

consite en derivar log likelihood e igualarlo a cero.

$\partial_\theta \ell = 0$

$\hookrightarrow \partial_{\mu_k} \ell = 0 \rightarrow \mu_k^* = \frac{1}{N_k} \sum \gamma(z_{nk}) \cdot x_n$

$\partial_{\Sigma_k} \ell = 0 \rightarrow \Sigma_k^* = \frac{1}{N_k} \sum \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^t$

$\partial_{\pi_k} \ell = 0 \rightarrow \pi_k^* = N_k / N$

N_k: número efectivo de ptos en la clase k

$N_k = \sum_n \gamma(z_{nK})$

Pasos a seguir:

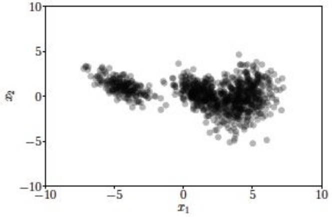
1. inicializamos μ_K, Σ_K, π_K

2. con estos param. vamos a evaluar $\gamma(z_{nk})$

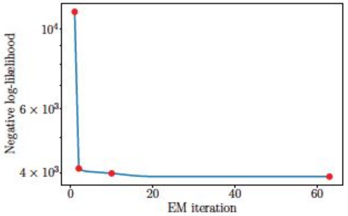
3. Actualizamos los parametros $\mu_K', \Sigma_K', \pi_K'$

4. calcular el log likelihood nuevo.

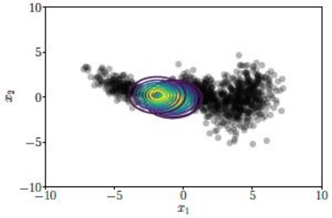
Gaussian Mixture Models - Teoría



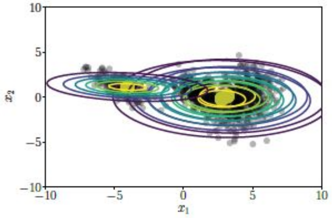
(a) Dataset.



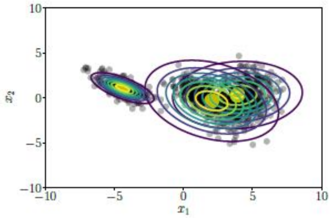
(b) Negative log-likelihood.



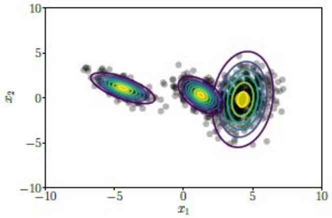
(c) EM initialization.



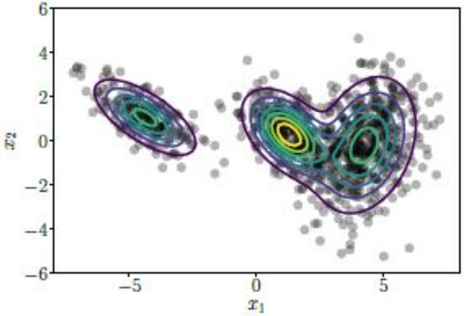
(d) EM after one iteration.



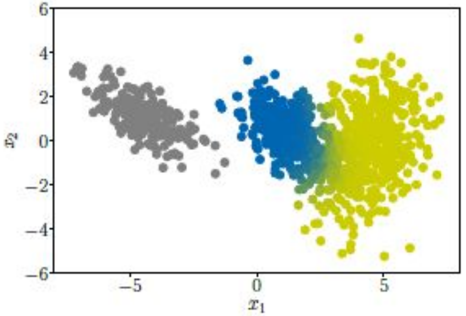
(e) EM after 10 iterations.



(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