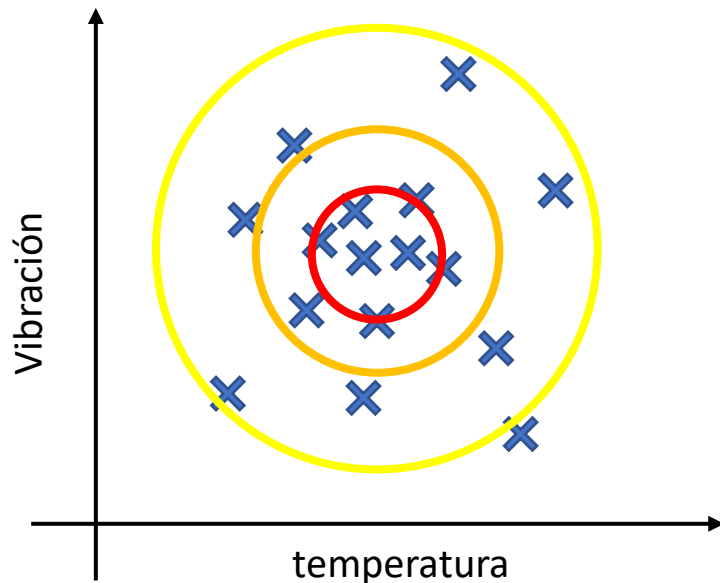


Gaussian Mixture Models

The background of the slide is a photograph of a clear night sky. The Milky Way galaxy is visible as a bright, hazy band of light stretching across the upper half of the frame. Numerous individual stars are scattered throughout the sky. In the lower portion of the image, the dark, silhouetted branches of evergreen trees are visible against the starry background.

Gaussian Models

- Un modo más “Soft” podría ser modelando cada cluster como una gaussiana.
- Ahora cada patrón no está dentro o fuera del cluster, sino que pertenece con cierta probabilidad.
- La media de la distribución podría considerarse el centro del cluster.
- Se agrega la matriz de covarianza.



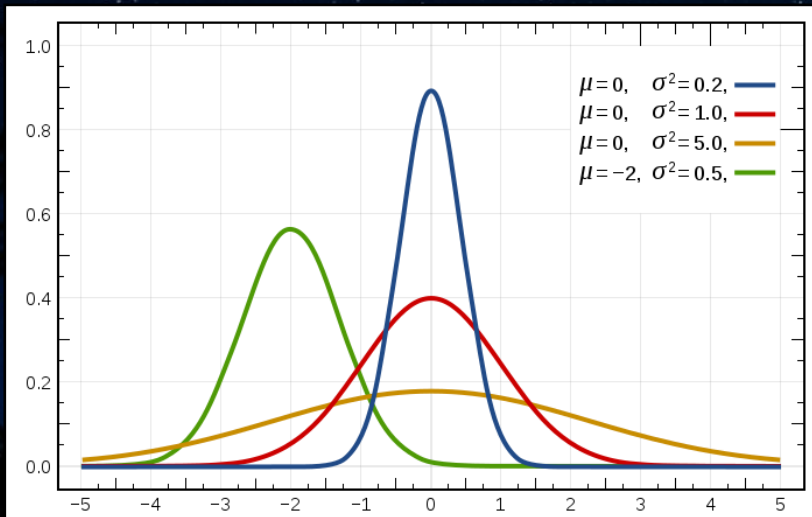
Función Gaussiana

En el caso de una dimensión:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}$$

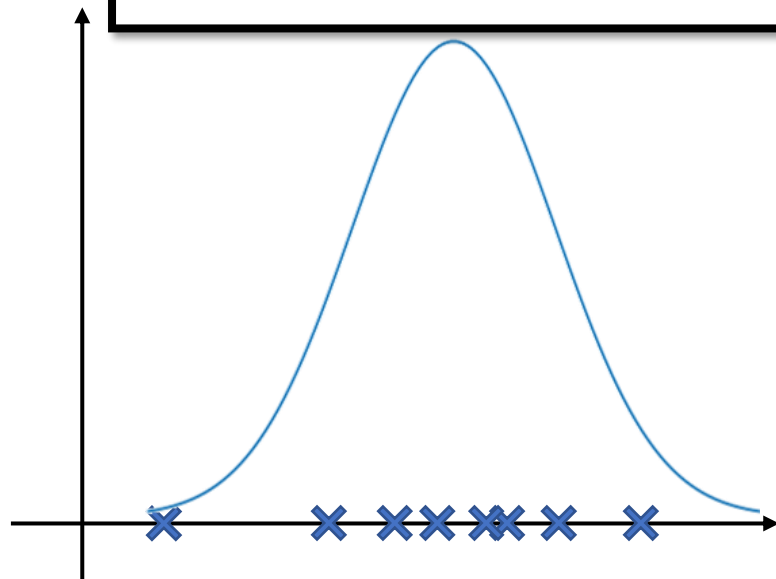
Media

Desviación estándar



$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\sigma^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n-1}$$

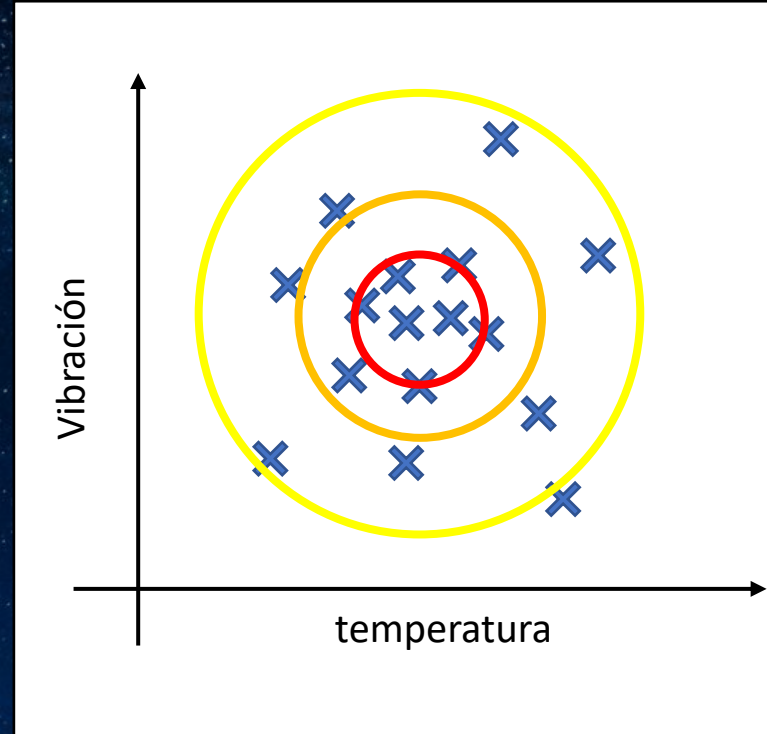
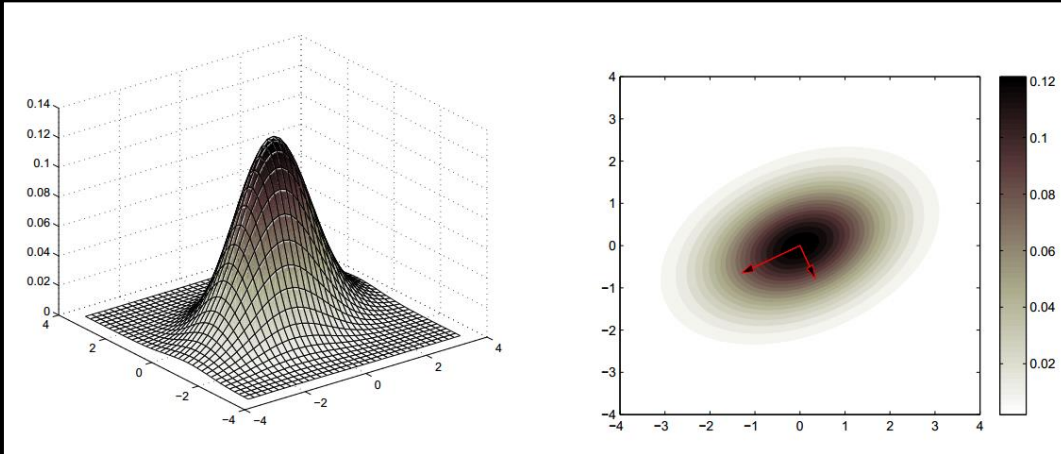


Función Gaussiana

En el caso n-dimensional:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Matriz de covarianza



Gaussian Mixture Models

La probabilidad de un dato x está dada por:

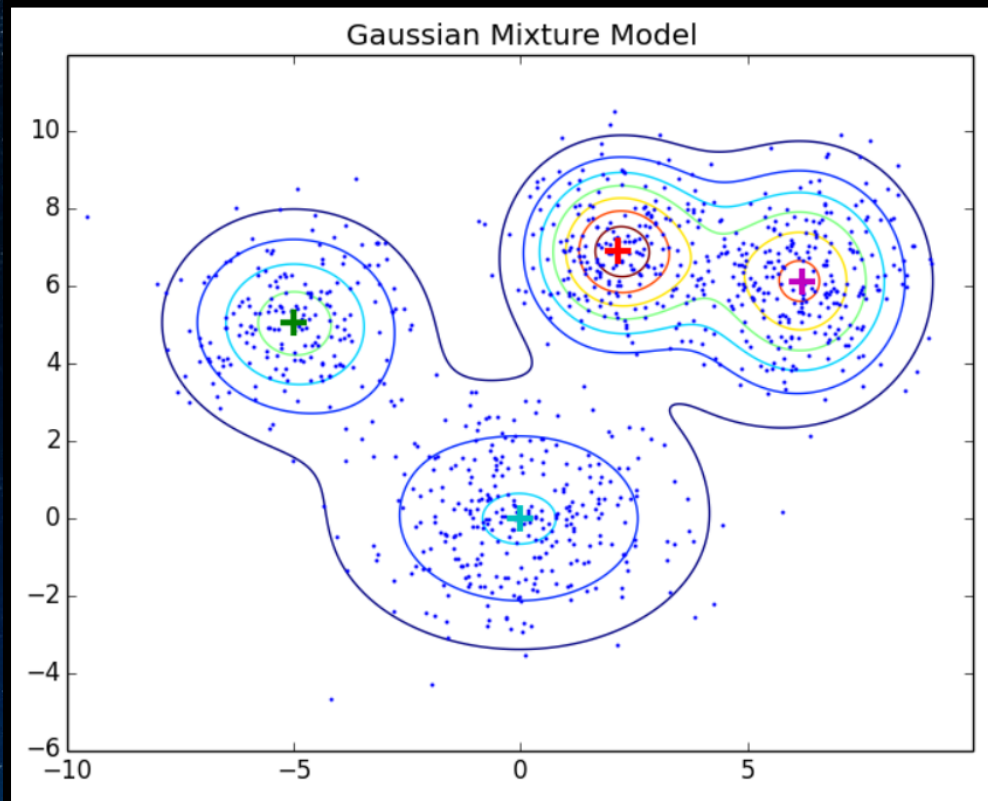
$$p(x) = \sum_{j=1}^K w_j \cdot N(x | \mu_j, \Sigma_j)$$

$$\sum_{j=1}^K w_j = 1$$

K = número de gaussianas.

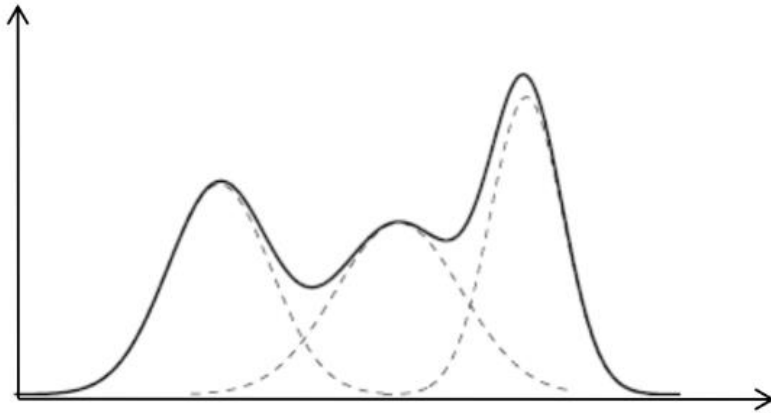
w_j = peso que se le da a cada función.

- Es posible utilizar la probabilidad general de un nuevo punto para detectar anomalías.
- También podemos utilizar el modelo como en k-mean, para identificar clusters, simplemente nos quedamos con la mayor probabilidad.

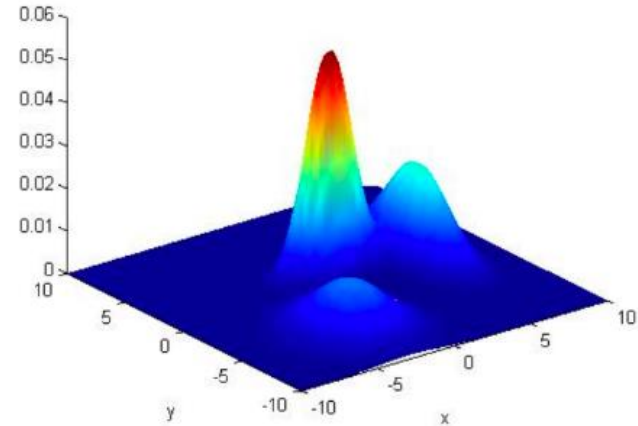


Gaussian Mixture Models

d=1:



d=2:



GMM – Algoritmo EM

¿Cómo encontrar los parámetros?

El algoritmo más clásico para optimizar los parámetros de la mixtura es el conocido como “EM” (Expectation-Maximization). El objetivo es maximizar el likelihood (verosimilitud).

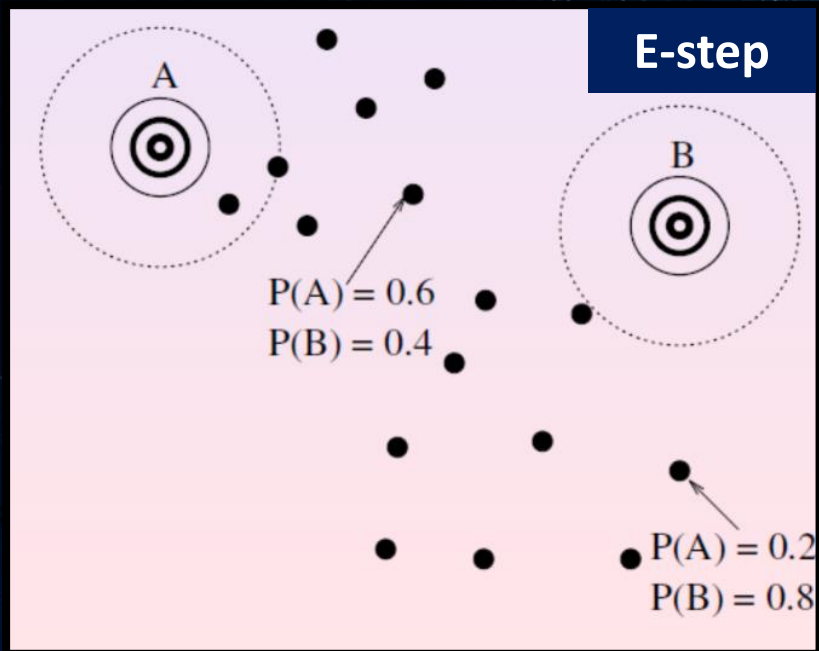
Definir cantidad de gaussianas.

En cada iteración:

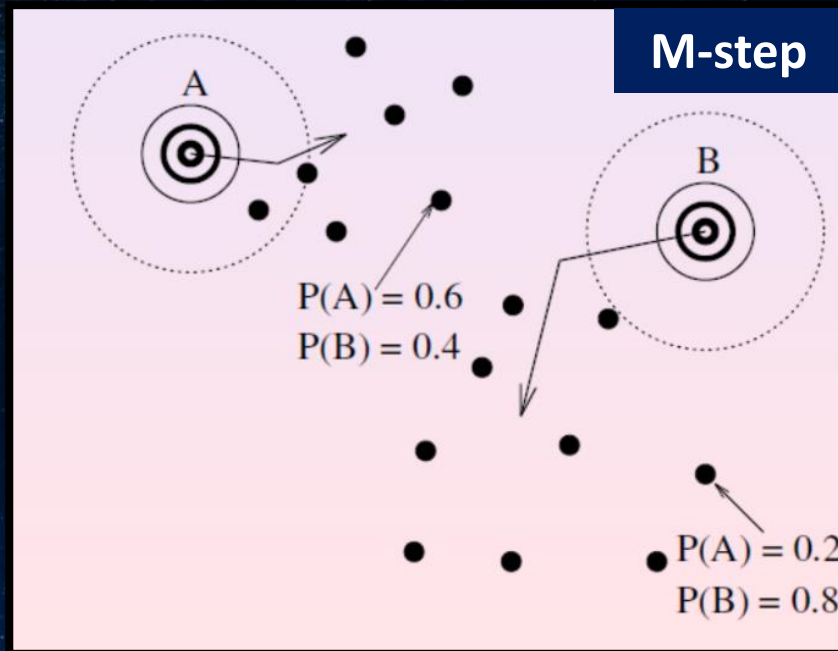
E-step: Estima la probabilidad de cada dato al conjunto de distribuciones.

M-step: recalcular media y desviación de las gaussianas.

GMM – Algoritmo EM

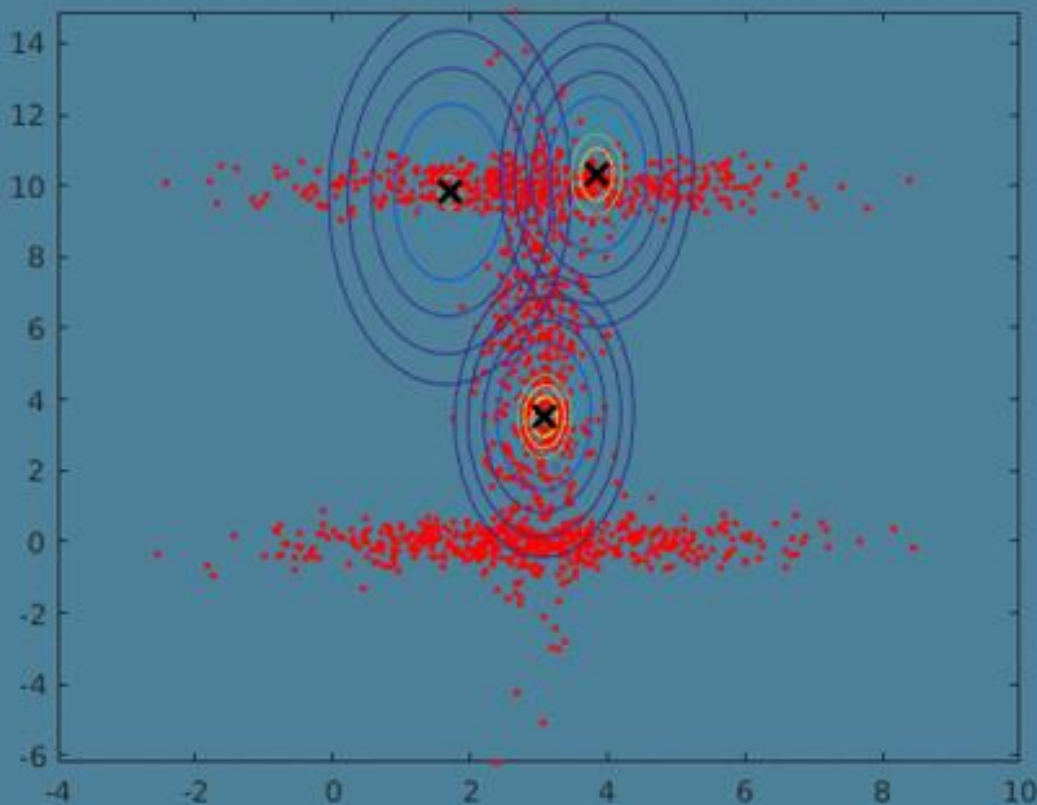
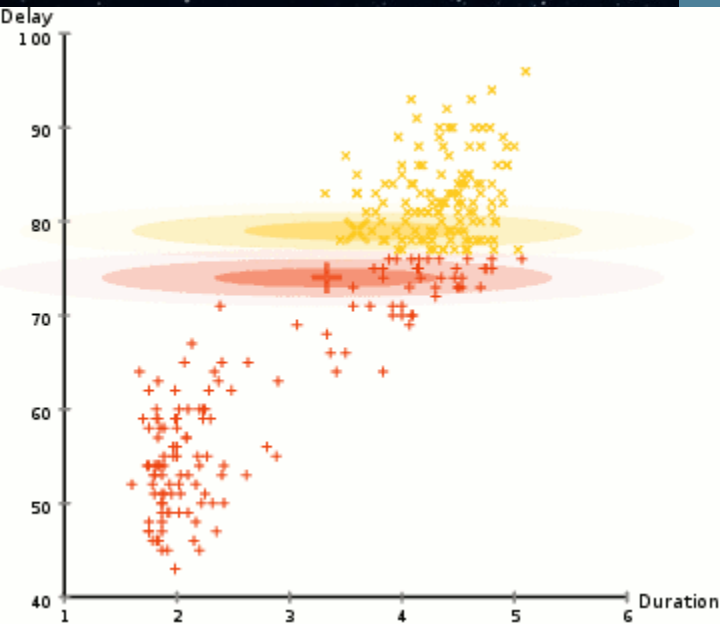


$$P(X_i \in C_k) = p(C_k|X_i) = \frac{p(C_k)p(X_i|C_k)}{p(X_i)}$$



$$\mu_A = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n}$$
$$\sigma_A^2 = \frac{w_1(x_1 - \mu)^2 + w_2(x_2 - \mu)^2 + \dots + w_n(x_n - \mu)^2}{w_1 + w_2 + \dots + w_n}$$

GMM – Algoritmo EM



GMM vs K-means

Gaussian Mixture Models es una generalización de k-means.

- Lloyd (K-means) intenta minimizar $(x - \mu_k)^2$
- EM intenta minimizar $\frac{(x - \mu_k)^2}{\sigma^2}$
- El denominador agrega un modo de medir distancias ponderadas.
- Ambos susceptibles a óptimos locales.

GMM en Sklearn

```
from sklearn.mixture import GaussianMixture

model = GaussianMixture(n_components=K) # crea el modelo

model.fit(data)    # entrena el modelo con EM

scores= model.score_samples(data) # log(verosimilitud) de cada patrón

labels= model.predict(data) # etiqueta de cada dato (clusters)

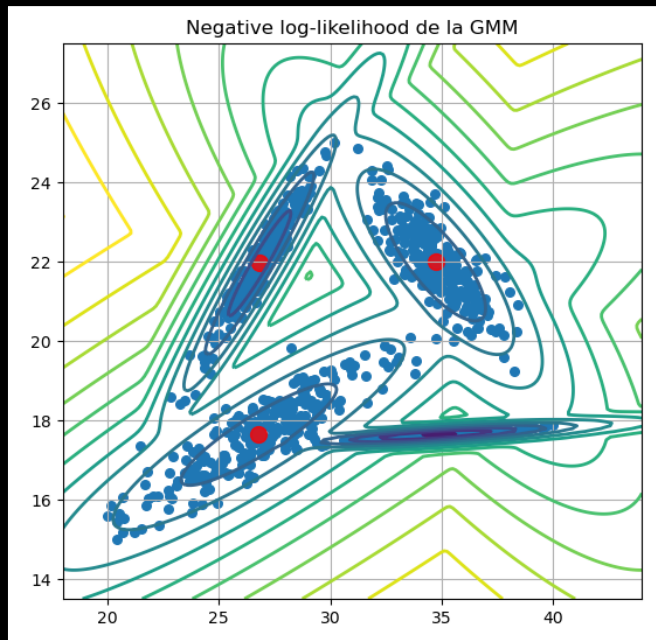
probs= model.predict_proba(data) # probabilidades para cada cluster -> sum(probs)==1
```


GMM – Definir número de componentes

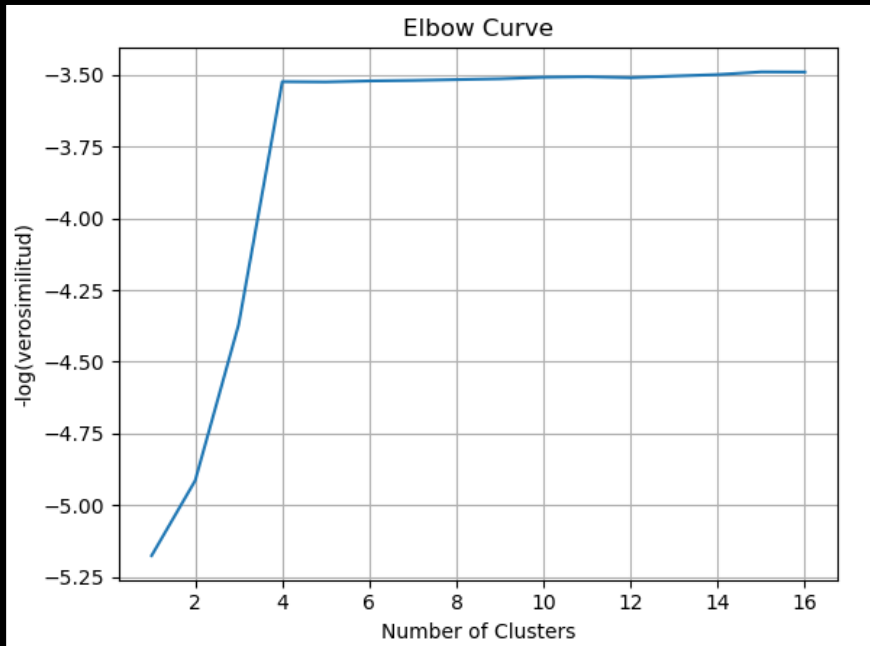
Mismo problema que en K-means.

No es trivial definir cantidad de componentes.

```
AA_utils.graficar_GMM(  
data, modelo, labels=False)
```



```
AA_utils.graficar_punto_elbow(data, 16, GMM=True)
```

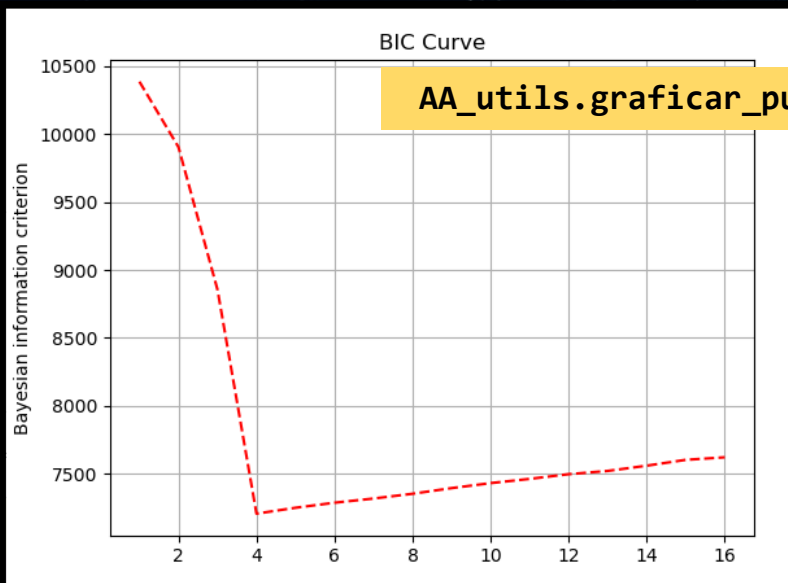


GMM – BIC

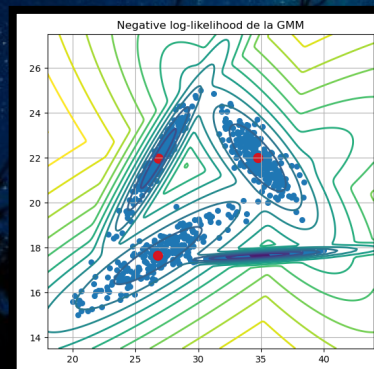
Bayesian information criterion (BIC) es un criterio para selección de modelos. Cuando más chico mejor.

$$BIC = k \log(n) - 2 \log(\mathcal{L})$$

\mathcal{L} = Verosimilitud, n = # datos, k = parámetros a estimar en el modelo

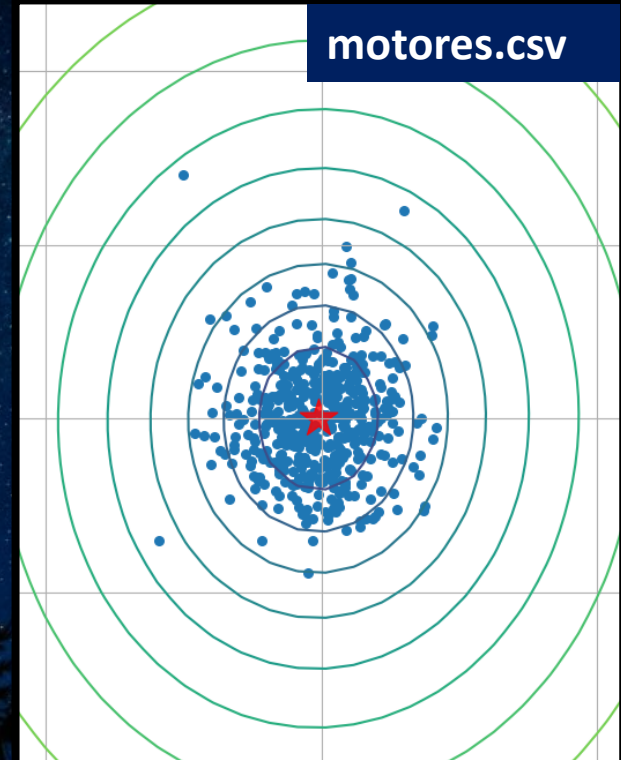


```
AA_utils.graficar_punto_elbow(data, 16, GMM=True)
```



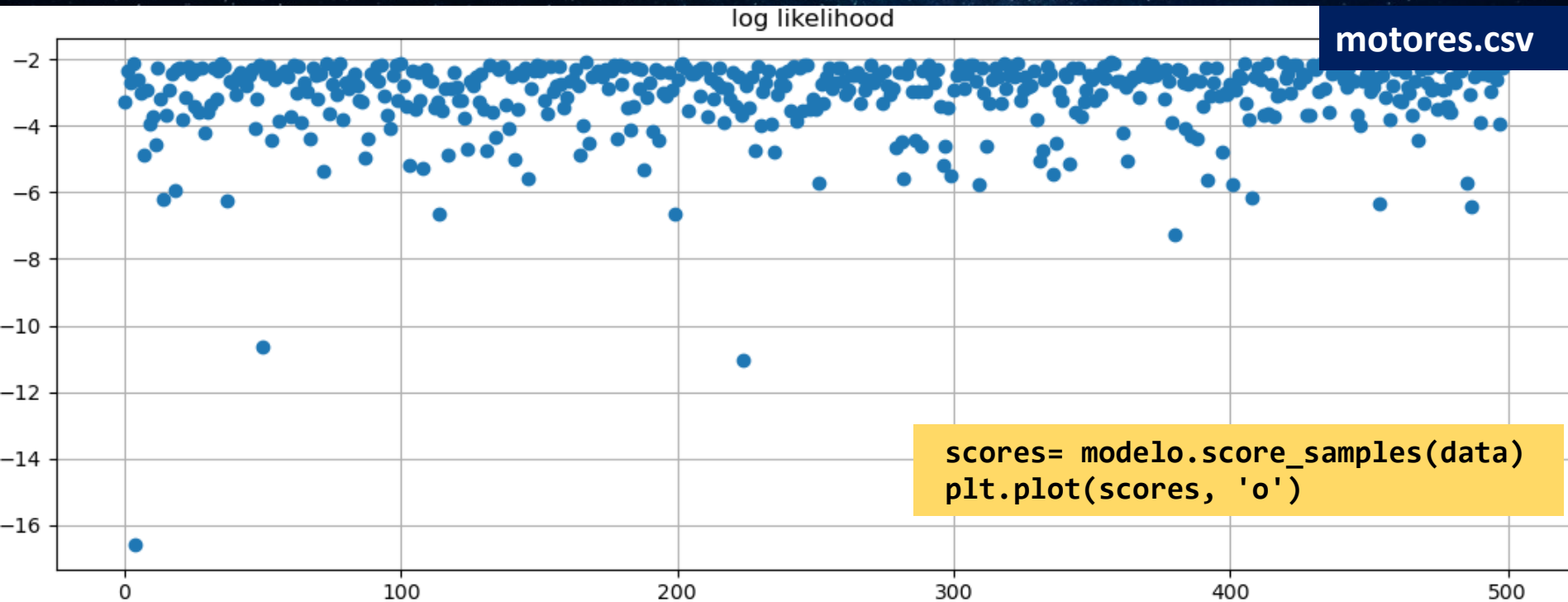
GMM – Detección de anomalías

- Supongamos que queremos detectar anomalías sin tenerlas etiquetadas.
- Utilizamos aprendizaje NO supervisado.
- Encontrar valores fuera de lo “normal”.



GMM – Detección de anomalías

- Log likelihood de cada elemento de mi dataset.



GMM – Detección de anomalías

- Establecemos una línea de corte (umbral) para decir que un dato es anómalo: Por ejemplo $\mu - 4\sigma$ (Es arbitrario)

