

Aprendizaje no supervisado

VC06: Agrupamiento basado en modelos probabilísticos -
Mixtura de Gaussianas y Algoritmo EM

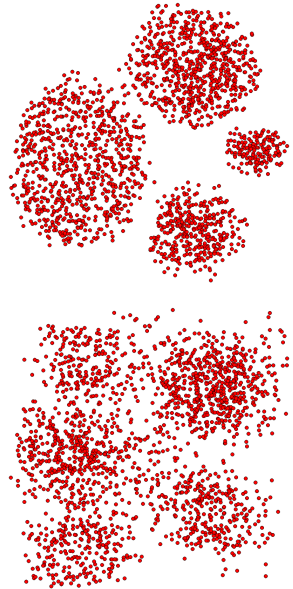
Félix José Fuentes Hurtado

felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia

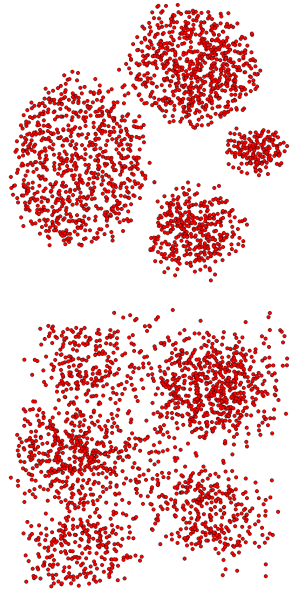
Tipos de algoritmos de agrupamiento

- ▶ Basados en particiones
- ▶ Jerárquicos
- ▶ Espectrales
- ▶ Basados en densidad
- ▶ Probabilísticos



Tipos de algoritmos de agrupamiento

- ▶ Basados en particiones
- ▶ Jerárquicos
- ▶ Espectrales
- ▶ Basados en densidad
- ▶ **Probabilísticos**



Agrupamiento probabilístico

- ▶ Se asume la existencia de un **modelo probabilístico** a partir del cual se han **generado** los datos observados.
- ▶ **Agrupamiento**: encontrar, a partir de los datos de entrenamiento, el mejor **ajuste del modelo generador** que se asume.
Procedimiento de estimación de los parámetros del modelo generador asumido
- ▶ Se asume la existencia de una variable oculta (no observada) que asigna los ejemplos a los clústeres
Sin esta información, es necesario acudir a técnicas de estimación en presencia de datos incompletos (algoritmo EM)

Agrupamiento probabilístico

Conceptos básicos

► Variable aleatoria, X

Función que asigna un valor al resultado de un experimento aleatorio

No se conoce el valor que tomará al ser medida, pero se sabe cuál es la distribución de probabilidad asociada al conjunto de valores posibles

Discretas vs. continuas

Ej., cara obtenida al lanzar un dado (todos los valores, misma probabilidad $1/6$)

Ej., peso de una persona tomada al azar en una población (escuela, pueblo, etc.)

Agrupamiento probabilístico

Conceptos básicos

► Variable aleatoria, X

Función que asigna un valor al resultado de un experimento aleatorio

No se conoce el valor que tomará al ser medida, pero se sabe cuál es la distribución de probabilidad asociada al conjunto de valores posibles

Discretas vs. continuas

Ej., cara obtenida al lanzar un dado (todos los valores, misma probabilidad $1/6$)

Ej., peso de una persona tomada al azar en una población (escuela, pueblo, etc.)

► Distribución de probabilidad, $p(X; \theta)$

Función que asigna, a cada valor posible x de la variable X , la probabilidad de que se obtenga dicho valor

Discretas vs. continuas

Ej., prob. de las caras de un dado (discreta, 6 valores)

Ej., prob. del peso las personas de una población (continua, infinitos valores)

Ej., prob. del número de quejas diarias de un negocio (discreta, infinitos valores)

Agrupamiento probabilístico

Conceptos básicos

► Variable aleatoria, X

Función que asigna un valor al resultado de un experimento aleatorio

No se conoce el valor que tomará al ser medida, pero se sabe cuál es la distribución de probabilidad asociada al conjunto de valores posibles

Discretas vs. continuas

Ej., cara obtenida al lanzar un dado (todos los valores, misma probabilidad $1/6$)

Ej., peso de una persona tomada al azar en una población (escuela, pueblo, etc.)

► Distribución de probabilidad, $p(X; \theta)$

Función que asigna, a cada valor posible x de la variable X , la probabilidad de que se obtenga dicho valor

Discretas vs. continuas



Ej., prob. de las caras de un dado (discreta, 6 valores)

Ej., prob. del peso las personas de una población (continua, infinitos valores)

Ej., prob. del número de quejas diarias de un negocio (discreta, infinitos valores)

Agrupamiento probabilístico

Conceptos básicos

► Variable aleatoria, X

Función que asigna un valor al resultado de un experimento aleatorio

No se conoce el valor que tomará al ser medida, pero se sabe cuál es la distribución de probabilidad asociada al conjunto de valores posibles

Discretas vs. continuas

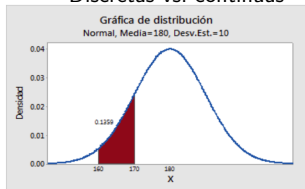
Ej., cara obtenida al lanzar un dado (todos los valores, misma probabilidad $1/6$)

Ej., peso de una persona tomada al azar en una población (escuela, pueblo, etc.)

► Distribución de probabilidad, $p(X; \theta)$

Función que asigna, a cada valor posible x de la variable X , la probabilidad de que se obtenga dicho valor

Discretas vs. continuas



Ej., prob. de las caras de un dado (discreta, 6 valores)

Ej., prob. del peso las personas de una población (continua, infinitos valores)

Ej., prob. del número de quejas diarias de un negocio (discreta, infinitos valores)

Agrupamiento probabilístico

Conceptos básicos

► Variable aleatoria, X

Función que asigna un valor al resultado de un experimento aleatorio

No se conoce el valor que tomará al ser medida, pero se sabe cuál es la distribución de probabilidad asociada al conjunto de valores posibles

Discretas vs. continuas

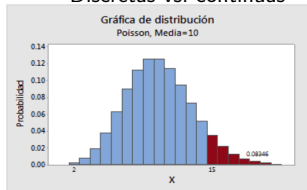
Ej., cara obtenida al lanzar un dado (todos los valores, misma probabilidad $1/6$)

Ej., peso de una persona tomada al azar en una población (escuela, pueblo, etc.)

► Distribución de probabilidad, $p(X; \theta)$

Función que asigna, a cada valor posible x de la variable X , la probabilidad de que se obtenga dicho valor

Discretas vs. continuas



Ej., prob. de las caras de un dado (discreta, 6 valores)

Ej., prob. del peso las personas de una población (continua, infinitos valores)

Ej., prob. del número de quejas diarias de un negocio (discreta, infinitos valores)

Agrupamiento probabilístico

Conceptos

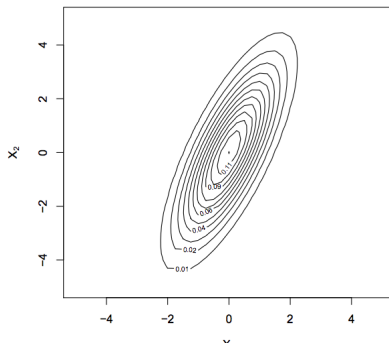
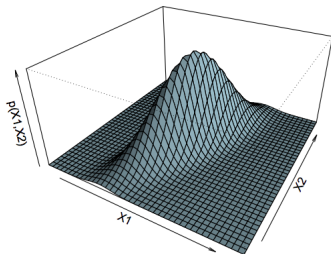
- Vector aleatorio (multi-variada), (X_1, \dots, X_v)

Función que asigna un vector de valores al resultado de un experimento aleatorio

La distribución de probabilidad asociada es multivariada, $p(X_1, \dots, X_v)$

Ej., coordenadas (x, y) en las que bota un balón al dejarlo caer desde un tercer piso

Ej., caras obtenidas al lanzar un dado y una moneda



Agrupamiento probabilístico

Conceptos

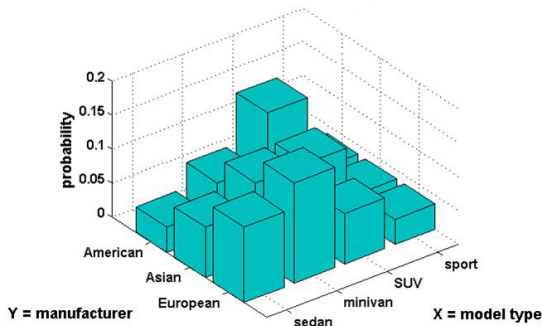
- Vector aleatorio (multi-variada), (X_1, \dots, X_v)

Función que asigna un vector de valores al resultado de un experimento aleatorio

La distribución de probabilidad asociada es multivariada, $p(X_1, \dots, X_v)$

Ej., coordenadas (x, y) en las que bota un balón al dejarlo caer desde un tercer piso

Ej., caras obtenidas al lanzar un dado y una moneda



Agrupamiento probabilístico

Conceptos

- ▶ Vector aleatorio (multi-variada), (X_1, \dots, X_v)

Función que asigna un vector de valores al resultado de un experimento aleatorio

La distribución de probabilidad asociada es multivariada, $p(X_1, \dots, X_v)$

Ej., coordenadas (x, y) en las que bota un balón al dejarlo caer desde un tercer piso

Ej., caras obtenidas al lanzar un dado y una moneda

- ▶ Modelo generativo

Es el modelo probabilístico que produjo un conjunto de datos dado

Agrupamiento probabilístico

Conceptos

- Vector aleatorio (multi-variada), (X_1, \dots, X_v)

Función que asigna un vector de valores al resultado de un experimento aleatorio

La distribución de probabilidad asociada es multivariada, $p(X_1, \dots, X_v)$

Ej., coordenadas (x, y) en las que bota un balón al dejarlo caer desde un tercer piso

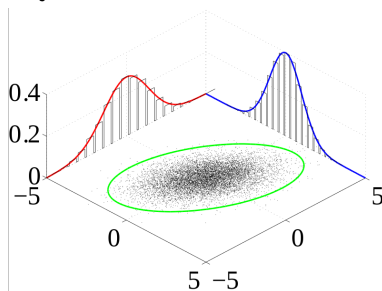
Ej., caras obtenidas al lanzar un dado y una moneda

- Modelo generativo

Es el modelo probabilístico que produjo un conjunto de datos dado

- Caso de una muestra

Es una instanciación del modelo generativo. Un valor (vector) producto de muestrear la distribución de prob. generadora.



Agrupamiento probabilístico

Conceptos

► Mezcla de distribuciones (Gaussianas)

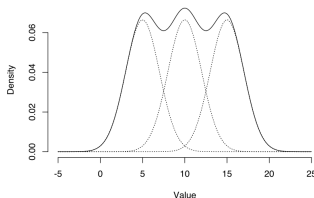
Función que asigna, a cada valor posible x de la variable X , la probabilidad de que se obtenga dicho valor

La probabilidad de cada valor se obtiene usando una serie de variables aleatorias (una de selección, y varias componentes):

- Una variable de selección que indica a qué componente pertenece (distribuida según $\{\pi_k\}_{k=1}^K$)
- La variable de la componente seleccionada (distribuida según $p_k(x; \theta_k)$)

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

Ej., edad de las personas de una escuela (alumnos/as, docentes)



Agrupamiento probabilístico

Conceptos

► Mixtura de distribuciones (Gaussianas)

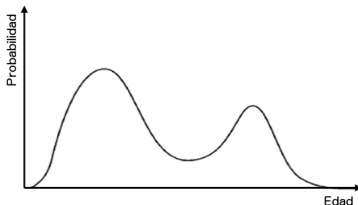
Función que asigna, a cada valor posible x de la variable X , la probabilidad de que se obtenga dicho valor

La probabilidad de cada valor se obtiene usando una serie de variables aleatorias (una de selección, y varias componentes):

- Una variable de selección que indica a qué componente pertenece (distribuida según $\{\pi_k\}_{k=1}^K$)
- La variable de la componente seleccionada (distribuida según $p_k(x; \theta_k)$)

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

Ej., edad de las personas de una escuela (alumnos/as, docentes)



Agrupamiento probabilístico

Mixtura de Gaussianas

Modelo generador de una Mixtura de Gaussianas sobre un vector aleatorio, (x_1, \dots, x_v) . La función de densidad de la mixtura es:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \cdot p_k(\mathbf{x}; \theta_k)$$

Coeficientes de mezcla, $\{\pi_1, \dots, \pi_K\}$: denotan la probabilidad de que un punto sea la instanciación de cada una de las componentes. Cumplen que:

$$0 \leq \pi_k \leq 1, \forall k \in \{1, \dots, K\} \wedge \sum_{k=1}^K \pi_k = 1$$

La distribución de probabilidad de la k -ésima componente:

$$p_k(\mathbf{x}; \theta_k) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^v |\boldsymbol{\Sigma}|}} e^{-(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2}$$

donde $\theta_k = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ son los parámetros de la componente.

Agrupamiento probabilístico

Mixtura de Gaussianas

Modelo generador de una Mixtura de Gaussianas sobre un vector aleatorio, (x_1, \dots, x_v) . La función de densidad de la mixtura es:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \cdot p_k(\mathbf{x}; \theta_k)$$

Coeficientes de mezcla, $\{\pi_1, \dots, \pi_K\}$: denotan la probabilidad de que un punto sea la instanciación de cada una de las componentes. Cumplen que:

$$0 \leq \pi_k \leq 1, \forall k \in \{1, \dots, K\} \wedge \sum_{k=1}^K \pi_k = 1$$

La distribución de probabilidad de la k -ésima componente:

$$p_k(\mathbf{x}; \theta_k) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^v |\Sigma|}} e^{-(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2}$$

donde $\theta_k = \{\boldsymbol{\mu}, \Sigma\}$ son los parámetros de la componente.

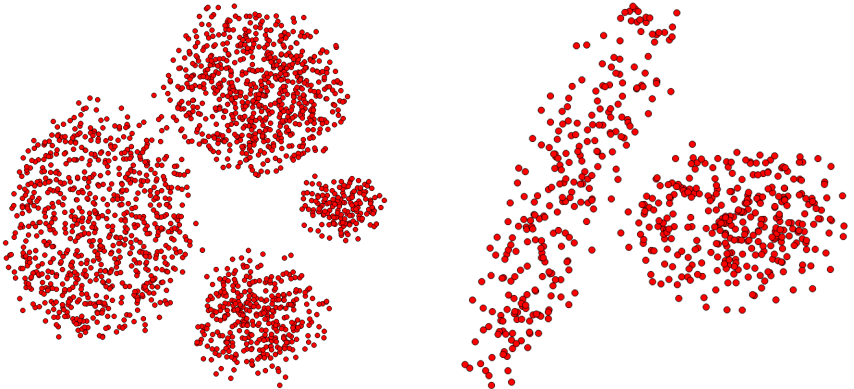
El proceso generador puede escalonarse en dos pasos:

- (i) seleccionar la componente k ; (ii) muestrear su distribución $p(\mathbf{x}; \theta_k)$

Agrupamiento probabilístico

Mixtura de Gaussianas

Relación con el análisis de agrupamiento:



Agrupamiento probabilístico

Mixtura de Gaussianas

¿Qué queremos hacer?

- ▶ Plantear un modelo generador y entenderlo
- ▶ Estimar los parámetros del modelo generador dados los datos observados

Agrupamiento probabilístico

Mixtura de Gaussianas

La aproximación más habitual:

Asumir como modelo probabilístico generador de los datos una mixtura de distribuciones de probabilidad

- ▶ Mixtura de Gaussianas
- ▶ Cada componente (distribución) de la mixtura, un clúster
- ▶ **Incertidumbre:** asignación de los ejemplos a una u otra componente
- ▶ Un ejemplo pertenece a todos los clústeres (con distinta probabilidad)
- ▶ Al final, cada ejemplo pertenece al clúster cuya componente le otorga mayor probabilidad

Agrupamiento probabilístico

Mixtura de Gaussianas

La aproximación más habitual:

Asumir como modelo probabilístico generador de los datos una mixtura de distribuciones de probabilidad

- ▶ Mixtura de Gaussianas
- ▶ Cada componente (distribución) de la mixtura, un clúster
- ▶ **Incertidumbre:** asignación de los ejemplos a una u otra componente
 - *** Algoritmo Esperanza-Maximización (EM) ***
- ▶ Un ejemplo pertenece a todos los clústeres (con distinta probabilidad)
- ▶ Al final, cada ejemplo pertenece al clúster cuya componente le otorga mayor probabilidad

Proceso generador escalonado:

1. Muestrear $p(z)$ para seleccionar la componente k desde la que se generará la muestra
2. Muestrear la distribución de probabilidad seleccionada

$$p_k(\mathbf{x}; \theta_k) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \Sigma_k) \text{ with } \theta_k = \{\boldsymbol{\mu}_k, \Sigma_k\}$$

Proceso generador escalonado:

1. Muestrear $p(z)$ para seleccionar la componente k desde la que se generará la muestra
*** Existencia de una variable aleatoria latente que determina la selección de la componente a muestrear***
2. Muestrear la distribución de probabilidad seleccionada

$$p_k(\mathbf{x}; \theta_k) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \Sigma_k) \quad \text{with} \quad \theta_k = \{\boldsymbol{\mu}_k, \Sigma_k\}$$

Asignación de casos a componentes

Variable aleatoria (multinomial con un único intento)

$$z \sim Mult(1, \{\pi_k\}_{k=1}^K)$$

Una muestra z es un vector K -dimensional con valor 1 para la componente seleccionada, y 0 para el resto

La distribución marginal de z es:

$$p(z) = \prod_{k=1}^K \pi_k^{z_k}$$

y la distribución condicional del caso \mathbf{x} dado z :

$$p(\mathbf{x}|z) = \prod_{k=1}^K p_k(\mathbf{x}; \theta_k)^{z_k}$$

Modelo completo

Distribución marginal de \mathbf{x} :

$$p(\mathbf{x}) = \sum_z p(\mathbf{x}, z) = \sum_z p(\mathbf{x}|z) \cdot p(z) = \sum_{k=1}^K \pi_k \cdot p(\mathbf{x}; \theta_k)$$

La probabilidad condicionada de z dado \mathbf{x} (usando la regla de Bayes):

$$\begin{aligned} p(z_k = 1|\mathbf{x}) &= \frac{p(\mathbf{x}|z_k = 1) \cdot p(z_k = 1)}{\sum_{k'=1}^K p(\mathbf{x}|z_{k'} = 1) \cdot p(z_{k'} = 1)} \\ &= \frac{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \cdot \pi_k}{\sum_{k'=1}^K \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'}) \cdot \pi_{k'}} \end{aligned}$$

Aprendizaje

Verosimilitud: Plausibilidad, dado un conjunto de datos, de un conjunto de parámetros Θ para un modelo.

Se calcula como la probabilidad asignada al conjunto de datos por el modelo parametrizado mediante Θ :

$$\mathcal{L}(\Theta|\{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = \prod_{i=1}^n p(\mathbf{x}_i; \Theta)$$

Parámetros del modelo: $\Theta = \{\pi_1, \dots, \pi_K, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K, \Sigma_1, \dots, \Sigma_K\}$

Aprendizaje: estimación máximo verosímil

$$\begin{aligned}\Theta_{ML} &= \arg \max_{\Theta} \log \mathcal{L}(\Theta | \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) \\ &= \arg \max_{\Theta} \sum_{i=1}^n \log p(\mathbf{x}_i; \Theta)\end{aligned}$$

Aprendizaje de parámetros máximo verosímiles: fórmula cerrada.

Aprendizaje: estimación máximo verosímil

$$\begin{aligned}\Theta_{ML} &= \arg \max_{\Theta} \log \mathcal{L}(\Theta | \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) \\ &= \arg \max_{\Theta} \sum_{i=1}^n \log p(\mathbf{x}_i; \Theta)\end{aligned}$$

Aprendizaje de parámetros máximo verosímiles: fórmula cerrada.

**** Probabilidad de cada componente:**

$$\hat{\pi}_k = \frac{\sum_{i=1}^n \hat{z}_{ik}}{n}$$

Aprendizaje: estimación máximo verosímil

$$\begin{aligned}\Theta_{ML} &= \arg \max_{\Theta} \log \mathcal{L}(\Theta | \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) \\ &= \arg \max_{\Theta} \sum_{i=1}^n \log p(\mathbf{x}_i; \Theta)\end{aligned}$$

Aprendizaje de parámetros máximo verosímiles: fórmula cerrada.

****** Centro de la normal de la k -ésima componente:

$$\hat{\mu}_k = \frac{\sum_{i=1}^n \hat{z}_{ik} \cdot \mathbf{x}_i}{\sum_{i=1}^n \hat{z}_{ik}}$$

Aprendizaje: estimación máximo verosímil

$$\begin{aligned}\Theta_{ML} &= \arg \max_{\Theta} \log \mathcal{L}(\Theta | \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) \\ &= \arg \max_{\Theta} \sum_{i=1}^n \log p(\mathbf{x}_i; \Theta)\end{aligned}$$

Aprendizaje de parámetros máximo verosímiles: fórmula cerrada.

****** Matriz de covarianza de la normal de la k -ésima componente:

$$\hat{\Sigma}_k = \frac{\sum_{i=1}^n \hat{z}_{ik} (\mathbf{x}_i - \hat{\mu}_k)(\mathbf{x}_i - \hat{\mu}_k)^T}{\sum_{i=1}^n \hat{z}_{ik}}$$

Aprendizaje: estimación máximo verosímil

$$\begin{aligned}\Theta_{ML} &= \arg \max_{\Theta} \log \mathcal{L}(\Theta | \{\mathbf{x}_1, \dots, \mathbf{x}_n\}) \\ &= \arg \max_{\Theta} \sum_{i=1}^n \log p(\mathbf{x}_i; \Theta)\end{aligned}$$

Aprendizaje de parámetros máximo verosímiles: fórmula cerrada.

Calcular la asignación prob. de cada ejemplo a componentes:

$$\hat{z}_{ik} = p(z_{ik} = 1 | \mathbf{x}_i) = \frac{\hat{\pi}_k \cdot \mathcal{N}(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k)}{\sum_{k'=1}^K \hat{\pi}_{k'} \cdot \mathcal{N}(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_{k'}, \hat{\boldsymbol{\Sigma}}_{k'})}$$

Agrupamiento probabilístico

Mixtura de Gaussianas

Aprendizaje: inter-dependencia

Calcular asignación prob. de cada ejemplo a componentes:

$$\hat{z}_{ik} = \hat{\pi}_k \cdot \mathcal{N}(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k) / \left(\sum_{k'=1}^K \hat{\pi}_{k'} \cdot \mathcal{N}(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_{k'}, \hat{\boldsymbol{\Sigma}}_{k'}) \right)$$

Aprender los parámetros:

- Probabilidad de cada componente:

$$\hat{\pi}_k = \sum_{i=1}^n \hat{z}_{ik} / n$$

- Centro de la normal de la k -ésima componente:

$$\hat{\boldsymbol{\mu}}_k = \left(\sum_{i=1}^n \hat{z}_{ik} \cdot \mathbf{x}_i \right) / \left(\sum_{i=1}^n \hat{z}_{ik} \right)$$

- Matriz de covarianza de la normal de la k -ésima componente:

$$\hat{\boldsymbol{\Sigma}}_k = \left(\sum_{i=1}^n \hat{z}_{ik} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)^T \right) / \left(\sum_{i=1}^n \hat{z}_{ik} \right)$$

Agrupamiento probabilístico

Mixtura de Gaussianas

Aprendizaje: algoritmo Esperanza-Maximización

Paso E: Calcular asignación prob. de cada ejemplo a componentes:

$$\hat{z}_{ik} = \hat{\pi}_k \cdot \mathcal{N}(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k) / \left(\sum_{k'=1}^K \hat{\pi}_{k'} \cdot \mathcal{N}(\mathbf{x}_i | \hat{\boldsymbol{\mu}}_{k'}, \hat{\boldsymbol{\Sigma}}_{k'}) \right)$$

Paso M: Aprender los parámetros

- Probabilidad de cada componente:

$$\hat{\pi}_k = \sum_{i=1}^n \hat{z}_{ik} / n$$

- Centro de la normal de la k -ésima componente:

$$\hat{\boldsymbol{\mu}}_k = \left(\sum_{i=1}^n \hat{z}_{ik} \cdot \mathbf{x}_i \right) / \left(\sum_{i=1}^n \hat{z}_{ik} \right)$$

- Matriz de covarianza de la normal de la k -ésima componente:

$$\hat{\boldsymbol{\Sigma}}_k = \left(\sum_{i=1}^n \hat{z}_{ik} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)^T \right) / \left(\sum_{i=1}^n \hat{z}_{ik} \right)$$

Agrupamiento probabilístico

Mixtura de Gaussianas

Aprendizaje: algoritmo Esperanza-Maximización

Inicialización aleatoria

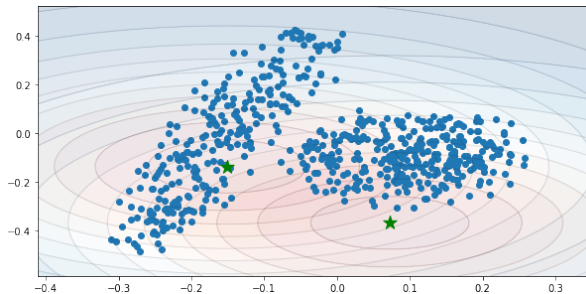
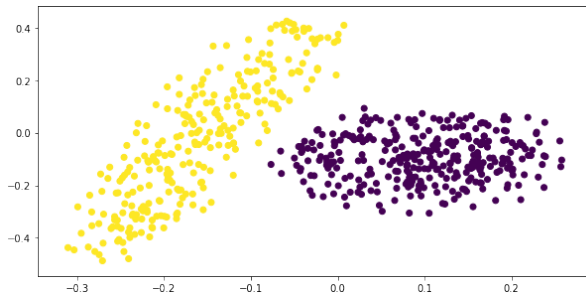
Paso E: Calcular asignación prob. de cada ejemplo a componentes

Paso M: Aprender los parámetros

Convergencia a un óptimo local de la función de verosimilitud

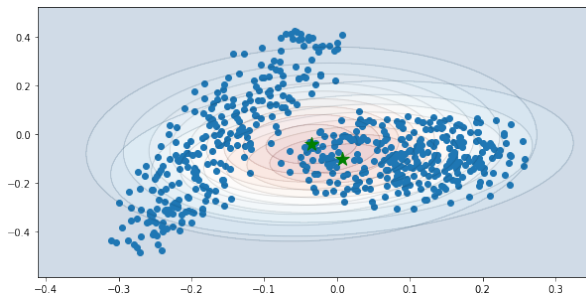
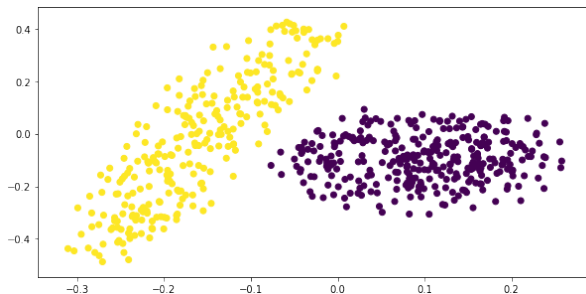
Agrupamiento probabilístico

Mixtura de Gaussianas



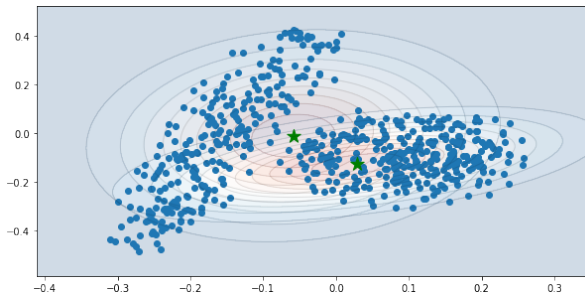
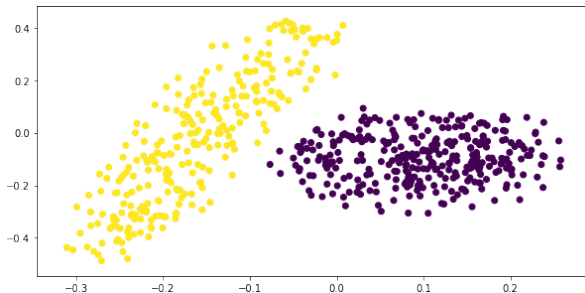
Agrupamiento probabilístico

Mixtura de Gaussianas



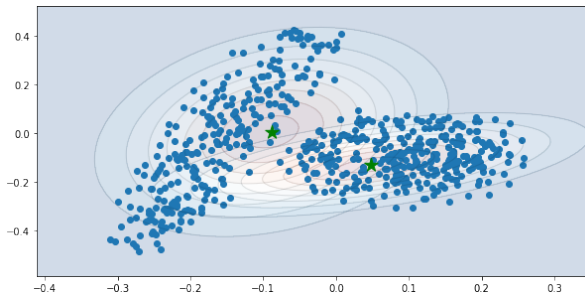
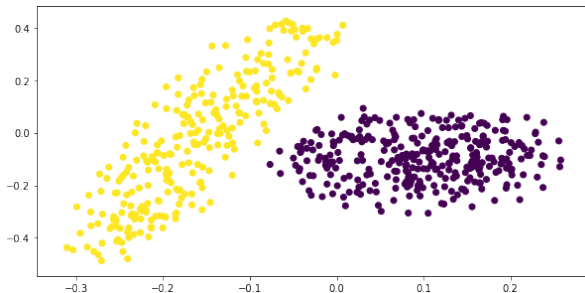
Agrupamiento probabilístico

Mixtura de Gaussianas



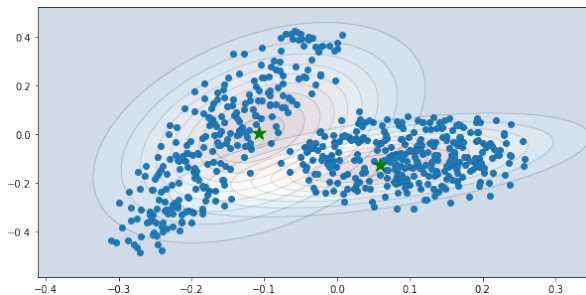
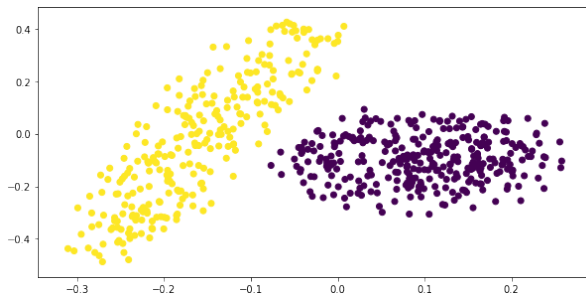
Agrupamiento probabilístico

Mixtura de Gaussianas



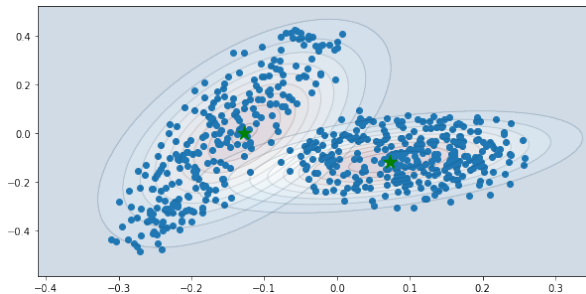
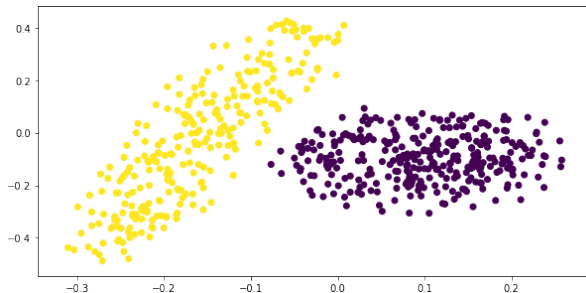
Agrupamiento probabilístico

Mixtura de Gaussianas



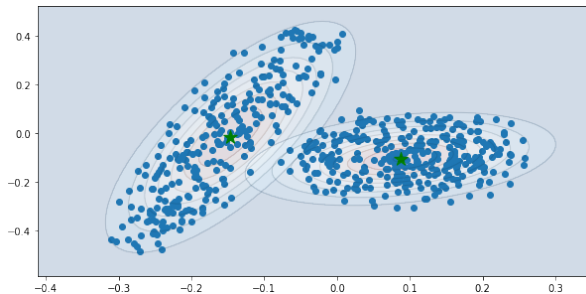
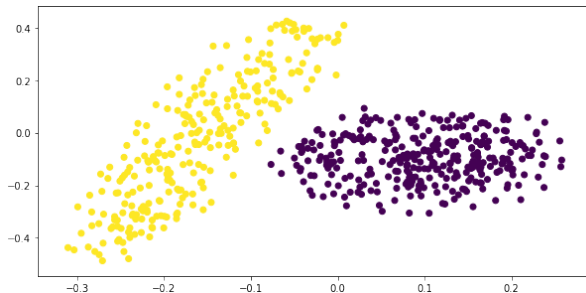
Agrupamiento probabilístico

Mixtura de Gaussianas



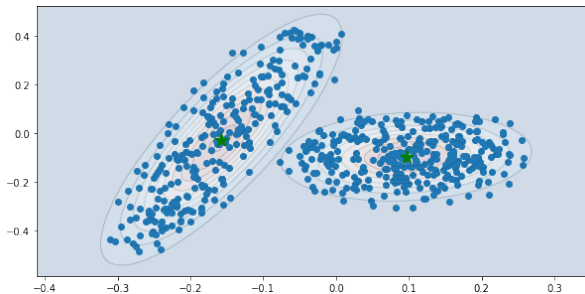
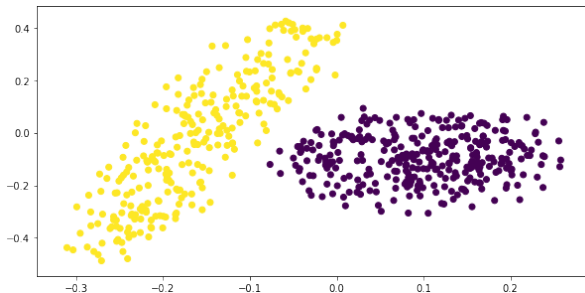
Agrupamiento probabilístico

Mixtura de Gaussianas



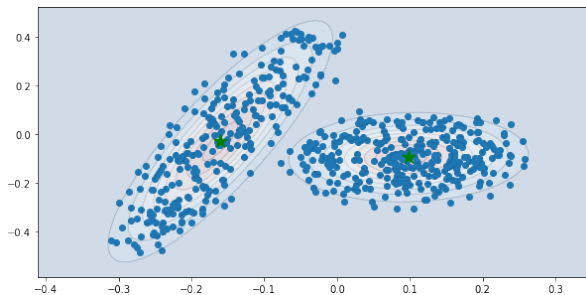
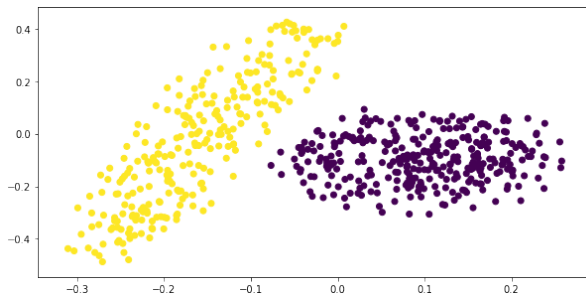
Agrupamiento probabilístico

Mixtura de Gaussianas



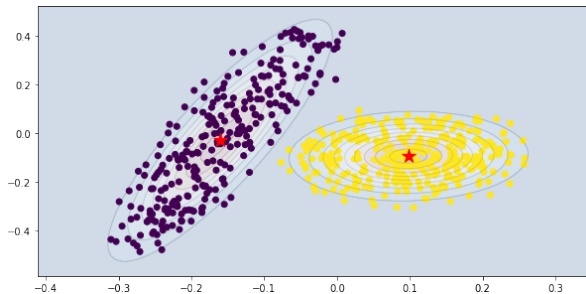
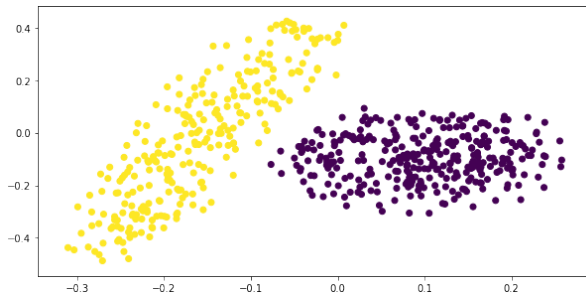
Agrupamiento probabilístico

Mixtura de Gaussianas



Agrupamiento probabilístico

Mixtura de Gaussianas



Agrupamiento probabilístico

Mixtura de Gaussianas

Ventajas

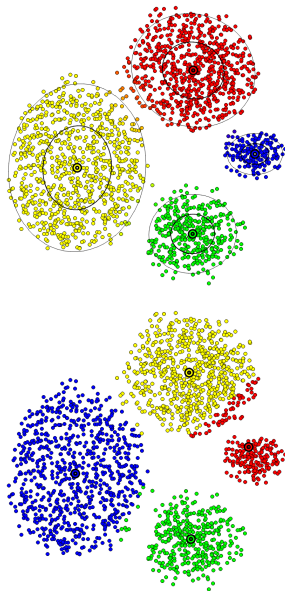
- ▶ Definición basada en probabilidad
- ▶ Definición sencilla
- ▶ Podría funcionar con diferentes distribuciones
- ▶ Funciona con clústeres cóncavos elípticos (vs. K -means)

Agrupamiento probabilístico

Mixtura de Gaussianas

Relación con K -means

- ▶ Asignación probabilística vs. determinista
- ▶ Clústeres elípticos vs. circulares

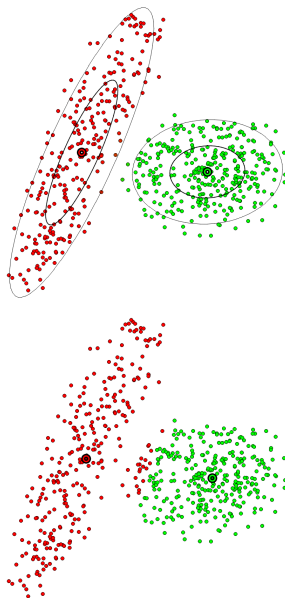


Agrupamiento probabilístico

Mixtura de Gaussianas

Relación con K -means

- Asignación probabilística vs. determinista
- Clústeres elípticos vs. circulares



Agrupamiento probabilístico

Mixtura de Gaussianas

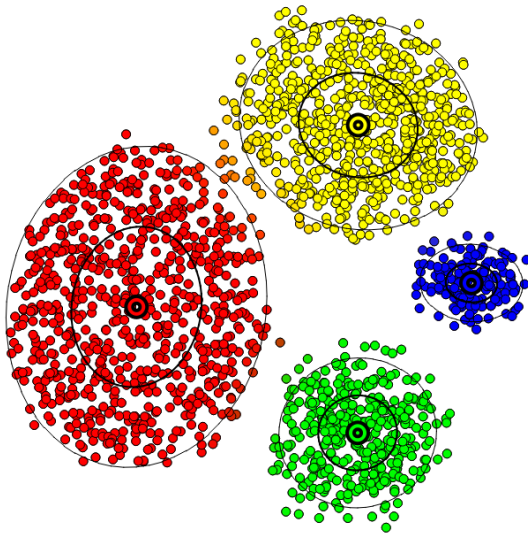
Desventajas

- ▶ Depende de K (parámetro)
- ▶ Asume clústeres cóncavos
- ▶ Problemas al lidiar con clústeres de diferente forma
- ▶ Depende de la inicialización

Agrupamiento probabilístico

Mixtura de Gaussianas

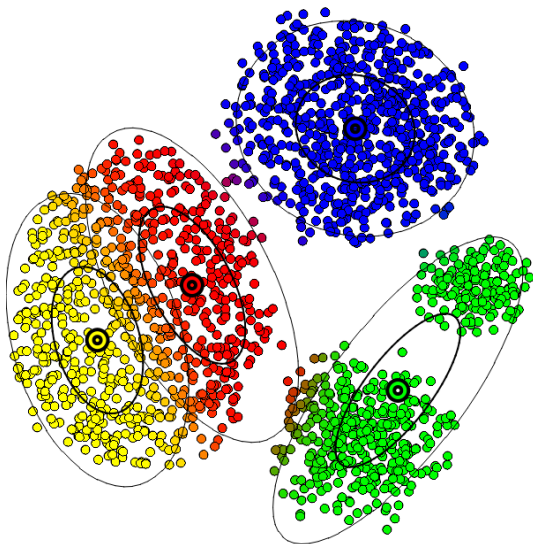
Dependencia de la inicialización



Agrupamiento probabilístico

Mixtura de Gaussianas

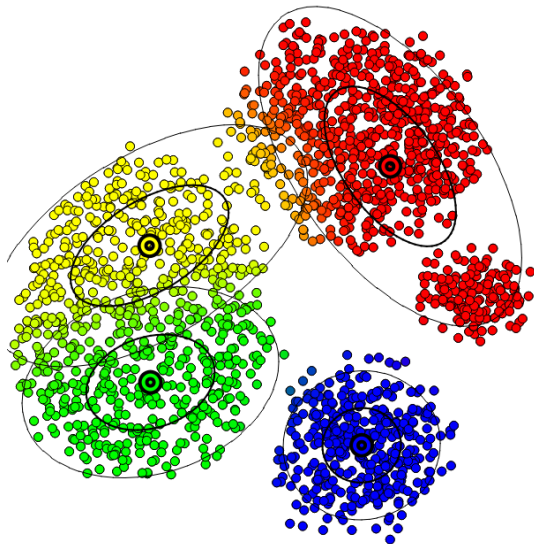
Dependencia de la inicialización



Agrupamiento probabilístico

Mixtura de Gaussianas

Dependencia de la inicialización



Aprendizaje no supervisado

VC06: Agrupamiento basado en modelos probabilísticos -
Mixtura de Gaussianas y Algoritmo EM

Félix José Fuentes Hurtado

felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia