

Modelos de mezclas Gaussianas

Héctor Selley

Universidad Anáhuac México

8 de septiembre de 2022

- 1 Modelos de Mezclas Gaussianas
- 2 ¿Qué hace el algoritmo EM en el GMM?
- 3 Algoritmo EM
- 4 Definición del GMM

Introducción

- El modelo de mezclas gaussianas (GMM) es conocido como un algoritmo de aprendizaje no supervisado para clustering.

Introducción

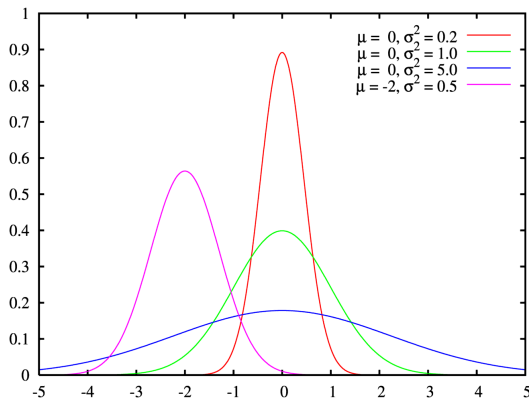
- El modelo de mezclas gaussianas (GMM) es conocido como un algoritmo de aprendizaje no supervisado para clustering.
- Utiliza la función de distribución Gaussiana descrita por su media y varianza.

Introducción

- El modelo de mezclas gaussianas (GMM) es conocido como un algoritmo de aprendizaje no supervisado para clustering.
- Utiliza la función de distribución Gaussiana descrita por su media y varianza.
- El término “mezclas” implica que se utilizará más de una distribución Gaussiana.

Introducción

- El modelo de mezclas gaussianas (GMM) es conocido como un algoritmo de aprendizaje no supervisado para clustering.
- Utiliza la función de distribución Gaussiana descrita por su media y varianza.
- El término “mezclas” implica que se utilizará más de una distribución Gaussiana.



¿Qué es?

- Suponga que se tiene un conjunto de datos y se sabe que pertenecen a un número de modelos Gaussianos.

¿Qué es?

- Suponga que se tiene un conjunto de datos y se sabe que pertenecen a un número de modelos Gaussianos.
- El modelo se describe por la media y varianza para datos de 1 característica o por el vector media y el vector varianza para datos de N características.

¿Qué es?

- Suponga que se tiene un conjunto de datos y se sabe que pertenecen a un número de modelos Gaussianos.
- El modelo se describe por la media y varianza para datos de 1 característica o por el vector media y el vector varianza para datos de N características.
- Podríamos saber la probabilidad de que cada dato pertenezca a uno de los 2 modelos gaussianos si conocemos sus funciones de densidad.

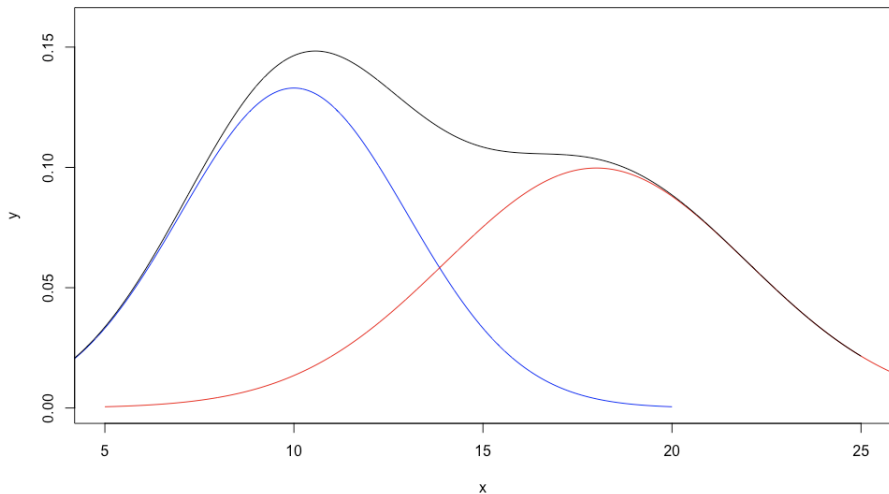
¿Qué es?

- Suponga que se tiene un conjunto de datos y se sabe que pertenecen a un número de modelos Gaussianos.
- El modelo se describe por la media y varianza para datos de 1 característica o por el vector media y el vector varianza para datos de N características.
- Podríamos saber la probabilidad de que cada dato pertenezca a uno de los 2 modelos gaussianos si conocemos sus funciones de densidad.
- Luego, podemos asignar el dato al modelo específico con la mayor probabilidad entre la mezcla Gaussiana.

¿Qué es?

- Suponga que se tiene un conjunto de datos y se sabe que pertenecen a un número de modelos Gaussianos.
- El modelo se describe por la media y varianza para datos de 1 característica o por el vector media y el vector varianza para datos de N características.
- Podríamos saber la probabilidad de que cada dato pertenezca a uno de los 2 modelos gaussianos si conocemos sus funciones de densidad.
- Luego, podemos asignar el dato al modelo específico con la mayor probabilidad entre la mezcla Gaussiana.

¿Qué es?



¿Qué es?

- Como se puede notar, hay dos cosas más importantes en el modelo de mezclas Gaussianas:

¿Qué es?

- Como se puede notar, hay dos cosas más importantes en el modelo de mezclas Gaussianas:
 - 1 Estimar los parámetros para cada componente Gaussiano dentro de la mezcla Gaussiana.

¿Qué es?

- Como se puede notar, hay dos cosas más importantes en el modelo de mezclas Gaussianas:
 - 1 Estimar los parámetros para cada componente Gaussiano dentro de la mezcla Gaussiana.
 - 2 Determinar a cual componente Gaussiano le corresponde algún dato.

¿Qué es?

- Como se puede notar, hay dos cosas más importantes en el modelo de mezclas Gaussianas:
 - 1 Estimar los parámetros para cada componente Gaussiano dentro de la mezcla Gaussiana.
 - 2 Determinar a cual componente Gaussiano le corresponde algún dato.
- Esta es la razón por la cual el clustering es una de las aplicaciones más importantes del modelo de mezclas Gaussianas pero la esencia del modelo de mezclas Gaussianas es la estimación de densidad.

¿Qué es?

- Para estimar los parámetros que describen cada componente en el modelo de mezclas Gaussianas se utiliza un método llamado algoritmo Expectation-Maximization (EM).

¿Qué es?

- Para estimar los parámetros que describen cada componente en el modelo de mezclas Gaussianas se utiliza un método llamado algoritmo Expectation-Maximization (EM).
- El algoritmo EM es ampliamente utilizado para la estimación de parámetros cuando un modelo depende de algunas variables latentes no observadas.

¿Qué es?

- Para estimar los parámetros que describen cada componente en el modelo de mezclas Gaussianas se utiliza un método llamado algoritmo Expectation-Maximization (EM).
- El algoritmo EM es ampliamente utilizado para la estimación de parámetros cuando un modelo depende de algunas variables latentes no observadas.
- Las variables latentes en el GMM son aquellas que describen a cual componente Gaussiano pertenece cada dato.

¿Qué es?

- Para estimar los parámetros que describen cada componente en el modelo de mezclas Gaussianas se utiliza un método llamado algoritmo Expectation-Maximization (EM).
- El algoritmo EM es ampliamente utilizado para la estimación de parámetros cuando un modelo depende de algunas variables latentes no observadas.
- Las variables latentes en el GMM son aquellas que describen a cual componente Gaussiano pertenece cada dato.
- Dado que sólo observamos los datos, esta es una variable latente no observada.

Contenido

- 1 Modelos de Mezclas Gaussianas
- 2 ¿Qué hace el algoritmo EM en el GMM?
- 3 Algoritmo EM
- 4 Definición del GMM

¿Qué hace el algoritmo EM en el GMM?

- El algoritmo EM se utiliza para estimar los parámetros Gaussianos.

¿Qué hace el algoritmo EM en el GMM?

- El algoritmo EM se utiliza para estimar los parámetros Gaussianos.
- Vayamos al inicio:

¿Qué hace el algoritmo EM en el GMM?

- El algoritmo EM se utiliza para estimar los parámetros Gaussianos.
- Vayamos al inicio:
- Se debe entender una distribución Gaussiana antes de entender una mezcla de ellas.

¿Qué hace el algoritmo EM en el GMM?

- El algoritmo EM se utiliza para estimar los parámetros Gaussianos.
- Vayamos al inicio:
- Se debe entender una distribución Gaussiana antes de entender una mezcla de ellas.
- Supóngase que tiene una secuencia de datos y cada uno tiene una sola característica.

¿Qué hace el algoritmo EM en el GMM?

- El algoritmo EM se utiliza para estimar los parámetros Gaussianos.
- Vayamos al inicio:
- Se debe entender una distribución Gaussiana antes de entender una mezcla de ellas.
- Supóngase que tiene una secuencia de datos y cada uno tiene una sola característica.
- Se puede graficar la densidad de los datos en un eje que represente esa característica.

Ejemplo

Ejemplo

Expliquemos este concepto a través de un ejemplo sencillo:

Ejemplo

Expliquemos este concepto a través de un ejemplo sencillo:

- Se tiene una cubeta con manzanas Gala.

Ejemplo

Expliquemos este concepto a través de un ejemplo sencillo:

- Se tiene una cubeta con manzanas Gala.
- Se desea describir las manzanas Gala mediante su diámetro.

Ejemplo

Expliquemos este concepto a través de un ejemplo sencillo:

- Se tiene una cubeta con manzanas Gala.
- Se desea describir las manzanas Gala mediante su diámetro.
- Podría determinarse el promedio de los diámetros de las manzanas para describirlas.

Ejemplo

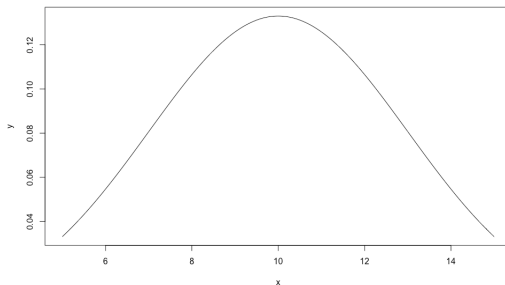
Expliquemos este concepto a través de un ejemplo sencillo:

- Se tiene una cubeta con manzanas Gala.
- Se desea describir las manzanas Gala mediante su diámetro.
- Podría determinarse el promedio de los diámetros de las manzanas para describirlas.
- Pero en lugar de eso se desea la distribución completa de cada una de las manzanas.

Ejemplo

Expliquemos este concepto a través de un ejemplo sencillo:

- Se tiene una cubeta con manzanas Gala.
- Se desea describir las manzanas Gala mediante su diámetro.
- Podría determinarse el promedio de los diámetros de las manzanas para describirlas.
- Pero en lugar de eso se desea la distribución completa de cada una de las manzanas.



Distribución Gaussiana

La gráfica de densidad puede ser descrita por la ecuación 1

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

donde μ es la media y σ es la desviación estándar, los cuales a su vez describen la distribución Gaussiana.

Ejemplo

- Suponga que se tienen dos cubetas con manzanas, una cubeta con manzanas Gala y otra cubeta con manzanas Fuji.

Ejemplo

- Suponga que se tienen dos cubetas con manzanas, una cubeta con manzanas Gala y otra cubeta con manzanas Fuji.
- Accidentalmente se han mezclado las manzanas en las cubetas y no nos es posible distinguirlas fácilmente.

Ejemplo

- Suponga que se tienen dos cubetas con manzanas, una cubeta con manzanas Gala y otra cubeta con manzanas Fuji.
- Accidentalmente se han mezclado las manzanas en las cubetas y no nos es posible distinguirlas fácilmente.
- Lo que si es posible hacer es medir el diámetro de las manzanas.

Ejemplo

- Suponga que se tienen dos cubetas con manzanas, una cubeta con manzanas Gala y otra cubeta con manzanas Fuji.
- Accidentalmente se han mezclado las manzanas en las cubetas y no nos es posible distinguirlas fácilmente.
- Lo que si es posible hacer es medir el diámetro de las manzanas.
- Este criterio sería útil siempre y cuando las manzanas Fuji y Gala tengan una diferencia sustancial en su tamaño.

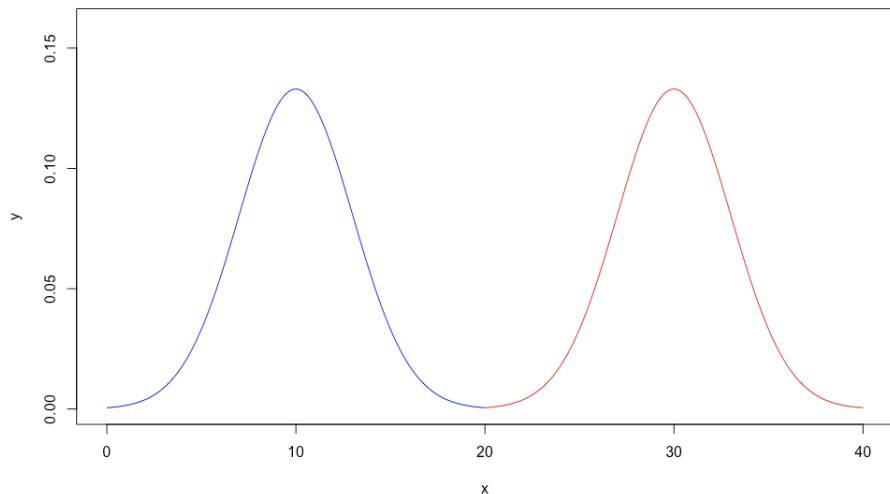
Ejemplo

- Suponga que se tienen dos cubetas con manzanas, una cubeta con manzanas Gala y otra cubeta con manzanas Fuji.
- Accidentalmente se han mezclado las manzanas en las cubetas y no nos es posible distinguirlas fácilmente.
- Lo que si es posible hacer es medir el diámetro de las manzanas.
- Este criterio sería útil siempre y cuando las manzanas Fuji y Gala tengan una diferencia sustancial en su tamaño.
- El diámetro entonces sería una *característica*.

Ejemplo

- Suponga que se tienen dos cubetas con manzanas, una cubeta con manzanas Gala y otra cubeta con manzanas Fuji.
- Accidentalmente se han mezclado las manzanas en las cubetas y no nos es posible distinguirlas fácilmente.
- Lo que si es posible hacer es medir el diámetro de las manzanas.
- Este criterio sería útil siempre y cuando las manzanas Fuji y Gala tengan una diferencia sustancial en su tamaño.
- El diámetro entonces sería una *característica*.
- Si los diámetros de las manzanas son lo suficientemente diferentes de acuerdo a su tipo, seguirían entonces dos distribuciones Gaussianas diferentes.

Dos distribuciones Gaussianas



Ejemplo

En este caso, se separarían las manzanas de acuerdo a su diámetro.

Ejemplo

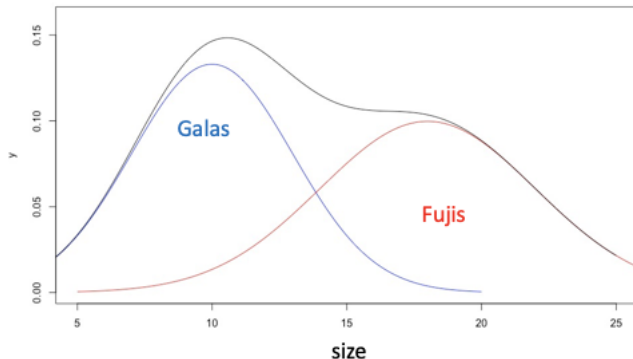
En este caso, se separarían las manzanas de acuerdo a su diámetro.

- 1 Fuji: $\text{diametro} > 2$ pulgadas
- 2 Gala: $\text{diametro} < 2$ pulgadas

Ejemplo

En este caso, se separarían las manzanas de acuerdo a su diámetro.

- 1 Fuji: $\text{diametro} > 2$ pulgadas
- 2 Gala: $\text{diametro} < 2$ pulgadas



Ejemplo

- En la gráfica anterior, la línea negra representa la densidad observada de manzanas.

Ejemplo

- En la gráfica anterior, la línea negra representa la densidad observada de manzanas.
- Nótese que la densidad sugiere la ubicación de las Gaussianas.

Ejemplo

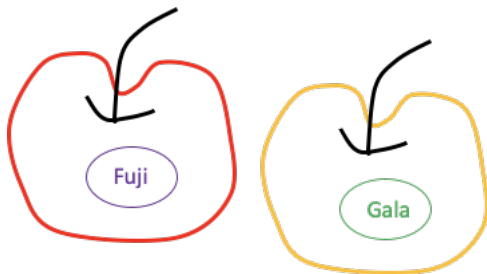
- En la gráfica anterior, la línea negra representa la densidad observada de manzanas.
- Nótese que la densidad sugiere la ubicación de las Gaussianas.
- Si supiéramos la media y varianza para cada función de densidad, se podría obtener la probabilidad de que una manzana pertenezca a algún tipo.

Ejemplo

- En la gráfica anterior, la línea negra representa la densidad observada de manzanas.
- Nótese que la densidad sugiere la ubicación de las Gaussianas.
- Si supiéramos la media y varianza para cada función de densidad, se podría obtener la probabilidad de que una manzana pertenezca a algún tipo.
- Si algunas de las manzanas tuvieran una etiqueta indicando su tipo, se podría estimar la media y varianza de acuerdo a ello.

Ejemplo

- En la gráfica anterior, la línea negra representa la densidad observada de manzanas.
- Nótese que la densidad sugiere la ubicación de las Gaussianas.
- Si supiéramos la media y varianza para cada función de densidad, se podría obtener la probabilidad de que una manzana pertenezca a algún tipo.
- Si algunas de las manzanas tuvieran una etiqueta indicando su tipo, se podría estimar la media y varianza de acuerdo a ello.



Contenido

- 1 Modelos de Mezclas Gaussianas
- 2 ¿Qué hace el algoritmo EM en el GMM?
- 3 Algoritmo EM**
- 4 Definición del GMM

La idea del algoritmo EM

El algoritmo EM (Expectation-Maximization) consiste de dos procesos:

La idea del algoritmo EM

El algoritmo EM (Expectation-Maximization) consiste de dos procesos:

- ① **Paso E:** Estima la variable latente, aquella variable clave que determina la clase a la que pertenece el dato. Esta variable latente influye en los datos pero no es observable.

La idea del algoritmo EM

El algoritmo EM (Expectation-Maximization) consiste de dos procesos:

- 1 **Paso E:** Estima la variable latente, aquella variable clave que determina la clase a la que pertenece el dato. Esta variable latente influye en los datos pero no es observable.
- 2 **Paso M:** Estima los parámetros de las distribuciones maximizando la probabilidad de los datos. La probabilidad describe cuanto coincide un conjunto de parámetros con los datos para lo cuál se busca obtener el máximo (mayor coincidencia con los datos, también llamado estimación de máxima verosimilitud).

La idea del algoritmo EM

El algoritmo EM (Expectation-Maximization) consiste de dos procesos:

- 1 **Paso E:** Estima la variable latente, aquella variable clave que determina la clase a la que pertenece el dato. Esta variable latente influye en los datos pero no es observable.
- 2 **Paso M:** Estima los parámetros de las distribuciones maximizando la probabilidad de los datos. La probabilidad describe cuanto coincide un conjunto de parámetros con los datos para lo cuál se busca obtener el máximo (mayor coincidencia con los datos, también llamado estimación de máxima verosimilitud).

Este algoritmo debe repetirse hasta conseguir una conjunto de parámetros adecuados para una separación también adecuada.

De vuelta al ejemplo

- Considere que a las manzanas que no tiene una etiqueta se les asigna una de algún tipo de manera aleatoria.

De vuelta al ejemplo

- Considere que a las manzanas que no tiene una etiqueta se les asigna una de algún tipo de manera aleatoria.
- La idea es que dado que no es posible separar los datos, se hace una suposición del tipo para la inicialización.

De vuelta al ejemplo

- Considere que a las manzanas que no tiene una etiqueta se les asigna una de algún tipo de manera aleatoria.
- La idea es que dado que no es posible separar los datos, se hace una suposición del tipo para la inicialización.
- Este proceso se llama: **inicialización del algoritmo EM**.

De vuelta al ejemplo

- Considere que a las manzanas que no tiene una etiqueta se les asigna una de algún tipo de manera aleatoria.
- La idea es que dado que no es posible separar los datos, se hace una suposición del tipo para la inicialización.
- Este proceso se llama: **inicialización del algoritmo EM**.
- Ahora se lleva a cabo el paso E del algoritmo EM.

De vuelta al ejemplo

- Considere que a las manzanas que no tiene una etiqueta se les asigna una de algún tipo de manera aleatoria.
- La idea es que dado que no es posible separar los datos, se hace una suposición del tipo para la inicialización.
- Este proceso se llama: **inicialización del algoritmo EM**.
- Ahora se lleva a cabo el paso E del algoritmo EM.
- A las manzanas con una etiqueta correcta Fuji se les asigna una probabilidad 1 de ser Fuji y probabilidad 0 de ser Gala.

De vuelta al ejemplo

- Considere que a las manzanas que no tiene una etiqueta se les asigna una de algún tipo de manera aleatoria.
- La idea es que dado que no es posible separar los datos, se hace una suposición del tipo para la inicialización.
- Este proceso se llama: **inicialización del algoritmo EM**.
- Ahora se lleva a cabo el paso E del algoritmo EM.
- A las manzanas con una etiqueta correcta Fuji se les asigna una probabilidad 1 de ser Fuji y probabilidad 0 de ser Gala.
- Viceversa para las manzanas Gala.

De vuelta al ejemplo

- Si algún **paso M** hubiese ocurrido antes, habría que recalcular las probabilidades para las manzanas.

De vuelta al ejemplo

- Si algún **paso M** hubiese ocurrido antes, habría que recalcular las probabilidades para las manzanas.
- Si alguna manzana Fuji tuviera una probabilidad mayor de ser Gala, se cambia la etiqueta.

De vuelta al ejemplo

- Si algún **paso M** hubiese ocurrido antes, habría que recalcular las probabilidades para las manzanas.
- Si alguna manzana Fuji tuviera una probabilidad mayor de ser Gala, se cambia la etiqueta.
- Viceversa para las manzanas Gala.

De vuelta al ejemplo

- Si algún **paso M** hubiese ocurrido antes, habría que recalcular las probabilidades para las manzanas.
- Si alguna manzana Fuji tuviera una probabilidad mayor de ser Gala, se cambia la etiqueta.
- Viceversa para las manzanas Gala.



De vuelta al ejemplo

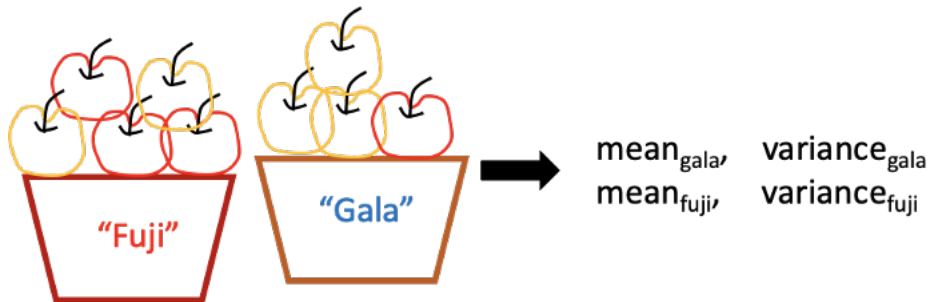
- Al realizar el **paso M**, se puede estimar los parámetros para la distribución del tipo de manzanas Fuji y Gala con las etiquetas correctas.

De vuelta al ejemplo

- Al realizar el **paso M**, se puede estimar los parámetros para la distribución del tipo de manzanas Fuji y Gala con las etiquetas correctas.
- Esto es, encontrar los parámetros que maximizan la probabilidad dadas las etiquetas de las manzanas.

De vuelta al ejemplo

- Al realizar el **paso M**, se puede estimar los parámetros para la distribución del tipo de manzanas Fuji y Gala con las etiquetas correctas.
- Esto es, encontrar los parámetros que maximizan la probabilidad dadas las etiquetas de las manzanas.



De vuelta al ejemplo

- Ahora se deben repetir los pasos E y M una y otra vez hasta que no haya cambios en la asignación de tipo de manzana.

De vuelta al ejemplo

- Ahora se deben repetir los pasos E y M una y otra vez hasta que no haya cambios en la asignación de tipo de manzana.
- Esto es, hasta que el cambio en la función de probabilidad sea muy pequeño.

Consideraciones para una buena ejecución

- Considere un proceso de etiquetado inicial completamente aleatorio.

Consideraciones para una buena ejecución

- Considere un proceso de etiquetado inicial completamente aleatorio.
- Teóricamente no está garantizado que cada inicialización aleatoria ocasione el mismo resultado del algoritmo EM para GMM.

Consideraciones para una buena ejecución

- Considere un proceso de etiquetado inicial completamente aleatorio.
- Teóricamente no está garantizado que cada inicialización aleatoria ocasione el mismo resultado del algoritmo EM para GMM.
- El proceso mejorará si se tienen manzanas correctamente etiquetadas.

Consideraciones para una buena ejecución

- Considere un proceso de etiquetado inicial completamente aleatorio.
- Teóricamente no está garantizado que cada inicialización aleatoria ocasione el mismo resultado del algoritmo EM para GMM.
- El proceso mejorará si se tienen manzanas correctamente etiquetadas.
- ¿Qué ocurre si las manzanas no difieren mucho en su tamaño de acuerdo a su tipo?

Consideraciones para una buena ejecución

- Considere un proceso de etiquetado inicial completamente aleatorio.
- Teóricamente no está garantizado que cada inicialización aleatoria ocasione el mismo resultado del algoritmo EM para GMM.
- El proceso mejorará si se tienen manzanas correctamente etiquetadas.
- ¿Qué ocurre si las manzanas no difieren mucho en su tamaño de acuerdo a su tipo?
- En ese caso no se podría hacer una separación apropiada.

Consideraciones para una buena ejecución

- Considere un proceso de etiquetado inicial completamente aleatorio.
- Teóricamente no está garantizado que cada inicialización aleatoria ocasione el mismo resultado del algoritmo EM para GMM.
- El proceso mejorará si se tienen manzanas correctamente etiquetadas.
- ¿Qué ocurre si las manzanas no difieren mucho en su tamaño de acuerdo a su tipo?
- En ese caso no se podría hacer una separación apropiada.
- Por ello se puede considerar tener más características de las manzanas, por ejemplo: sabor, olor o color.

Dos características

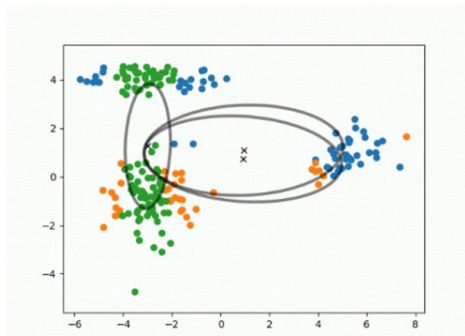
- En el caso de que se tengan dos características, las distribuciones Gaussianas 2-D se pueden visualizar como una elipse en el espacio de características.

Dos características

- En el caso de que se tengan dos características, las distribuciones Gaussianas 2-D se pueden visualizar como una elipse en el espacio de características.
- La figura muestra el proceso del algoritmo EM para un modelo de mezclas Gaussianas con 3 componentes Gaussianas.

Dos características

- En el caso de que se tengan dos características, las distribuciones Gaussianas 2-D se pueden visualizar como una elipse en el espacio de características.
- La figura muestra el proceso del algoritmo EM para un modelo de mezclas Gaussianas con 3 componentes Gaussianas.
- Siguiendo con las manzanas, sería la tarea de separar manzanas Fuji, Gala y Honeycrisp de acuerdo a las características tamaño y sabor.



Contenido

- 1 Modelos de Mezclas Gaussianas
- 2 ¿Qué hace el algoritmo EM en el GMM?
- 3 Algoritmo EM
- 4 Definición del GMM**

Mezcla Gaussiana

Mezcla Gaussiana

Una mezcla Gaussiana es una función compuesta de un conjunto de Gaussianas, cada una identificada por $k \in \{1, 2, \dots, K\}$, donde K es el número de clusters de los datos.

Mezcla Gaussiana

Mezcla Gaussiana

Una mezcla Gaussiana es una función compuesta de un conjunto de Gaussianas, cada una identificada por $k \in \{1, 2, \dots, K\}$, donde K es el número de clusters de los datos.

Cada Gaussiana k en la mezcla está compuesta de los siguientes parámetros:

Mezcla Gaussiana

Una mezcla Gaussiana es una función compuesta de un conjunto de Gaussianas, cada una identificada por $k \in \{1, 2, \dots, K\}$, donde K es el número de clusters de los datos.

Cada Gaussiana k en la mezcla está compuesta de los siguientes parámetros:

- La media μ que define su centro.

Mezcla Gaussiana

Una mezcla Gaussiana es una función compuesta de un conjunto de Gaussianas, cada una identificada por $k \in \{1, 2, \dots, K\}$, donde K es el número de clusters de los datos.

Cada Gaussiana k en la mezcla está compuesta de los siguientes parámetros:

- La media μ que define su centro.
- La covarianza Σ que define su ancho. Esta sería equivalente a las dimensiones de la elipsoide en un caso multivariado.

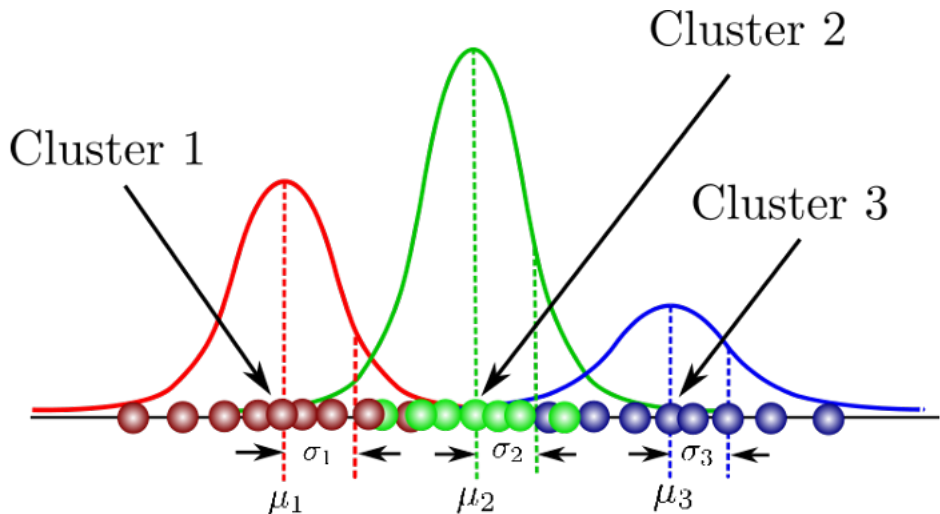
Mezcla Gaussiana

Una mezcla Gaussiana es una función compuesta de un conjunto de Gaussianas, cada una identificada por $k \in \{1, 2, \dots, K\}$, donde K es el número de clusters de los datos.

Cada Gaussiana k en la mezcla está compuesta de los siguientes parámetros:

- La media μ que define su centro.
- La covarianza Σ que define su ancho. Esta sería equivalente a las dimensiones de la elipsoide en un caso multivariado.
- La probabilidad de mezcla π que define cuan pequeña o grande será la función Gaussiana.

Mezcla Gaussiana



Los coeficientes de mezcla son probabilidades que deben satisfacer la condición:

$$\sum_{k=1}^K \pi_k = \pi_1 + \pi_2 + \pi_3 + \cdots + \pi_K = 1 \quad (2)$$

Para determinar los valores óptimos, se debe asegurar que cada Gaussiana cubra todos los datos del cluster correspondiente. Esto es lo que maximizar la probabilidad (semejanza) hace.

Función de densidad Gaussiana

La función de densidad Gaussiana está dada por:

$$N(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad (3)$$

donde \mathbf{x} representa los datos, D es el número de dimensiones para cada punto, μ es la media y Σ la covarianza.

Función de densidad Gaussiana

Si se tiene un conjunto de datos compuesto de $N = 1000$ puntos en tres dimensiones ($D = 3$), entonces \mathbf{x} sería una matriz 1000×3 , μ sería un vector 1×3 y Σ sería una matriz 3×3 . Si se aplica la función logaritmo a la ecuación (3) se tiene:

$$\ln N(\mathbf{x}|\mu, \Sigma) = -\frac{D}{2} \ln 2\pi - \frac{1}{2} \ln \Sigma - \frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \quad (4)$$

Si se deriva esta expresión con respecto a μ y Σ , y se iguala a cero entonces se puede obtener los valores óptimos para esos parámetros. La solución corresponderá a la Estimación Máxima de Semejanza (MLE).