Algoritmo Maximización de Esperanzas _

Motivación

¿Qué es?

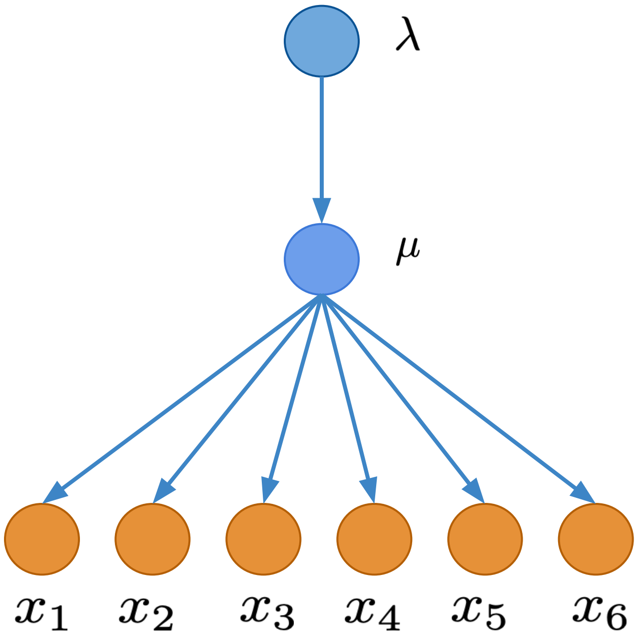
- Es un algoritmo que a grandes rasgos permite encontrar información incompleta en un modelo.
- Presenta una serie de implementaciones. Por efectos expositivos nos concentraremos en un modelo de **Mezcla de Gaussianas** (GMM).

Casos de uso:

- **Variables Latentes**: Patrones subyacentes en los datos.
- **Problemas de datos perdidos**: patrones de datos faltantes existentes en una matriz de datos.
- **Procesamiento de Imágenes**: reconstrucción de tomografías.
- Manejo de Riesgo de portafolios.

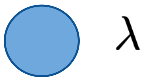
El problema con Máxima Verosimilitud

Estimación por Máxima Verosimilitud (EMV)



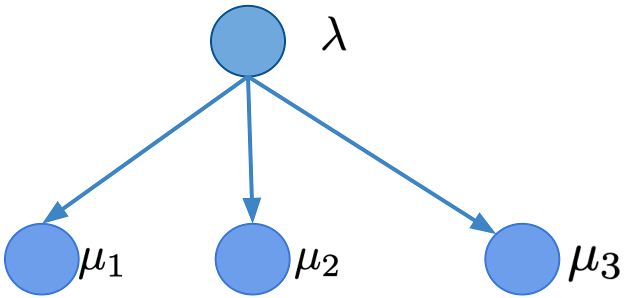
- Nuestro objetivo es obtener algún parámetro dado los datos, que refleje el fenómeno latente.
- EMV presentará la solución más eficiente cuando sólo exista un parámetro a identificar.

Limitantes de EMV



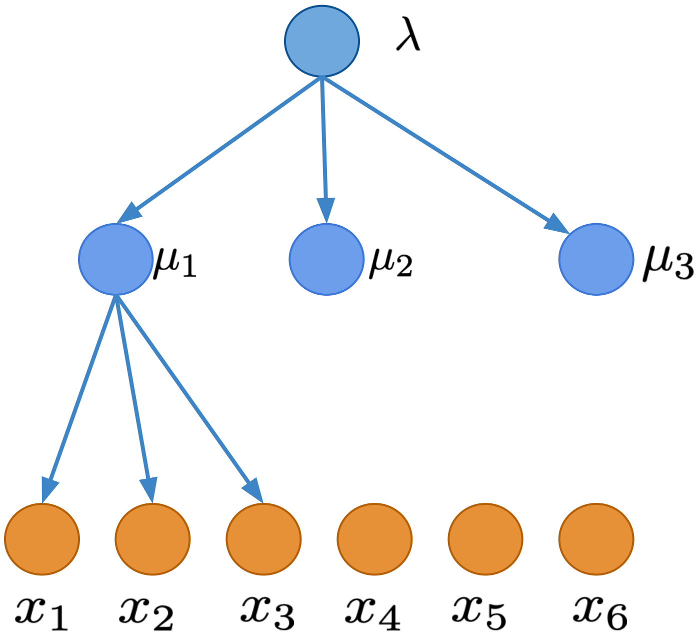
- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.

Limitantes de EMV



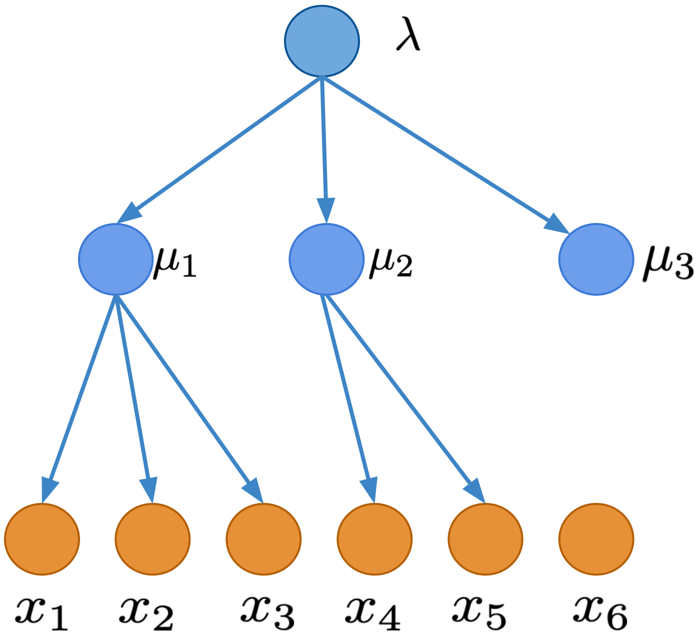
- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Limitantes de EMV



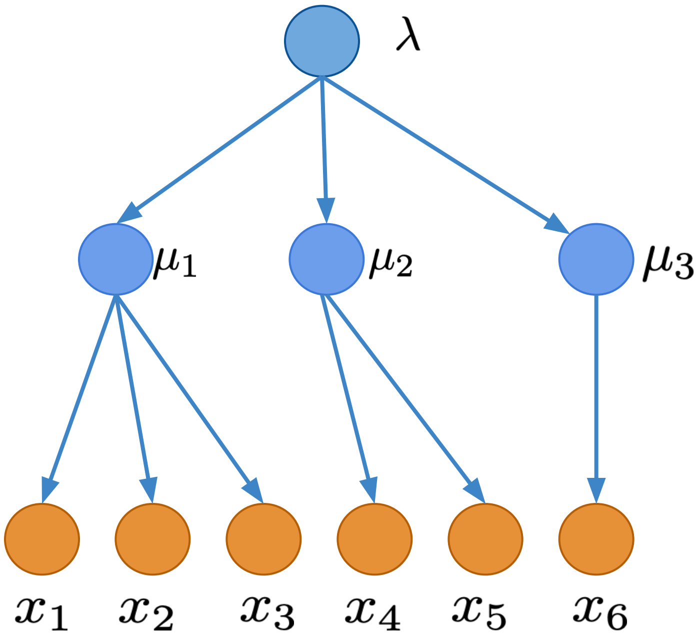
- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Limitantes de EMV



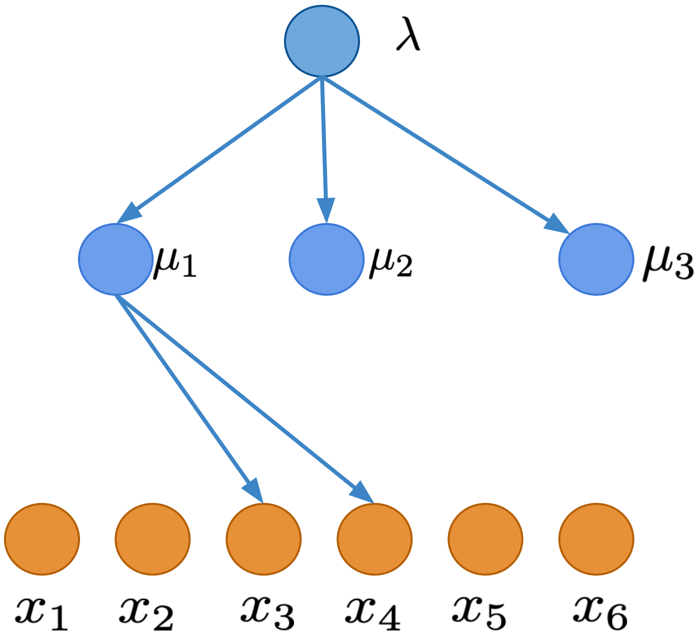
- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Limitantes de EMV



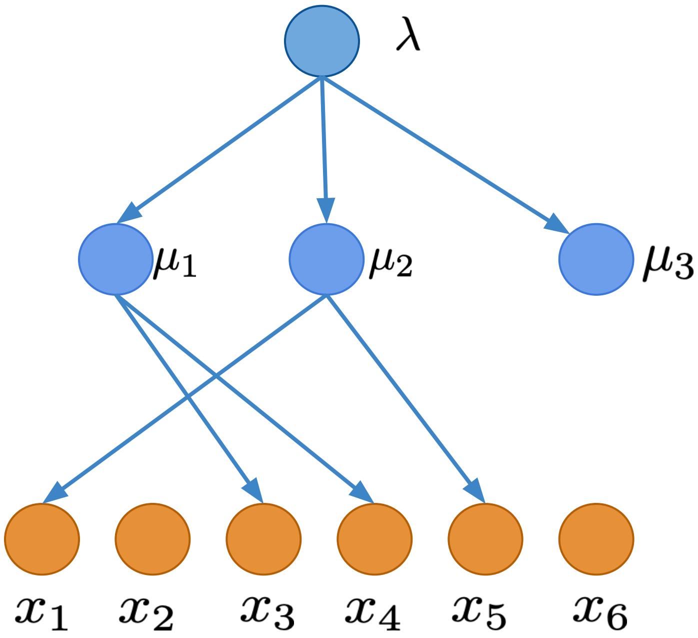
- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Limitantes de EMV



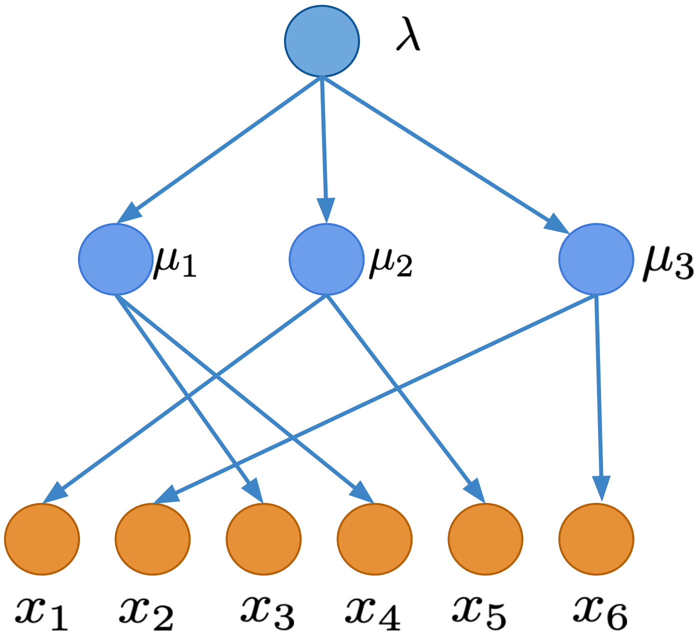
- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Limitantes de EMV



- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Limitantes de EMV



- El problema de EMV surge cuando no tenemos suficiente información para observar todo lo que deseamos rescatar.
- Lo que necesitamos es encontrar un método que permita identificar la cantidad de parámetros cuando no tengamos conocimiento previo de éstos.

Mecanismo de Acción

Ejemplo: Identificación de dos clusters en 1-d

- En un caso supervisado, podemos perfilar la pertenencia de cada observación dado que tenemos información sobre las clases (cantidad y afiliación).

Ejemplo: Identificación de dos clusters en 1-d

$$\phi_{x|y=1} = \text{Normal}(\mu_1, \sigma_1^2)$$

- En un caso supervisado, podemos perfilar la pertenencia de cada observación dado que tenemos información sobre las clases (cantidad y afiliación).
- De esta manera, la identificación de un modelo simplemente dependerá la evaluación (condicional) de los parámetros.

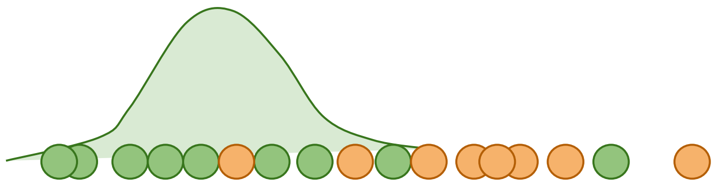
Ejemplo: Identificación de dos clusters en 1-d

$\phi_{x|y=1} = \text{Normal}(\mu_1, \sigma_1^2)$ $\phi_{x|y=2} = \text{Normal}(\mu_2, \sigma_2^2)$

- En un caso supervisado, podemos perfilar la pertenencia de cada observación dado que tenemos información sobre las clases (cantidad y afiliación).
- De esta manera, la identificación de un modelo simplemente dependerá la evaluación (condicional) de los parámetros.

Ejemplo: Identificación de dos clusters en 1-d

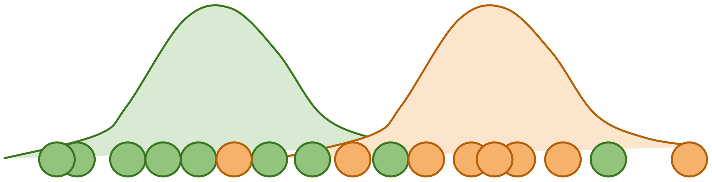
$\phi_{x|y=1} = \text{Normal}(\mu_1, \sigma_1^2)$ $\phi_{x|y=2} = \text{Normal}(\mu_2, \sigma_2^2)$



- En un caso supervisado, podemos perfilar la pertenencia de cada observación dado que tenemos información sobre las clases (cantidad y afiliación).
- De esta manera, la identificación de un modelo simplemente dependerá la evaluación (condicional) de los parámetros.

Ejemplo: Identificación de dos clusters en 1-d

$\phi_{x|y=1} = \mathrm{Normal}(\mu_1, \sigma_1^2)$ $\phi_{x|y=2} = \mathrm{Normal}(\mu_2, \sigma_2^2)$



- En un caso supervisado, podemos perfilar la pertenencia de cada observación dado que tenemos información sobre las clases (cantidad y afiliación).
- De esta manera, la identificación de un modelo simplemente dependerá la evaluación (condicional) de los parámetros.

Ejemplo: Identificación de dos clusters en 1-d

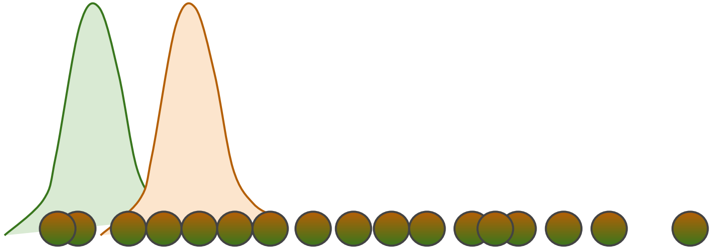
$$\phi_{x|y=?} = \mathsf{Normal}(?,?)$$

En un caso no supervisado, el trabajo de clasificar conlleva una serie de pasos previos:

- Definir cuál es la cantidad óptima de clases.
- Inferir cuáles son los principales atributos de una clase.
- Asignar la clase más probable.

Ejemplo: Identificación de dos clusters en 1-d

$$\phi_{x|y=?} = \text{Normal}(?, ?)$$

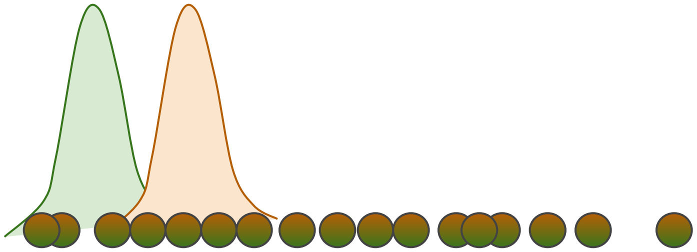


En un caso no supervisado, el trabajo de clasificar conlleva una serie de pasos previos:

- Definir cuál es la cantidad óptima de clases
- Inferir cuáles son los principales atributos de una clase.
- Asignar la clase más probable.

Ejemplo: Identificación de dos clusters en 1-d

$$\phi_{x|y=?} = \mathsf{Normal}(?,?)$$

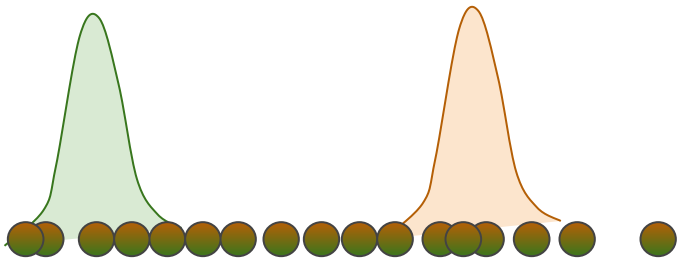


En un caso no supervisado, el trabajo de clasificar conlleva una serie de pasos previos:

- Definir cuál es la cantidad óptima de clases.
- Inferir cuáles son los principales atributos de una clase.
- Asignar la clase más probable.

Ejemplo: Identificación de dos clusters en 1-d

$$\phi_{x|y=?} = \text{Normal}(?, ?)$$

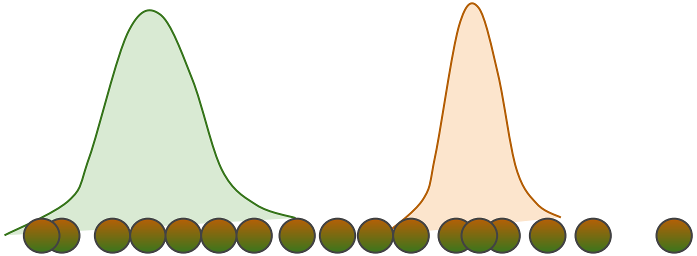


En un caso no supervisado, el trabajo de clasificar conlleva una serie de pasos previos:

- Definir cuál es la cantidad óptima de clases
- Inferir cuáles son los principales atributos de una clase.
- Asignar la clase más probable.

Ejemplo: Identificación de dos clusters en 1-d

$$\phi_{x|y=?} = \text{Normal}(?, ?)$$

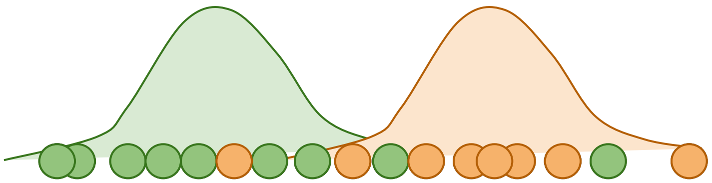


En un caso no supervisado, el trabajo de clasificar conlleva una serie de pasos previos:

- Definir cuál es la cantidad óptima de clases
- Inferir cuáles son los principales atributos de una clase.
- Asignar la clase más probable.

Ejemplo: Identificación de dos clusters en 1-d

$$\phi_{x|y=?} = \mathsf{Normal}(?,?)$$

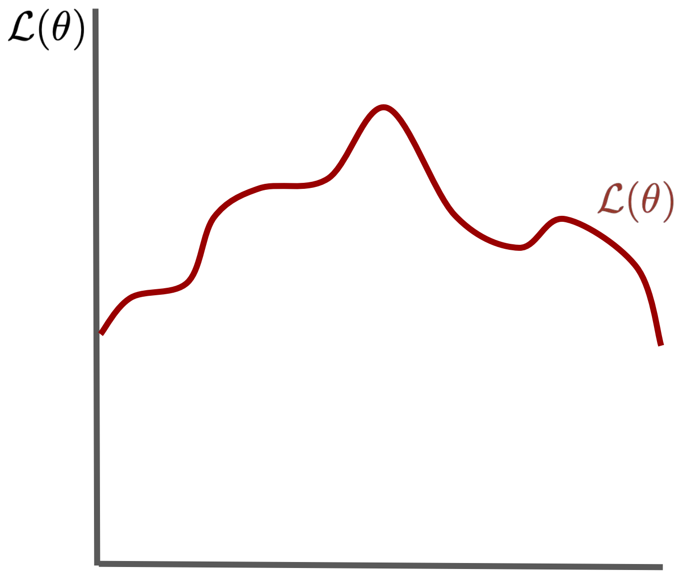


En un caso no supervisado, el trabajo de clasificar conlleva una serie de pasos previos:

- Definir cuál es la cantidad óptima de clases
- Inferir cuáles son los principales atributos de una clase.
- Asignar la clase más probable.

El proceso iterativo de EM

Proceso Iterativo de EM

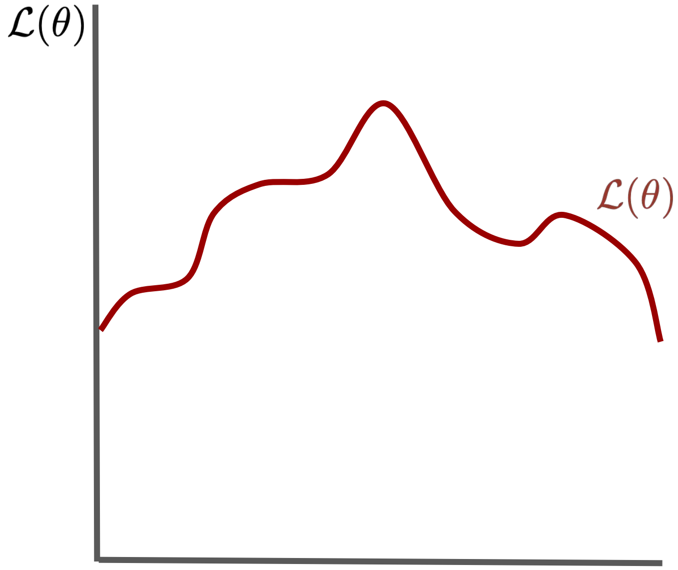


El algoritmo de Maximización de Esperanzas es un proceso iterativo.

Este se compone de dos pasos:

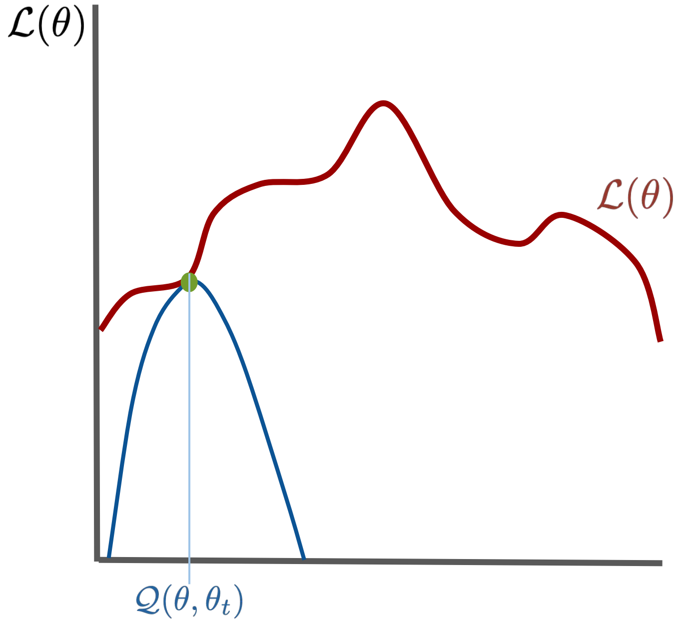
- **Paso E:** Obtención de la esperanza y un intervalo inferior en la verosimilitud.
- **Paso M:** Tomando el punto donde ambas funciones presentan la misma gradiente, actualizamos.

Proceso Iterativo de EM



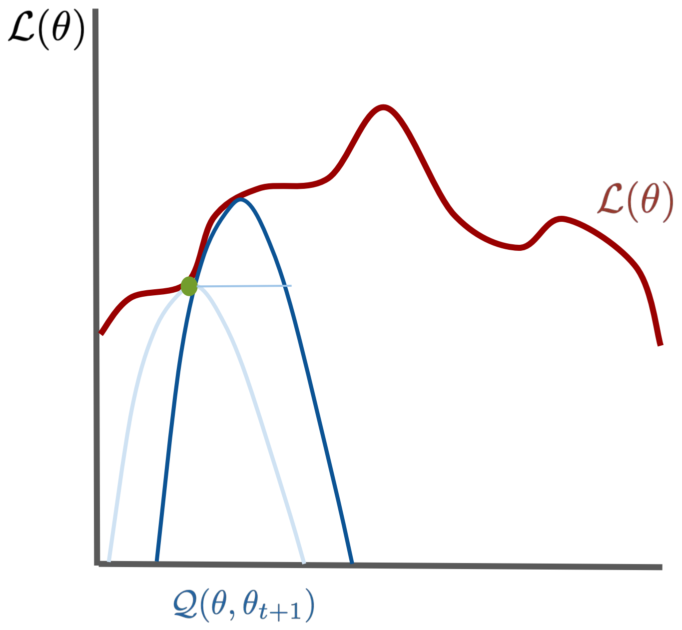
- Tenemos una función de verosimilitud observada incompleta.

Proceso Iterativo de EM - Paso E



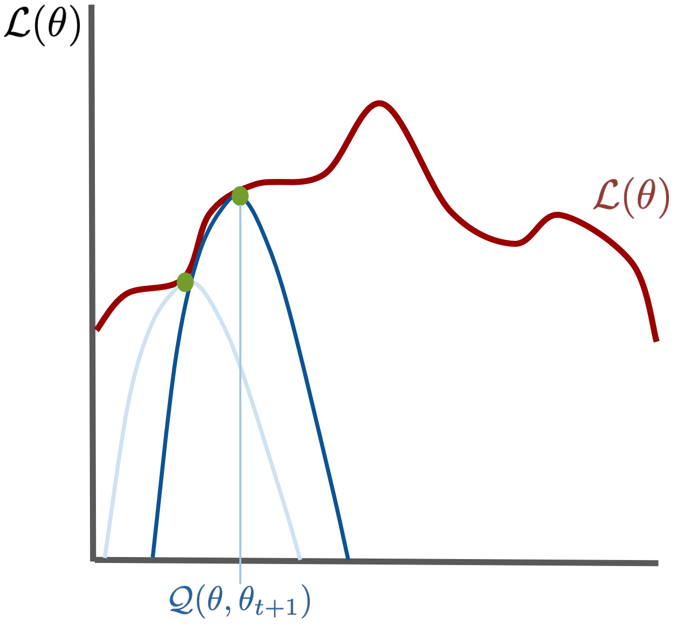
- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.

Proceso Iterativo de EM - Paso M



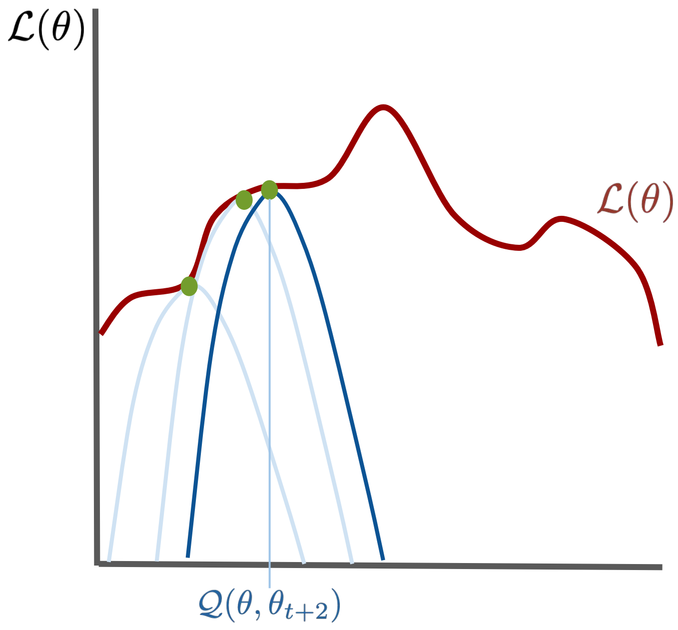
- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.

Proceso Iterativo de EM



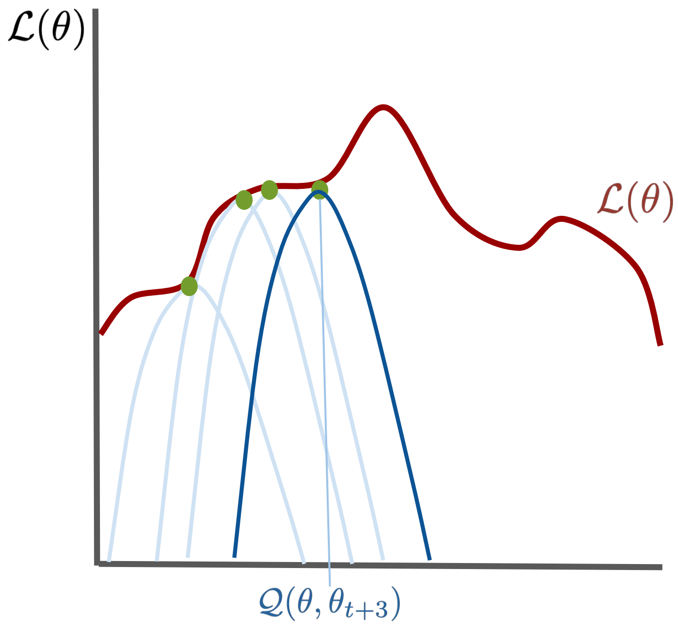
- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.

Proceso Iterativo de EM



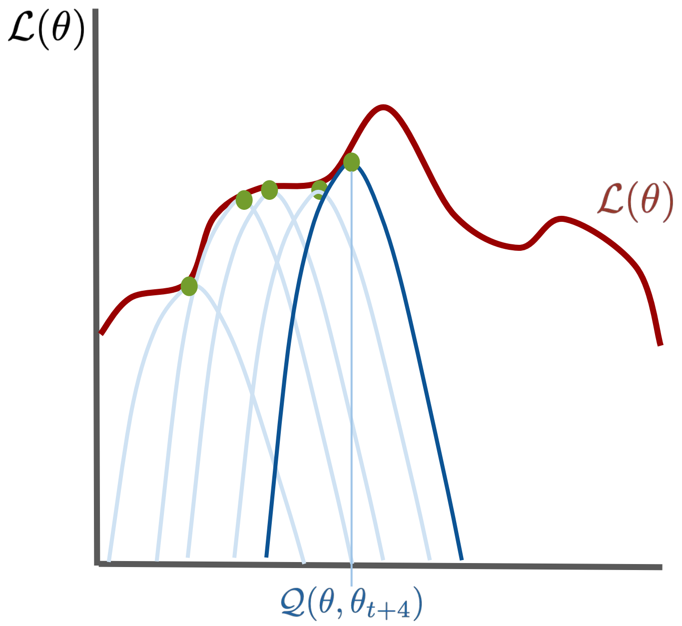
- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.
- Iterar hasta que EM alcance un punto de estabilidad.

Proceso Iterativo de EM



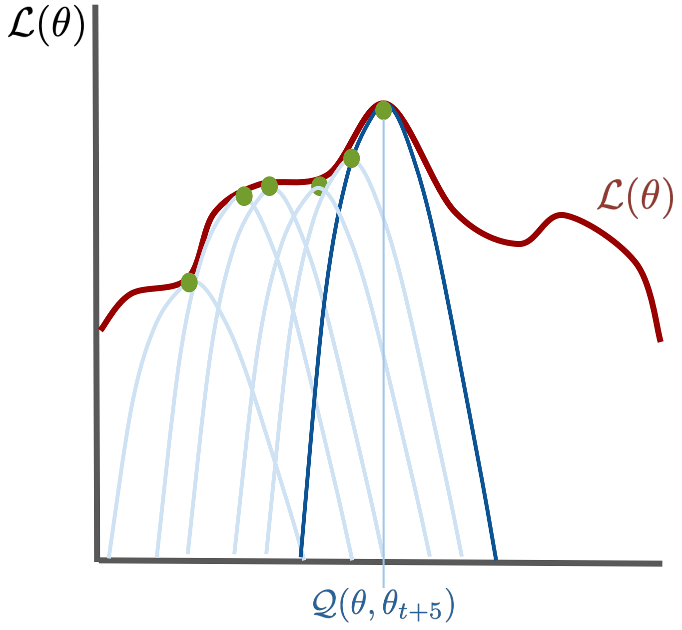
- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.
- Iterar hasta que EM alcance un punto de estabilidad.

Proceso Iterativo de EM



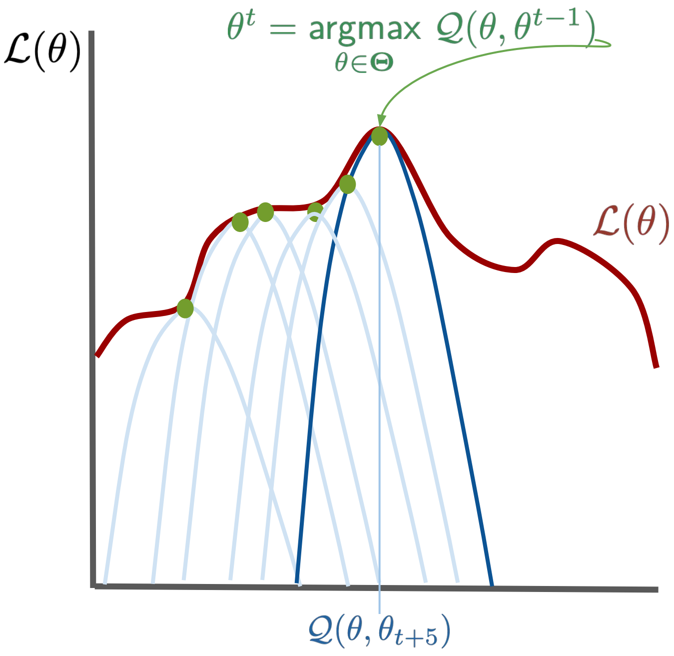
- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.
- Iterar hasta que EM alcance un punto de estabilidad.

Proceso Iterativo de EM



- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.
- Iterar hasta que EM alcance un punto de estabilidad.

Proceso Iterativo de EM



- Tenemos una función de verosimilitud observada incompleta.
- El objetivo es perfilar un mínimo en la función de verosimilitud local.
- En base al intervalo inferior de éste, estimar el siguiente punto.
- Iterar hasta que EM alcance un punto de estabilidad.
- Se dice que EM aumenta de forma monotónica la verosimilitud observada.
- De esta manera, garantiza la obtención de por lo menos un mínimo local.

Criterios de Información

Elementos a considerar

- La aplicación de Mezcla de Gaussianas es un modelo **no supervisado**.
- No tenemos información sobre la cantidad de clusters existentes.
- Fundamentado en el algoritmo EM, evaluamos la logverosimilitud de cada modelo propuesto mediante criterios de información.

AIC y (S)BIC

Criterio de Información de Akaike

$$\text{AIC} = \underbrace{-2}_{\text{Penalización}}(\underbrace{\mathcal{L}_{\text{modelo}}}_{\text{Verosimilitud del modelo propuesto}} - \underbrace{\#\ \text{params}}_{\text{Cantidad de parámetros estimados}})$$

Criterio de Información de Schwarz-Bayes

$$\text{(S)BIC} = \underbrace{\ln(\#\ \text{params})}_{\text{Penalización}}(\underbrace{\mathcal{L}_{\text{modelo}}}_{\text{Verosimilitud del modelo propuesto}} - \underbrace{\#\ \text{params}}_{\text{Cantidad de parámetros estimados}})$$

- Ambos estadísticos de prueba permiten evaluar el ajuste en la muestra para estimar la verosimilitud de un modelo en su capacidad predictiva.
- **Diagnóstico**: encontrar el modelo que presente un menor puntaje asociado.
- Las soluciones propuestas por (S)BIC tienden a ser más parsimoniosas.

{desafío}
latam_

Academia de
talentos digitales

www.desafiolatam.com