

# Trabajo práctico 0: Algoritmo de Maximización de la Esperanza

Ph. D. Saúl Calderón Ramírez  
Instituto Tecnológico de Costa Rica,  
Escuela de Computación  
PAttern Recongition and MACHine Learning Group (PARMA-Group)

14 de febrero de 2024

**Fecha de entrega:** Miercoles 13 de Marzo.

**Entrega:** Un archivo .zip con el código fuente LaTeX o Lyx, el pdf, y un jupyter en Pytorch, debidamente documentado, con una función definida por ejercicio. A través del TEC-digital.

**Modo de trabajo:** Grupos de 2 a 3 personas.

## Resumen

En el presente trabajo práctico introduce al algoritmo de maximización de esperanza, y su aplicación para la segmentacion de imagenes. El trabajo practico consta de 100 puntos, mas 30 puntos extra.

## 1. (80 puntos) Algoritmo de Maximización de la Esperanza con datos artificiales (modelo Gaussiano)

A continuación, implemente el algoritmo de maximización de la esperanza (descrito en el material del curso), usando la definición y descripción de las siguientes funciones como base:

1. **(15 puntos)** Implemente la función *generate\_data* la cual reciba la cantidad de observaciones unidimensionales total a generar  $N$ , y los parámetros correspondientes a  $K = 2$  funciones de densidad **Gaussianas**. Genere los datos siguiendo tales distribuciones, y retorne tal matriz de datos  $X \in \mathbb{R}^{N \times 1}$ .
  - a) Grafique los datos usando un *scatter plot* junto con las gráficas de los histogramas de los datos y las funciones de densidad de probabilidad **Gaussianas** usando los parámetros para inicializar los datos, en la misma figura (gráfico). Escoja 3 conjuntos de parámetros para mostrar los 3 dataset resultantes.

2. **(5 puntos)** Implemente la función *init\_random\_parameters* la cual genere una matriz de  $P \in \mathbb{R}^{K \times 2}$  dimensiones, con los parámetros de las funciones de densidad **Gaussiana** generados completamente al azar.
  - a) Muestre un pantallazo donde verifique su funcionamiento correcto con los comentarios asociados.
3. **(20 puntos)** Implemente la función *calculate\_likelihood\_gaussian\_observation*( $x_n$ ,  $\mu_k$ ,  $\sigma_k$ ) la cual calcule la verosimilitud de una observación específica  $x_n$ , para una función de densidad Gaussiana con parámetros  $\mu_k$  y  $\sigma_k$ . Realice la corrección pertinente al calculo de la función de verosimilitud para evitar el problema de *under flow*.
  - a) Diseñe y ejecute una prueba unitaria donde verifique su funcionamiento correcto con los comentarios asociados.
4. **(10 puntos)** Implemente la función *calculate\_membership\_dataset*( $X_{dataset}$ ,  $Parameters_{matrix}$ ), la cual, usando la matriz de parámetros  $P$  y la función anteriormente implementada *calculate\_likelihood\_gaussian\_observation*, defina por cada observación  $x_n \in X$  la pertenencia o membresía a cada cluster  $k = 1, \dots, K$ , en una matriz binaria  $M \in \mathbb{R}^{N \times K}$ . Retorne tal matriz de membresía  $M$ .
  - a) Diseñe y ejecute una prueba unitaria donde verifique su funcionamiento correcto con los comentarios asociados.
5. **(10 puntos)** Implemente la función *recalculate\_parameters*( $X_{dataset}$ ,  $Membership_{data}$ ), la cual recalculé los parámetros de las funciones de densidad **Gaussianas** representadas en la matriz  $P$ , de acuerdo a lo representado en la matriz de membresía  $M$ .
  - a) Use las funciones *mean* y *std* de pytorch para ello. Intente prescindir al máximo de estructuras de repetición tipo *for*.
6. **(10 puntos)** Ejecute 5 corridas diferentes del algoritmo, donde por cada una documente los parámetros a los que se arribó, junto con los parámetros iniciales (describalo en una tabla). Hagalo para los 3 conjuntos de datos generados anteriormente.
  - a) Grafique las funciones de densidad de probabilidad a las que convergió el algoritmo. Puede graficar también las funciones de densidad obtenidas en 2 o 3 pasos intermedios. Presente una tabla de gráficas donde en cada entrada se identifique el número de iteración y los parámetros iniciales.
  - b) Comente los resultados.
7. **(10 puntos)** Proponga una mejor heurística para inicializar los parámetros del modelo aleatoriamente.
  - a) Compruebe la mejora obtenida con el método propuesto, corriendo las pruebas del punto anterior.

## 2. (40 puntos) Algoritmo de Maximización de la Esperanza con datos artificiales (modelo Gamma)

En esta sección, modificará el algoritmo de maximización de la esperanza anteriormente implementado para un modelo Gaussiano, para el modelo Gamma.

1. **(20 puntos)** Enliste e implemente todas las modificaciones necesarias tanto conceptualmente como en código para hacer que el algoritmo de maximización de la esperanza utilice un modelo Gamma.
2. **(5 puntos)** Implemente la función *generate\_data* la cual reciba la cantidad de observaciones unidimensionales total a generar  $N$ , y los parámetros correspondientes a  $K = 2$  funciones de densidad **Gamma**. Genere los datos siguiendo tales distribuciones, y retorne tal matriz de datos  $X \in \mathbb{R}^{N \times 1}$ .
  - a) Grafique los datos usando un *scatter plot* junto con las gráficas de los histogramas de los datos y las funciones de densidad de probabilidad **Gamma** usando los parámetros para inicializar los datos, en la misma figura (gráfico). Los parámetros serán escogidos por usted, muestre al menos 3 generaciones del dataset y comente el efecto de los parámetros.
3. **(10 puntos)** Ejecute 5 corridas diferentes del algoritmo, donde por cada una documente los parámetros a los que se arribó, junto con los parámetros iniciales (describalos en una tabla). Hagalo para los 3 conjuntos de datos generados anteriormente.
  - a) Grafique las funciones de densidad de probabilidad a las que convergió el algoritmo. Puede graficar también las funciones de densidad obtenidas en 2 o 3 pasos intermedios. Presente una tabla de gráficas donde en cada entrada se identifique el número de iteración y los parámetros iniciales.
    - 1) Comente los resultados.
4. **(5 puntos)** Explique cual es el comportamiento esperado de utilizar la implementación del algoritmo de maximización de esperanza para una distribución Gamma, con un conjunto de datos que sigue una distribución Gaussiana y viceversa.
  - a) **(10 puntos)** Realice un diseño experimental donde verifique lo anterior