

# Aprendizaje Automático: Trabajo práctico 2

PATtern Recongition and MACHine Learning Group (PARMA-Group)

24 de mayo de 2022

**Fecha de entrega:** Martes 31 de Mayo del 2022.

**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.

## 1. Implementación del algoritmo K-medias

1. Genere un conjunto de datos  $X \in \mathbb{R}^{D \times N}$  con  $D = 2$  y  $N = 200$ , de  $K = 2$  etiquetas, donde:
  - a) 100 observaciones correspondan a un *cluster* y las otras 100 a otro, generados por una VA Gaussiana con dos medias  $\vec{\mu}_1, \vec{\mu}_2$  y matrices de covarianza  $\Sigma_1$  y  $\Sigma_2$  (escoja los valores de forma que haya covarianzas no nulas, y varianzas diferentes en cada dimension, haciendo que los datos roten).
  - 1) Genere un caso en el que las medias y las covarianzas faciliten la separación de los datos,  $X_1$ .
  - 2) Genere otro caso en el que las medias y covarianzas generen datos traslapados,  $X_2$ .
  - b) Grafique las muestras, con las etiquetas correctas y sin las etiquetas, usando una simbología de su preferencia.
2. Implemente alguno de los algoritmos de aprendizaje no supervisado vistos en clase (K-medias), para realizar la clasificación de los dos conjuntos  $X_1$  y  $X_2$ . **Hagalo de forma completamente matricial usando Pytorch.**
  - a) Realice una implementación usando la distancia Euclidiana para medir la disimilitud de las muestras.

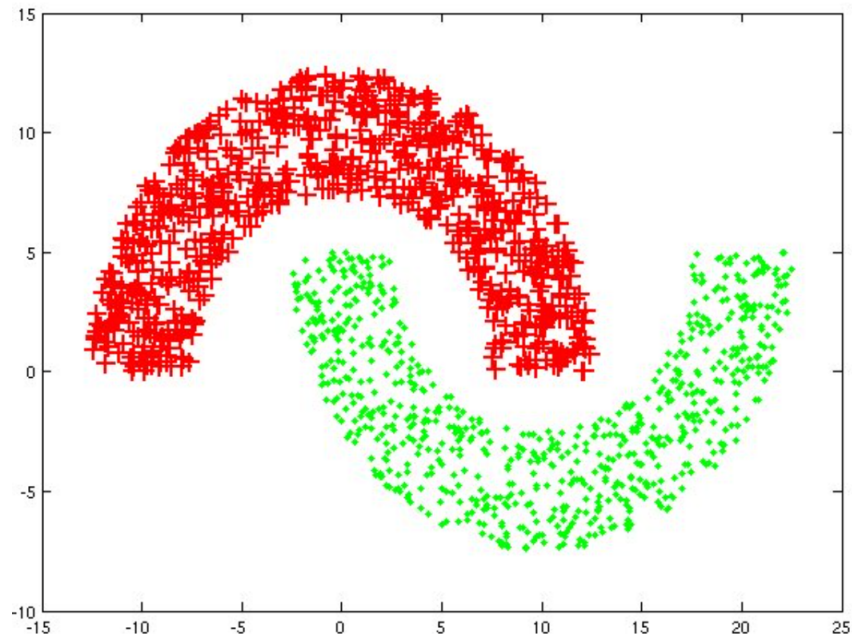


Figura 1: Datos en configuracion de *dos medias lunas*.

- b) Proponga e implemente el algoritmo usando otra distancia, como la distancia L1 o de Mahalanobis.
- c) Reporte los resultados midiendo la tasa de aciertos (media y desviacion estandar), con gráficas que muestren las etiquetas generadas por el algoritmo. Realicelo para 10 corridas.
  - 1) Reporte ademas el tiempo de ejecucion promedio y la desviacion estandar para cada variante con cada distancia.

## 2. (20 puntos extra) Implementación del algoritmo de Corte de grafos

1. Implemente el algoritmo de corte de grafos visto en clase.
2. Investigue como generar el conjunto de datos conocido como las *dos medias lunas* e implementelo en pytorch, como se muestra en la Figura 1.
3. Compruebe el resultado de realizar el agrupamiento usando el algoritmo

de corte de grafos, usando el conjunto de datos de las *dos medias lunas*, usando una separacion *grande* y otra *pequena*.

- a) Grafique los resultados del agrupamiento.
- b) Reporte la tasa de aciertos respecto a una particion distinta de datos (particion de *test*) generada con la misma distribucion.
- c) Realice los dos pasos anteriores para el algoritmo de K-medias previamente implementado. Compare y comente los resultados.