

Matemática para ciencias de los datos:

Trabajo práctico 2

M. Sc. Saúl Calderón Ramírez
Instituto Tecnológico de Costa Rica,
Programa de Ciencias de los Datos,
PAttern Recognition and MACHine Learning Group (PARMA-Group)

20 de mayo de 2021

El presente proyecto introduce conceptos de optimización y aprendizaje automático.

- **Fecha de entrega:** 31 de May. 2021
- **Modo de trabajo:** Grupo de tres personas.
- **Tipo de entrega:** digital, por medio de la plataforma TEC-digital (Jupyter y pdf).

Parte I

Proyecciones y distancia de Mahalanobis

1. (15 puntos) La distancia Euclidiana

1. Genere los datos usando la función *create_data_one_mode* y grafique los puntos en la matrix $X \in \mathbb{R}^{n \times 2}$ (con $n = 100$), además de los vectores (como puntos, usando la función *scatter*) \vec{y}_1 , \vec{y}_2 y $\vec{\mu}$, con un símbolo gráfico distinto para cada uno.
 - a) Calcule y reporte en el documento las siguientes distancias euclidianas $d_{\ell_2}(\vec{\mu}, \vec{y}_1)$ y $d_{\ell_2}(\vec{\mu}, \vec{y}_2)$ (Puede usar la librería de Pytorch para ello) ¿Qué tan similares son según esa distancia?
 - b) A partir de la **distribución de los puntos**, explique cual de los dos vectores \vec{y}_1 y \vec{y}_2 es más *verosímil* o posible.

- 1) Realice 3 corridas (generaciones de los datos), para corroborar su observación. Muestre los gráficos de dispersión de las 3 corridas.

2. (25 puntos) Blanqueo PCA, sin normalización por la varianza

A continuación, se detalla el procedimiento a seguir para normalizar la rotación de los datos, a implementarse en la función `calcular_blanqueo_PCA1(X, y1, y2)`.

1. Calcule la matriz de covarianza para los datos, Σ , usando el **método programado por usted anteriormente**.
 - a) Transcribala en el reporte.
 - b) Porqué la matriz de covarianza calculada es similar a la matriz de covarianza usada para generar los datos?
2. Calcule los autovectores \vec{v}_1 y \vec{v}_2 de la matriz Σ .
 - a) Transcribalos en el reporte.
3. Para todos los vectores $\vec{x}_i \in X$, calcule las muestras con la media sustraída $\vec{u}_i = \vec{x}_i - \vec{\mu}$.

a) Acomode los datos formando la matriz $U = \begin{bmatrix} - & \vec{u}_1 & - \\ - & \vec{u}_2 & - \\ & \vdots & \\ - & \vec{u}_n & - \end{bmatrix}$.

4. Proyecte los vectores \vec{u}_i usando como conjunto base los auto-vectores \vec{v}_1 y \vec{v}_2 , con la ecuación para cada componente del vector proyectado \vec{u}'_i :

$$\vec{u}'_{i,1} = \frac{\vec{u}_i \cdot \vec{v}_1}{\sigma_1}$$

$$\vec{u}'_{i,2} = \frac{\vec{u}_i \cdot \vec{v}_2}{\sigma_2}$$

la matriz con todos los puntos proyectados viene dada entonces por $U' =$

$$\begin{bmatrix} - & \vec{u}'_1 & - \\ - & \vec{u}'_2 & - \\ & \vdots & \\ - & \vec{u}'_n & - \end{bmatrix}. \text{ Los coeficientes } \sigma_1 \text{ y } \sigma_2 \text{ corresponden a las desviaciones}$$

estandar, la cuales están contenidas en la matriz de covarianza Σ .

- a) **Realice tal proyección expresada de forma matricial, para el conjunto de observaciones representado en la matriz U .**

- 1) Escriba la fórmula de la proyección matricial en el documento.
5. Projete también los vectores $\vec{w}_1 = \vec{y}_1 - \vec{\mu}$ y $\vec{w}_2 = \vec{y}_2 - \vec{\mu}$ usando el espacio base conformado por los auto-vectores \vec{v}_1 y \vec{v}_2 , para lograr los vectores proyectados \vec{y}_1' y \vec{y}_2' .
 - a) Escriba e implemente también tal proyección de forma matricial, con

$$W = \begin{bmatrix} - & w_1 & - \\ - & \vec{w}_2 & - \end{bmatrix}.$$
 Use la función *cat* para ello.
6. Grafique todos los puntos en U' además de \vec{y}_1' y \vec{y}_2' , usando la simbología similar a la usada anteriormente.
7. Calcule la matriz de covarianza Σ para los datos en U' , y relacione los valores de covarianza con lo graficado en el punto anterior
 - a) ¿Porqué las covarianzas son nulas en tal matriz? Justifique su respuesta con base a la gráfica anterior.
8. Calcule y reporte en el documento las siguientes distancias euclidianas $d_{\ell_2}(\vec{\mu}', \vec{y}_1')$ y $d_{\ell_2}(\vec{\mu}', \vec{y}_2')$ con $\vec{\mu}'$ el centroide de los datos U' . ¿Qué tan similares son al centroide según esa distancia? Relacione los datos con lo graficado en el apartado 6.

3. (20 puntos) La distancia de Mahalanobis

1. Programe la función *calcular_distancia_Mahalanobis(vector1, vector2)* la cual calcula la distancia de mahalanobis entre dos vectores:

$$d_M(\vec{x}_1, \vec{x}_2) = \sqrt{(\vec{x}_1 - \vec{x}_2)^T \Sigma^{-1} (\vec{x}_1 - \vec{x}_2)}$$

- a) Calcule y reporte en el documento las siguientes distancias de Mahalanobis $d_M(\vec{\mu}, \vec{y}_1)$ y $d_M(\vec{\mu}, \vec{y}_2)$. ¿Qué tan similares son según esta distancia? ¿Explica mejor la distancia de Mahalanobis la verosimilitud de que cada uno de los puntos haya sido generado por una distribución Gaussiana? ¿Porqué?

Parte II

Optimización

4. Optimización de funciones

Para la siguiente función:

1.

$$f_1(x_1, x_2) = (x_1 - 0,7)^2 + (x_2 - 0,5)^2$$

con $x_1, x_2 \in [-4, 4]$.

Realice lo siguiente:

1. **(10 puntos)** Según tales gráficas, grafique las funciones usando la función *contour* de *matplotlib* o similar, y distinga si las funciones son convexas o no, y los puntos mínimos y regiones o puntos silla.
2. **(30 puntos)** Implemente el algoritmo del **descenso del gradiente**, para cada función:
 - a) Escoja un coeficiente de aprendizaje α que permita la convergencia y reporte los resultados para 10 corridas:
 - 1) La cantidad de iteraciones necesarias para converger, y el punto de convergencia.
 - 2) Reporte si convergió al punto correcto.
 - 3) Escoja una de las corridas y en una gráfica muestre los puntos probados por el algoritmo, usando la función *contour* o similar, donde se muestren todos los puntos probados durante el proceso de optimización.
 - b) Escoja un α relativamente grande respecto al valor seleccionado ¿Qué sucede? ¿Permite un α muy grande la convergencia?
 - 1) ¿Qué sucede si escoge un α muy pequeño?
3. **(20 extra puntos)** Implemente el algoritmo de **Newton-Raphson**, para cada función:
 - a) Calcule la matriz Hessiana demostrando cada paso intermedio.
 - b) Reporte los resultados para 10 corridas:
 - 1) La cantidad de iteraciones necesarias para converger.
 - 2) El punto de convergencia y reporte si convergió al punto correcto.
 - 3) Escoja una de las corridas y en una gráfica muestre los puntos probados por el algoritmo, usando la función *contour* o similar, donde se muestren todos los puntos probados durante el proceso de optimización.

Para la siguiente función:

1.

$$f_2(x_1, x_2) = x_1 e^{(-x_1^2 - x_2^2)}$$

con $x_1, x_2 \in [-2, 2]$.

2. **(10 puntos extra)** Implemente el algoritmo de **Newton-Raphson**, para cada función:
- a) Calcule la matriz Hessiana demostrando cada paso intermedio.
 - b) Reporte los resultados para 10 corridas:
 - 1) La cantidad de iteraciones necesarias para converger.
 - 2) El punto de convergencia y reporte si convergió al punto correcto.
 - 3) Escoga una de las corridas y en una gráfica muestre los puntos probados por el algoritmo, usando una gráfica de las curvas de nivel para la función optimizada.