Ayudantía 4

1. Implemente en Python una función que, dada una matriz A, determine su norma de Frobenius.

2. El código mostrado a continuación, dada una matriz A, calcula su primer vector singular izquierdo y derecho (u_1, v_1), asi como su primer valor singular (σ_1):

```
import numpy.linalg as nl


def find_u_v_sigma(A):
    U, S, VH = nl.svd(A)
    return U[:,0], S[0], VH[0]
```

Use esta función para encontrar la matriz B de rango 1 que mejor aproxima las siguientes matrices, y calcule el error de aproximación (en norma de Frobenius):

- $A = \begin{pmatrix} 4 & 4 & 5 \\ 1 & 2 & 3 \\ 3 & 0 & -1 \\ 0 & 7 & -3 \end{pmatrix}$
- $A = \begin{pmatrix} 1 & 2 & 4 & 10 & -2 \\ -1 & -2 & 3 & -4 & 1 \end{pmatrix}$
- $A = \begin{pmatrix} -2 & 4 & 0 \\ 1 & 2 & 4 \\ 3 & 0 & 6 \end{pmatrix}$

Corrobore que en todos los casos anteriores se cumplen las siguientes igualdades

a) $\|A - B\|_F^2 = \|A\|_F^2 - \sigma_1^2$
b) $u_1 = \frac{1}{\sigma_1} A v_1$

3. El dataset `HVI.csv`[1] contiene información relacionada al índice de desarrollo humano de los diez países de latinoamérica. Usaremos los contenidos del curso para intentar predecir el ranking de los países.

[1]Disponible en Canvas.

a) Lea los datos, realice una exploración de estos, y elimine las columnas no numéricas.

b) Tratando cada fila como un vector, podemos interpretar los datos como varios puntos $a_1, \ldots, a_{10}$. Encuentre el vector v (de norma 1) que genere el espacio de dimensión 1 que mejor aproxima a estos puntos, es decir, tal que

$$\sum_{i=1}^{10} \|a_i - \mathrm{proj}_v(a_i)\|^2 = \sum_{i=1}^{10} \|a_i - \langle a_i, v \rangle v\|^2$$

sea lo menor posible. Ordene los puntos de acuerdo al valor de su proyección sobre v. ¿Que tal es la aproximación del ranking?

c) Ahora buscaremos el espacio afín de dimensión 1 que mejor aproxima los puntos. Para ello repetimos el procedimiento anterior, pero partiendo por centrar los datos. Para ello, calculamos el centro de los datos

$$c = \frac{1}{10} \sum_{i=1}^{10} a_i$$

y repetiremos los mismos pasos que en el punto anterior, pero con los puntos $\tilde{a}_i = a_i - c$. ¿Que tal es la aproximación ahora?