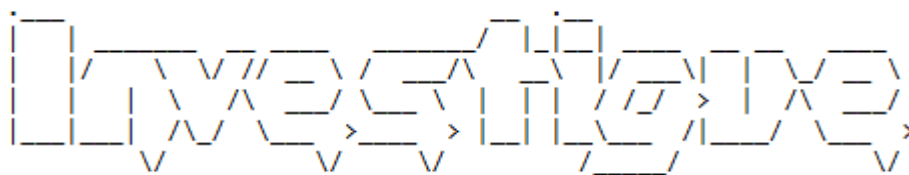


Proyecto #4 – INFO 1157

By Alberto Caro

1.- Dado el algoritmo **KNN** visto en clases, haga los cambios necesarios para clasificar el data test.

```
1 import math
2
3 dLabel = {0:'Iris-Setona',1:'Iris-Versicolor',2:'Iris-Virginica' }
4
5 def KNN(dData,aTest,K=3):
6     aD = []
7     for c in dData:
8         for f in dData[c]:
9             dE = math.sqrt('Calculo Distancia Euclidiana')
10            aD.append((dE,c))
11            aD = sorted(aD[:k])
12            ...
13            for d in aD:
14                ...
15                ... -> Completar
16                ...
17            return
18
19 Obs:
20 1.- Cargue en dD <-- data_training.txt
21 2.- Cargue en aT <-- data_test.txt
22 3.- Probar clasificación con K = 2,3,4 y 5.
23
24     for aTest in aT:
25         print(KNN(dD,aTest,nK))
26
```

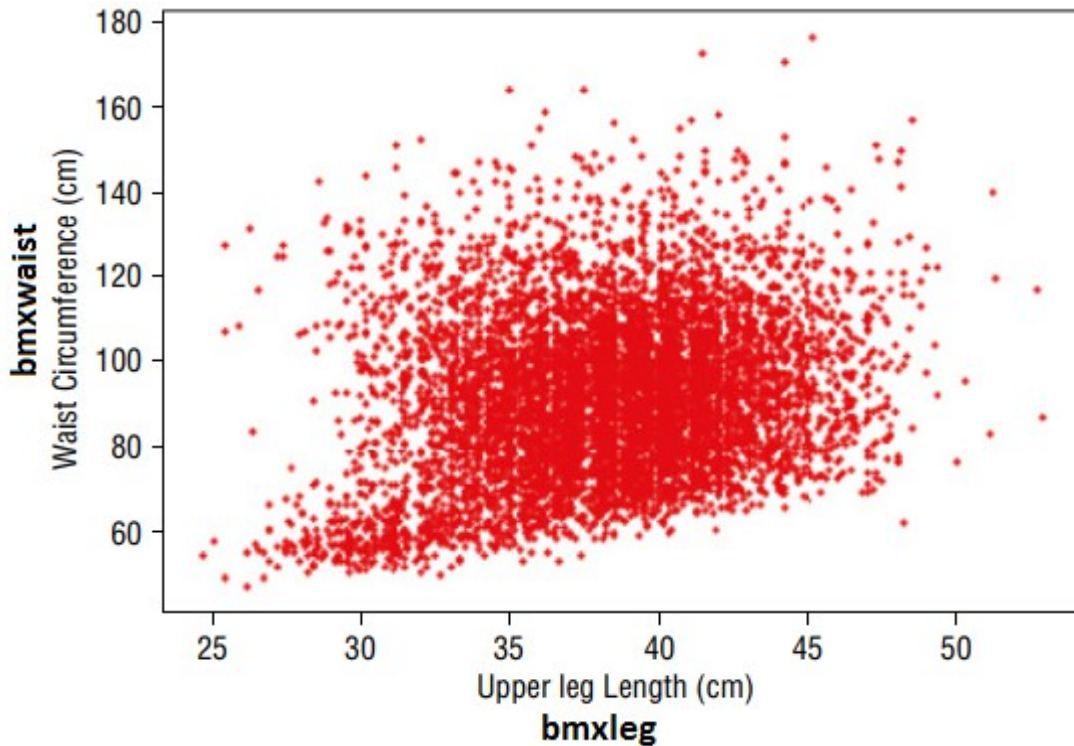


Obs: No se puede utilizar ninguna librería. Implemente **KNN** en **python**. Puede utilizar **numpy**.

2.- Implemente una **Métrica de Accuracy** en la clasificación de los **data test** y complete la **Tabla**.
Pruebe con **k = 2,3,4** y **5**. Explique claramente sus resultados.

Tabla Accuracy (%)				
Data Test	k = 2	3	4	5
Data Test				

2.- Utilizando **Kmeans** en un problema real. Suponga que usted es un diseñador de ropa (**Bermudas**) y tiene que decidir sobre el tamaño óptimo de los nuevos **Bermudas** (*Circunferencia de Cintura y Largo de Piernas*) para que lo puedan usar la mayoría de personas (*clientes potenciales*). ¿Cómo puede encontrar la correcta combinación de tamaño de estos **Bermudas**? Usted dispone de la siguiente base de datos ← **body.csv**, la cual posee **27 columnas y 9.338 filas**. Sin embargo a usted le sirven sólo **2** columnas: **BMXWAIST** (Circunferencia Cintura) y **BMXLEG** (Largo de Piernas), ambas en centímetros. Utilice la librería **pandas** para procesar los datos. Además, deberá limpiar los datos pues algunas filas de **BMXWAIST** y **BMXLEG** están vacías (**NaN**). Debe eliminar todas estas filas con operaciones de **pandas** (**isnull()**). Al aplicar **scatter** (matplotlib) sobre los datos limpios se obtiene la siguiente figura:



Obs: DataFrame Original → Shape → **(9.338,27)**. DataFrame Limpio → Shape → **(6.899,27)**.

Aplique **Kmean** de **sklearn.cluster**. Utilice **k = 2**. Responda,

1. ¿Cuáles son las dimensiones de los nuevos diseños de **Bermudas** a producir?
2. ¿Si la empresa quiere producir **4** nuevos diseños de **Bermudas**, cuáles serían sus dimensiones?
3. Fundamente sus respuestas. Grafique los **Cluster** obtenidos!

Fecha de entrega y defensa → **25 Noviembre Sección 1 y 26 Noviembre Sección 2**. Horario de clases y en oficina del profesor. Informe impreso y anillado.

