

part1

April 7, 2021

1 Machine Learning Project 1 (Part 1)

1.1 UNICAMP

1.2 Dataset 1

Load Datasets

```
[30]: import pandas as pd
import numpy as np

df = pd.read_table("cluster.dat",header=None, sep="\s+")
datos = pd.DataFrame(df)

datos
```

```
[30]:      0      1
0    1555.0  28.65
1    1490.0  27.55
2    1445.0  28.35
3    1415.0  28.80
4    1375.0  28.05
..      ...   ...
568  3420.0  27.95
569  3465.0  26.85
570  3525.0  26.00
571  3570.0  26.15
572  3440.0  25.60

[573 rows x 2 columns]
```

Transformamos el dataset a un array de nummpy

```
[31]: datos=datos.to_numpy()
datos
```

```
[31]: array([[1555. ,  28.65],
          [1490. ,  27.55],
          [1445. ,  28.35],
```

```
...,
[3525. , 26. ],
[3570. , 26.15],
[3440. , 25.6 ]])
```

Shuffle data

```
[32]: np.random.shuffle(datos)
      datos
```

```
[32]: array([[1160. , 25.9 ],
[3025. , 19.4 ],
[2070. , 5.15],
...,
[1425. , 8.25],
[3155. , 25.2 ],
[1635. , 5.35]])
```

Split data in train and test: - 90% train - 10% test

```
[33]: datos_train = datos[:int(0.9*len(datos))]
      datos_test = datos[int(0.9*len(datos)):]

      print("Train", datos_train.shape)
      print("Test", datos_test.shape)
```

Train (515, 2)

Test (58, 2)