**MACHINE LEARNING FROM DATA**

**Report: Lab Session 0 – Exploratory data analysis**

**Names:**

**Group:**

Instructions

* Answer the questions in a document **Lab0\_report\_team\_surnames.docx**
* Write the new code in a Colab Notebook **Mlearn\_lab0\_3\_team\_surnames.ipynb**.

Questions

Q1. Briefly describe the conclusions of your analysis (you can insert plots)

Q2. For each class and each feature, analyze histograms, cdfs and normal plots. Can we assume a Gaussian distribution for any of the features?

Q3. Analyze kurtosis and skewness values for each feature and class.

Q4. Analyze boxplots by feature. Are there ‘significant’ differences between the classes?

Q5. Analyze the scatter plot. Are features related in any way? What can you say about the separability of the classes?

Create a new Colab Notebook: **Mlearn\_lab0\_3\_team\_surnames.ipynb**.

Q6. Choose one feature (among the four available), write the code to compute the feature mean and confidence intervals at confidence levels 95%, 99% and 99.9% for the three classes.

Feature selected:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mean | CI at 95% | CI at 99% | CI at 99,9% |
| Class 1 |  |  |  |  |
| Class 2 |  |  |  |  |
| Class 3 |  |  |  |  |

Q7. Write the code to conduct the following hypothesis tests, using the Shapiro-Wilk test and the Anderson Darling test, for **all** the features K and classes J.

* Null hypothesis : Feature K from class J comes from a Gaussian distribution at the significance level **

For each test complete the corresponding table with the decisions (acceptance/rejection) for the null hypothesis H0 (feature Gaussianity), and the p-value or the critical and statistic values, respectively, for ** = 0,05 and ** = 0,01

Explain the meaning of the p-value / critical value and interpret the results accordingly.

Table for Shapiro-Wilk test

|  |  |  |
| --- | --- | --- |
| Feature # | Acceptance / rejection of *H*0 | *p*-value |
| class 1 |  |  |
| class 2 |  |  |
| class 3 |  |  |

Table for Anderson Darling test

|  |  |  |  |
| --- | --- | --- | --- |
| Feature # | Acceptance / rejection of *H*0 | critical value | stat |
| class 1 |  |  |  |
| class 2 |  |  |  |
| class 3 |  |  |  |