**MACHINE LEARNING FROM DATA**

**Fall 2018**

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

**Names:**

**Group:**

Instructions

* Answer the questions
* Save the report, convert to pdf and upload the pdf file

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?

Q6. Edit the script ML\_Lab0\_irisdataset.m. Choose one feature (among the four available) and compute the feature mean and confidence intervals at confidence levels 95%, 99% and 99.9% for the three classes.

Hint: use Matlab functions tinv and var

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

Q7. Copy the code used to answer Q6.

Q8. Choose one feature K (among the four available). Edit the script ML\_Lab0\_irisdataset.m to conduct the following hypothesis tests, using a chi-squared test

* Null hypothesis : Feature K from class 1 comes from a Gaussian distribution at the significance level 0.001
* Null hypothesis : Feature K from class 2 comes from a Gaussian distribution at the significance level 0.001
* Null hypothesis : Feature K from class 3 comes from a Gaussian distribution at the significance level 0.001

Complete the following table with the decisions (acceptance/rejection) for the null hypothesis H0 (feature Gaussianity), p-value and degrees of freedom for ** = 0,001.

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

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

Q9. Copy the code used to answer Q8.