

Data Science Small Test

This is a small test to check the basic understand of machine learning developed by HIAE.
Please do not spend too much time in it.

Instructions

- 1 – Choose the a language of interest (R / Python / Matlab)
- 2 – Read the dataset, it is tab separated, contains header and the first column is the index.
- 3 – The results can be presented in any format. We suggest a iPython notebook for python users and RMarkdown for R users.
- 4 – Please comment the main steps of your code.
- 5 – Imagine that the results will be presented for a non-technical audience so try to emphasize the results of your work.

For R:

```
df_points <- read.delim("df_points.txt", row.names=1, quote="", stringsAsFactors=FALSE)
```

For Python:

Import pandas as pd

```
dfPoints = pd.read_csv("df_points.txt", delimiter="\t")
```

Problem

- 1 - Build a **classification** problem, using the columns x, y and z, trying to classify the label column.
 - a) Segregate a test and training frame.
 - b) Use a GLM or Logistic Regression model and show the results.
 - c) Use other method of your choice to handle the problem
 - d) Compare and comment the results on the models used from b) and c)