**README OF MULTIPLE LINEAR CLASSIFICATION EQUATION SYSTEM DATABASES THAT WERE GENERATED**

made by: César Miranda Meza

email: cmirandameza3@hotmail.com

The multiple linear classification equation that was used as a reference to create the databases labeled as “the multiple linear classification equation systems” is the following:

Where is the dependent variable (output of the current sample); represents the independent variables (inputs of the current sample); and stand for the coefficient values of the equation. Furthermore, the values that were selected for are the following:

such that the Eq. will turn into the following:

However, the Eq. was modified by adding to it a bias component , that would represent a random value and should be generated each time a new sample is calculated:

Where the independent variable was restricted to be sampled with values according to the following way .

Nevertheless, several regression databases governed by the term from the Eq. have already been created (see databases in the directory databases/regressionDBs/multipleLinearEquationSystem). Therefore, it was decided to recycle them and use a copy of those files in Excel to apply to them the threshold defined in the Eq. , which is . As a consequence, the following .csv (comma delimited) files were generated for the creation of the multiple linear classification equation systems:

* linearClassificationSystem\_1systems\_10samplesPerSys.csv
* linearClassificationSystem\_10systems\_10samplesPerSys.csv
* linearClassificationSystem\_10systems\_100samplesPerSys.csv
* linearClassificationSystem\_100systems\_100samplesPerSys.csv
* linearClassificationSystem\_100systems\_1000samplesPerSys.csv
* linearClassificationSystem\_1000systems\_1000samplesPerSys.csv

For all these files, note that they try to mimic how a real database would normally be organized by a professional and in which you will encounter four columns, whose headers and purpose are the following:

1. **id:** Represents the unique identifier for the current row of the database.
2. **system\_id:** Represents the unique identifier for the current system sampled. This is because the databases will contemplate having several samples for several systems that manifest the same phenomenon.
3. **dependent\_variable:** Represents the output value of the current sample.
4. **independent\_variable\_1:** Represents the input value 1 that generated the current sample.
5. **independent\_variable\_2:** Represents the input value 2 that generated the current sample.

**Created in:** September 21, 2021.

**Last update in:** None.