README OF LOGISTIC EQUATION SYSTEM DATABASES THAT WERE GENERATED

made by: César Miranda Meza

email: cmirandameza3@hotmail.com

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

$$y = \frac{1}{1 + e^{-(b_0 + b_1 x)}} \quad | \, 0 < y < 1 \tag{1}$$

Where y is the dependent variable (output of the current sample); x represents the independent variable (input of the current sample); and b_0, b_1 stand for the coefficient values of the equation. Furthermore, the values that were selected for b_0, b_1 are the following:

- $b_0 = -8.26$
- $b_1 = 0.165$

such that the Eq. (1) will turn into the following:

$$y = \frac{1}{1 + e^{-(-8.26 + 0.165x)}} = \frac{1}{1 + e^{(8.26 - 0.165x)}} \tag{2}$$

However, the Eq. (2) was modified by adding to it a bias component r, that would represent a random value and should be generated each time a new sample is calculated:

$$y = \frac{1}{1 + e^{(8.26 - 0.165x)}} + r \quad | \, -0.1 \leq r \leq 0.1 \tag{3}$$

Where the independent variable was restricted to be sampled with values according to the following way $0 < x \leq 100$ and where if no random bias value is needed, then it should be negated by setting $r = 0$ or, Ec. (2) should be used instead. In addition, the Eq. (3) will be restricted so that the output values generated in it, through y, do not exceed the value of 1 and so that it is never negative, as explained in the Pseudocode 1:

Algorithm 1: Restriction to the Eq. (3).

---

1. **If** $(y > 1)$ **then**
2. $\quad y = 1$
3. **end if**
4. **If** $(y < 0)$ **then**

5. $y = 0$
6. **end if**

With the help of the Excel “text to columns” function, for the creation of the logistic equation system databases, the Eq. (3) was employed to generate each of the samples contained in the following .csv (comma delimited) files:

- randLogisticEquationSystem/1systems_10samplesPerSys.csv
- randLogisticEquationSystem/10systems_10samplesPerSys.csv
- randLogisticEquationSystem/10systems_100samplesPerSys.csv
- randLogisticEquationSystem/100systems_100samplesPerSys.csv
- randLogisticEquationSystem/100systems_1000samplesPerSys.csv
- randLogisticEquationSystem/1000systems_1000samplesPerSys.csv

And for the ones made from the Eq. (2):

- logisticEquationSystem/1systems_10samplesPerSys.csv
- logisticEquationSystem/10systems_10samplesPerSys.csv
- logisticEquationSystem/10systems_100samplesPerSys.csv
- logisticEquationSystem/100systems_100samplesPerSys.csv
- logisticEquationSystem/100systems_1000samplesPerSys.csv
- logisticEquationSystem/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 that generated the current sample.

Moreover, the samples generated aimed to attempt mimicking how several real life systems behave in real life due to the bias component r. On the other hand, each listed database was generated through a separated file which was developed in Python programming language (v3.7.1) in order to display a friendly and simple code:

- 1systems_10samplesPerSys.py
- 10systems_10samplesPerSys.py
- 10systems_100samplesPerSys.py
- 100systems_100samplesPerSys.py
- 100systems_1000samplesPerSys.py

- 1000systems_1000samplesPerSys.py

Created in: September 21, 2021.

Last update in: June 24, 2022.