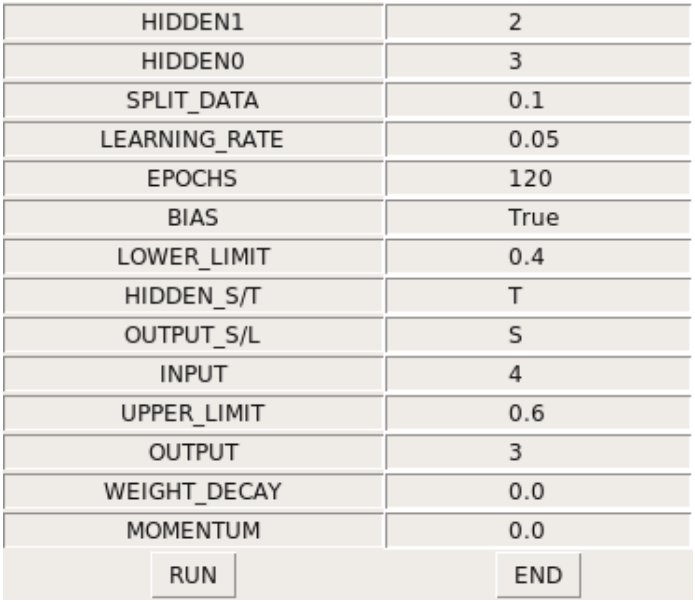
Some Further Notes for Guidance in your Execution of Assignment 2

1. You were provided with the PyBrain Simulator (Assignment2012Part2R1.py). This should run on remote Linux, but you could run it on your computer at home. You were also provided with three basic data files, Iris, Mushroom, Steel Faults and Blood. These were obtained from the machine learning database, http://archive.ics.uci.edu/ml/
2. Today (11th May 2012), a file Assignment2Data.zip has been place in the resource area of CECIL. This contains a number of data files which have been normalised in various ways (see the README file).
3. Consider again, the GUI



Unfortunately, when the new options were added, they separated LOWER_LIMIT from UPPER_LIMIT and INPUT from OUTPUT. This was a default action by the TkInter graphics package. It could be fixed, but I don't want any more revisions unless it is absolutely necessary.

Important Please be aware that it is very important that the numbers in the INPUT and OUTPUT fields represent the data set that you are working with. The values shown above pertain to the Iris Dataset. If you are using the Faults data, then INPUT must be 27 and OUTPUT must be 7. Similarly, for the Blood Dataset, INPUT must be 4 and OUTPUT must be 2. Otherwise, the simulator will not run properly (and my diagnostics are pretty minimal).

4. One of your main objectives is to find the highest percentage correct for the Blood data test set. This doesn't necessarily get you the highest mark. You should experiment with the parameters described in earlier (Assignment2012Part2Extra1.pdf – in earlier zip file) and also look at any effects that the choice of normalisation and squashing function may have.