NAME

regress – Fit straight lines to (x,y) data using L2 and L1 regression.

SYNOPSIS

```
cat data | ${HOME}/bin/exe/regress [plot]
```

DESCRIPTION

The program reads and counts up to 1000 (x,y) data pairs from standard-in. Then it uses library routines L2REGW and L1REGW (with weights of 1) to compute the least-squares and least-absolute-values regressions on the data, and writes the equations to standard-out.

OPTIONS

If the option is supplied, coordinates on the two regression lines are written to files /tmp/l2data and /tmp/l1data. Either or both lines can then be plotted by using ${HOME}/bin/plot, or they can be plotted with the data points by using gnuplot commands.

UNITS and FILES

0	error and informational messages
5	input coordinate pairs
6	output equations of best-fit lines
9	plotting coordinates in /tmp/l2data and /tmp/l1data

DIAGNOSTICS

These are the possible return codes.

0	all went well
1	too much data
2	bad data
3	too little data
4	L2 regression failed
5	L1 regression failed

PERFORMANCE

If there are N data pairs, the linear programming tableau for the L1 regression has N+1 rows and 2N+4 columns. The time required to complete a pivot is proportional to the size of the tableau, and the number of pivots required is proportional to the number of rows in the tableau, so the running time used by the simplex method grows as (N+1)(2N+4)(N+1) or approximately as the cube of the number of data pairs. For 500 data pairs regress uses about 22 seconds of CPU time on a machine clocking 1 GHz.

EXAMPLE

```
unix[1] cat data
0 1
1 3
2 5
3 7
4 9
5 27
unix[2] cat data | regress
L2: y=[-2.047619E+00] + [ 4.285714E+00]x
L1: y=[ 1.000000E+00] + [ 2.000000E+00]x
unix[3]
```

The coordinate pairs fall on the line y=2x+1, except for the last data point. Least-absolute-values regression ignores this outlier, while least-squares yields a regression line that responds strongly to it.