

Michael Pollack: LING 572 Hw2

Due: 11pm on Jan 24, 2025

Q1 (60 points):

Run **build_dt.sh** with **train.vectors.txt** as the training data and **test.vectors.txt** as the test data:

- Fill out Table 1 (where min_gain is set to 0) and Table 2 (where min_gain is set to 0.1).
- submit model_file, sys_output, acc_file produced by running
build_dt.sh train.vectors.txt test.vectors.txt 4 0.1 model_file sys_output > acc_file

Table 1: Your decision tree results when min_gain=0

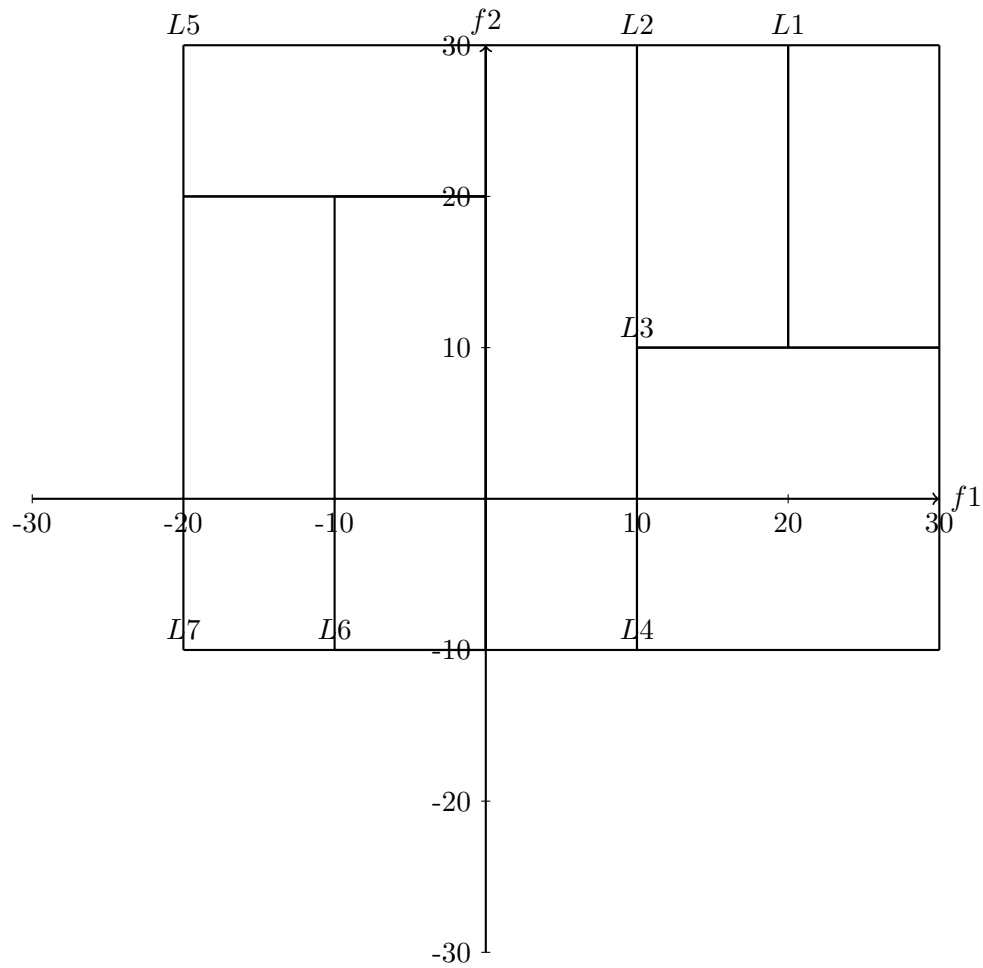
Depth	Training accuracy	Test accuracy	CPU time (in minutes)
1	0.45296296296296296	0.4166666666666667	0.10634603333333331
2	0.5207407407407407	0.5266666666666666	0.17833113333333333
4	0.6377777777777778	0.5233333333333333	0.49956506666666666
10	0.7514814814814815	0.6	2.3777599333333333
20	0.8555555555555555	0.6833333333333333	5.2909486833333334
50	0.9681481481481482	0.7	10.313614333333334

Table 2: Your decision tree results when min_gain=0.1

Depth	Training accuracy	Test accuracy	CPU time (in minutes)
1	0.45296296296296296	0.4166666666666667	0.10800393333333334
2	0.52	0.53	0.18130926666666666
4	0.6014814814814815	0.54	0.31964771666666664
10	0.6014814814814815	0.54	0.31633571666666666
20	0.6014814814814815	0.54	0.31752396666666666
50	0.6014814814814815	0.54	0.32321163333333333

Note: I am not sure why my performance is plateauing after a max depth of 4. I re-ran these several times and got the same results. This does not appear to be happening with a min gain of 0.

Q2 (10 points): Slide #10 of class2_DT.pdf shows a DT: f1 and f2 are two features; f1 is in [-20, 30]; f2 is in [-10, 30]. L_i (i=1, ..., 7) represents a leaf node. Each leaf node corresponds to a rectangle in a 2-dimension space, where f1 is the x-axis and f2 is the y-axis. Draw a graph that shows the boundary of the seven rectangles in this 2-dimension space.



Q3 (5 “free” points): If you are not familiar with Patas or Condor submit, please go over the condor information at <https://wiki.ling.washington.edu/bin/view.cgi/Main/HowToUseCondor> and <https://wiki.ling.washington.edu/bin/view.cgi/Main/CondorClusterHomepage>. You can run condor submit for the code in Q1 and other assignments.