Ling 572 HW2

Michael Roylance, Olga Whelan

Q1 (a) mallet import-svmlight --input examples/train.vectors.txt --output train.vectors
mallet import-file --input examples/test.vectors.txt --output test.vectors --use-pipe-from train.vectors
vectors2classify --training-file train.vectors --testing-file test.vectors --trainer DecisionTree > dt.stdout 2>dt.stderr

(b) train accuracy mean = 0.6377777777777778
test accuracy mean = 0.5233333333333333

Q2 It looks like Mallet DT learner treats the features as binary. After running the commands below, we get the same accuracy results as in Q1.

./binarize.sh train.vectors.txt train.vectors.bin.txt
./binarize.sh test.vectors.txt test.vectors.bin.txt
mallet import-svmlight --input train.vectors.bin.txt --output train.vectors.bin
mallet import-svmlight --input test.vectors.bin.txt --output test.vectors.bin --use-pipe-from train.vectors.bin
vectors2classify --training-file train.vectors.bin --testing-file test.vectors.bin --trainer DecisionTree > dt.stdout.bin 2>dt.stderr.bin
diff dt.stdout dt.stdout.bin