LING 572 Hw5 (MaxEnt decoder)
Due: 11:45pm on Feb 11, 2010

The example files are under dropbox/09-10/572/hw5/examples/.

Q1 (10 points): Run the Mallet MaxEnt learner (i.e., the trainer's name is MaxEnt) with train2.vectors.txt as the training data and test2.vectors.txt as the test data.

- Run the *vectors2train* command with the –model option and save the MaxEnt model to a file called **mallet_model**.
- Run the *classify* command with the saved model on the test data
- Make sure that running the two commands above yields the same result as running *vectors2classify*.
- Convert the model into the text format with the following command: classifier2info –classifer mallet_model_q1 > mallet_model_q1.txt
- In your note file, write down the training accuracy and the test accuracy.

Q2 (40 points): Write a MaxEnt classifier, called **maxent_classify.sh**, that classifies test data given a MaxEnt model learned from training data.

- The format is: maxent_classify.sh test_data model_file sys_output > acc_file
- test_data, sys_output, and acc_file have the same format as in Hw2-Hw4.
- model_file has the same format as mallet_model.txt created in Q1.
- Run "maxent_classify.sh test2.vectors.txt mallet_model_q1.txt output_q2 > acc_q2". What is the test accuracy? Is it the same as the test accuracy in Q1? Why or why not?

Q3 (20 points): Calculating empiricial expectation.

- The format is: calc_emp_exp.sh training_data output_file
- training_data has the same format as before.
- output_file has the format "class_label feat_name expectation"
- Run "calc_emp_exp.sh train2.vectors.txt output_q3" and include output_q3 in your submission.

Q4 (30 points): Calculating model expectation.

- The format is: calc_model_exp.sh training_data output_file {model_file}

- training_data and output_file have the same format as in Q3.
- model_file is optional. If it is given, it has the format as in Q2 and it is used to calculate $p(y|x_i)$. If it is not given, $p(y|x_i) = 1/|C|$, where $|C|$ is the number of class labels.
- Run "calc_model_exp.sh train2.vectors.txt output_q4 mallet_model_q1" and include output_q4 in your submission.

Submission: Submit a tar file via CollectIt. The tar file should include the following.

- If your team has two people, please submit only one copy. In your note file, please list the names of team members.
- In your note file hw5.*, include your answers to Q1-Q4, and any notes that you want the TA to read.
- Shell scripts for Q2-Q4, and related source and binary code.
- The model and output files created in Q1-Q4. (e.g., output_q2, output_q3, and output_q4).