Mehryar Mohri
Speech Recognition
Courant Institute of Mathematical Sciences
Homework assignment 2
Due: November 30, 2007

A. N-gram Language Models

For several of the questions of this section, it is recommended that you use the GRM library.

1. Download the following training corpus S and test corpus $\hat{S}$:
   `http://www.cs.nyu.edu/~eugenew/asr/train-norm.txt`
   `http://www.cs.nyu.edu/~eugenew/asr/test-norm.txt`.

2. Extract the vocabulary Σ_1 of S and define a start and end symbol.

3. Create a bigram back-off language model automaton A for S. What is the number of states and transitions of the resulting weighted automaton? How many bigrams were found in S? How many ϵ-transitions does that automaton have?

4. Same questions with a trigram back-off language model automaton B (with number of trigrams instead of bigrams).

5. Use $\hat{S}$ to estimate the perplexity of A and B using the test corpus $\hat{S}$ and compare them.

6. In this question, we wish to create a simple class-based model where each class is either reduced to a single word, or a single bigram. We can use the average mutual information of $\Pr[w_1 w_2]$ and $\Pr[w_1]\Pr[w_2]$ to determine classes. Thus, a bigram $w_1 w_2$ forms a class of his own when
$$I(w_1, w_2) = \log \frac{\Pr[w_1\, w_2]}{\Pr[w_1]\Pr[w_2]} \tag{1}$$
is positive and relatively large.

   (a) List the ten bigrams $w_1 w_2$ in S with the largest $I(w_1, w_2)$.

   (b) Define the 1,000 bigrams with the largest $I(w_1, w_2)$ as bigram classes and create a class-based bigram back-off model (describe how the class-based model is defined using transducers and mention which FSM or OpenFst library commands were used).

(c) Compute the perplexity of your class-based model on $\hat{S}$ and compare it with the previous perplexity measures obtained.

C. Maxent Models

The maximum entropy principle consists of selecting a distribution that is as close as possible to the uniform distribution, with the expected feature values matching the empirical ones. In this section, we examine alternatives to the standard maxent model where closeness is measured by a Bregman divergence other than the relative entropy.

1. Assume that the distributions are defined over a finite set X, $|X| = n < +\infty$. Show that the following

$$B(p||q) = (p-q)^T \mathbf{K}(p-q), \tag{2}$$

where $\mathbf{K} \in \mathbb{R}^{n \times n}$ is a positive definite symmetric matrix, defines a Bregman divergence.

2. Use the Lagrangian to determine the form of the solution to the maxent problem with this Bregman divergence.

3. Give the statement of the duality theorem for this divergence.