11-751: Speech recognition and understanding, Fall 2008

Assignment 2

Teaching assistant: Wilson Tam (yct@cs.cmu.edu) *Due on: Oct 15 before class*

Problem 1: Pattern recognition

Consider the following decision rule for a two-category $\{\omega_1, \omega_2\}$ one-dimensional problem on feature x: Decide ω_1 if x $> \theta$; otherwise decide ω_2 where θ is a threshold.

1. Show that the probability of error for this rule is given by:

$$P(error) = P(\omega_1) \int_{-\infty}^{\theta} p(x|\omega_1)dx + P(\omega_2) \int_{\theta}^{\infty} p(x|\omega_2)dx \quad (1)$$

2. By differentiating, show that a necessary condition to minimize P(error) is that θ satisfies

$$p(\theta|\omega_1)P(\omega_1) = p(\theta|\omega_2)P(\omega_2) \quad (2)$$

3. Does Eqn 2 define θ uniquely?
4. Give an example where a value of θ satisfying the equation actually *maximizes* the probability of error.
5. So, what is the correct decision rule which minimizes the probability of error? Explain briefly.

Problem 2: Hidden Markov Model

Assume Jessie lives on a small island. The weather is very simple on the island. In some days it is sunny and in some days it is rainy. But the weather is invisible to you. All you know is the information whether she brings umbrella with her, whether she wears skirt and whether she wears cap (assuming she only has one of these 3 items per day). We start on day 1 in sunny state and there is one transition per day.

Define Q_t as the state on day t. Therefore the initial state probability is $P(Q_1 = sunny) = 1$ and $P(Q_1 = rainy) = 0$.

The emission probability for sunny state is $P(O = skirt|sunny) = 0.6$, $P(O = umbrella|sunny) = 0.25$, and $P(O = cap|sunny) = 0.15$. The emission probability for rainy state is $P(O = skirt|rainy) = 0.25$, $P(O = umbrella|rainy) = 0.6$, and $P(O = cap|rainy) = 0.15$.

The transition probabilities are $P(sunny|sunny) = 0.8$, $P(rainy|sunny) = 0.2$, $P(sunny|rainy) = 0.4$ and $P(rainy|rainy) = 0.6$.

Denote O_t as Jessie's appearance on day t. For example, since the initial state is sunny, we already know that $P(O_1 = umbrella) = 0.25$.

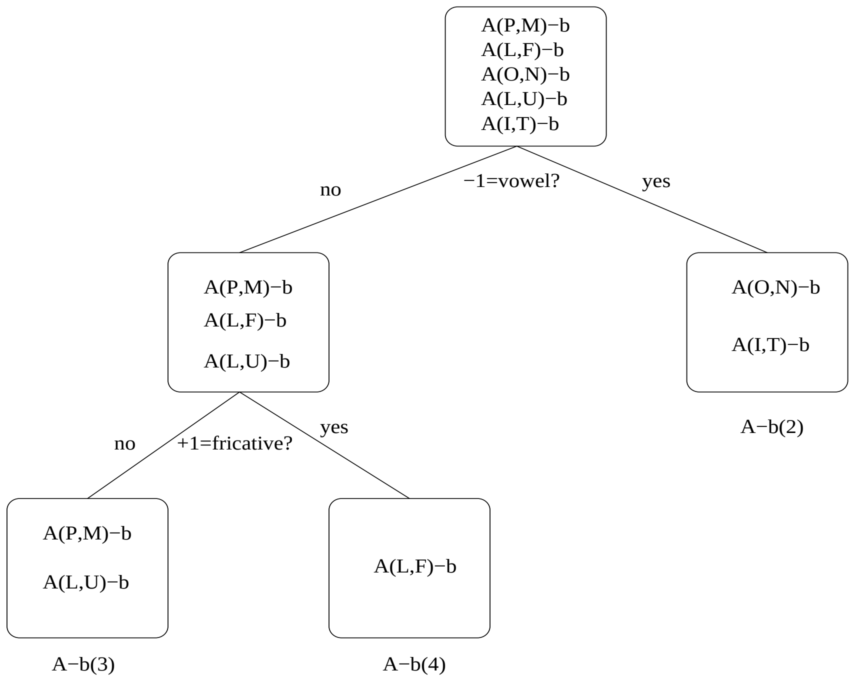

Figure 1: Phonetic tree for clustering context dependent phones. A(P,M) denotes that the left context (-1) of the current phone A is P and the right context (+1) is M.

1. Draw the hidden Markov model in a state diagram. Write the emission probabilities inside a state.
2. What is $P(Q_2 = rainy)$?
3. What is $P(O_2 = skirt)$?
4. What is $P(Q_2 = rainy|O_2 = skirt)$?
5. Denote Y(t) = $P(Q_t = Sunny)$. Note that Y(1) = 1. Then Y(t+1) can be defined inductively from Y(t) by an expression Y(t+1) = A + B·Y(t). What is A and B?
6. Assume you observe $\{O_1 = umbrella, O_2 = umbrella, O_3 = umbrella, O_4 = umbrella, O_5 = umbrella\}$, what is the most likely state sequence $\{Q_1Q_2Q_3Q_4Q_5\}$?

Problem 3: Acoustic modeling

1. What are the differences between discrete, continuous and semi-continuous HMMs? Compare the advantage and disadvantage of discrete HMM and continuous HMM.
2. Suppose the monophone representation of "CARNER" is /kh ab r9 n ETr/. Write down the left biphone, right biphone and right triphone representations of this word.
3. Given a context decision tree in Figure 1, which model on the leaf node does the context A(P,D) fall into?
4. Suppose we want to do context clustering for the phone /t/. We already have six discrete models for that phone. They are:

T1:[1/4,3/4] with 16 training samples
T2:[1/8,7/8] with 8 training samples
T3:[3/8,5/8] with 32 training samples
T4:[5/8,3/8] with 8 training samples
T5:[7/8,1/8] with 16 training samples
T6:[3/4,1/4] with 4 training samples

where all models share the same Gaussians but with different distribution weights $[p_1, 1-p_1]$ over 2 Gaussians.

Use the definition of Entropy of a discrete distribution:

$$H(p) = \sum_j p_j \log \frac{1}{p_j} \tag{3}$$

Write a program to perform the bottom-up clustering of the 6 models above into 3 clusters using the **weighted** discrete entropy distance. Attach the log file of your program as answers. [Hint: You need to consider the size of the training samples when you combine the distributions of two models.]