Week 7 Homework:
Due November 19, 2004

November 12, 2004

Problem 1: Viterbi decoding

Part A

Write a program to compute the most likely state sequence based on a sequence of observations and a model. You may use your programming language of choice, though ones with hash tables and arrays will make things easier.

You should hand in a printout of your commented code, as well as the execution results requested below.

You should implement the Hidden Markov Model for the ice cream data described in class. The observation sequence may be found in

`http://people.cs.uchicago.edu/~levow/bigideas/hmm_hw.data`

Compute the most likely state sequence and its probability for two conditions:

1. The original parameter settings given in Table 1 below.
2. The final parameter setting - trained by Baum-Welch - given in Table 2 below.

	$P(..\|C)$	$P(..\|H)$	$P(..\|Start)$
$P(1\|..)$	0.7	0.1	
$P(2\|..)$	0.2	0.2	
$P(3\|..)$	0.1	0.7	
$P(C\|..)$	0.8	0.1	0.5
$P(H\|..)$	0.1	0.8	0.5
$P(Stop\|..)$	0.1	0.1	0

Table 1: Original model parameters

	$P(..\|C)$	$P(..\|H)$	$P(..\|Start)$
$P(1\|..)$	0.614	7.1e-5	
$P(2\|..)$	0.148	0.534	
$P(3\|..)$	0.211	0.466	
$P(C\|..)$	0.934	0.072	5.1e-15
$P(H\|..)$	0.066	0.865	1
$P(Stop\|..)$	1e-15	0.0.063	0

Table 2: Trained model parameters

Part B

Evaluation of many tagging tasks that are performed using Hidden Markov models employs either an overall accuracy or precision recall measures that depend on matching the output state sequences with some small hand-labeled ground truth set of observation sequences. Explain why this may be problematic for Viterbi decoding.

Construct an example of an HMM and observation such that state j of the output of the Viterbi algorithm is not the individually most likely state.

Problem 2: Baum-Welch Reestimation

Part A

You may have noted in the ice cream example in class that the transition probability from START to STOP never changed from 0. Is there some observation sequence that could change the transition probability using the Baum-Welch algorithm? If so, give an example. If not, provide a proof.

Part B

Consider an analogous case where the original observation probability $P(1|H) = 0$. Is there some sequence of observations that could change this observation probability through the Baum-Welch algorithm? If so, give an example. If not, provide a proof.

Part C

Consider an original transition probability estimate as in Table 3. Describe and explain the effect of Baum-Welch reestimation on this initial condition.

Problem 3: Forward/Backward Computation

Compute $P(O|\lambda)$ in terms of β (without α). Prove that the formulation is equivalent to $\sum_{i=1}^{N} \alpha_i(T)$.

	$P(..\|C)$	$P(..\|H)$	$P(..\|Start)$
$P(1\|..)$	0.3	0.3	
$P(2\|..)$	0.4	0.4	
$P(3\|..)$	0.3	0.3	
$P(C\|..)$	0.45	0.45	0.5
$P(H\|..)$	0.45	0.45	0.5
$P(Stop\|..)$	0.1	0.1	0

Table 3: Revised initial model parameters