6.824 Artificial Intelligence, Fall 1994
Problem Set 4

Due: Beginning of class Wednesday, Oct. 12

Problem 1. Consider a Markov model with two internal states s_0 and s_1. We assume that whatever the internal state (whether it is s_0 or s_1) the transition probability matrix of the Markov model is such that the probability that the next state is different from the current state is P_c (the probability of change). We assume that $P_c < \frac{1}{2}$, i.e., the system is more likely to stay in the same state than to change states. Note that the probability of remaining in the same state is $1 - P_c$. Now suppose this Markov model is part of a Hidden Markov process with two possible output tokens, 0 and 1. Let P_a (for probability of agreement) be the probability that the output symbol is 1 when the internal state is s_1 and also the probability that the output symbol is 0 when the internal state is s_0. We assume that $P_a > \frac{1}{2}$. Intuitively, the internal state s_1 favors the output symbol 1 and the internal state s_0 favors the output symbol 0. Assume that the internal states s_0 and s_1 are equally likely in the initial state and the first output token corresponds to the initial state. Now Consider a bit string O that is an observed sequence of output 1s and 0s. Suppose that O is n 0s followed by k 1s. We wish to find the sequence I of internal states which maximizes $P_r(I|O)$. Depending on the value of n, k, P_c and P_a there are three possible correct answers. Give three values for S such that for some values of n, k, P_c and P_a (subject to the constraint $P_a > \frac{1}{2}$ and $P_c < \frac{1}{2}$) the expression $P_r(I|O)$ is maximized. For each answer give conditions on n, k, P_c and P_a which guarantee that $P_r(I|O)$ is maximized for that answer. Your conditions should cover all possibilties, i.e., any values of n, m, P_a and P_c with $P_c < \frac{1}{2}$ and $P_a > \frac{1}{2}$ should satisfy the conditions of one of your answers.

Problem 2. Consider an HMM λ with four internal state a_1, a_2, b_1 and b_2 and two external symbols a and b. We assume that if the internal state is either a_1 or a_2 then the external symbol is always a and if the internal state is either b_1 or b_2 then the external symbol is always b. We also assume that the state transition probability matrix is determined by four numbers A_1, A_2, B_1 and B_2. A_1 is the probability that when the machine is in state a_1 then the next machine state remains a_1. A_2, B_1 and B_2 are the analogous "no change" probabilities for states a_2, b_1 and b_2 respectively. If the machine does change state it must change from an "a" state to a "b" state or vice versa, e.g., there is no chance of a transition from a_1 to a_2. When the machine changes state the two possible successor states are equally likely. Consider the HMM defined above. Let O be a sequence of k_1 a's followed by k_2 b's followed by k_3 a's and then k_4 b's. Give a sequence I of $k_1 + k_2 + k_3 + k_4$ internal states and values for A_1, A_2, B_1, and B_2 such that $P_r(O|\lambda \wedge I)$ is maximized.

Problem 3. The model you constructed in problem 1 generates the external symbols deterministically — only one external symbol is possible for a given internal state. We might call an HMM with this property noise-free — the external symbols are generated in a noise-free way from the internal states. Consider an arbitrary noise-free HMM λ and an arbitrary sequence O of training data. Suppose that we use Baum-Welch reestimation to compute a new HMM $\overline{\lambda}$ such that $P_r(O|\overline{\lambda}) \geq P_r(O|\lambda)$.

Part a. Will $\overline{\lambda}$ be noise-free? Explain your answer.

Part b. Suppose that the HMM λ which maximizes $P_r(O|\lambda)$ is not noise-free but that we start with a noise-free HMM and run Baum-Welch reestimation. Will Baum-Welch restimation converge on an HMM which finds the global maximum (the maximum over all HMMs λ) of $P_r(O|\lambda)$? Explain your answer.

Problem 4. Consider the following stochastic grammar.

$$S \xrightarrow{\frac{1}{2}} AB$$

$$S \xrightarrow{\frac{1}{2}} aSb$$

$$A \xrightarrow{\frac{1}{2}} aA$$

$$A \xrightarrow{\frac{1}{2}} a$$

$$B \xrightarrow{\frac{1}{2}} bB$$

$$B \xrightarrow{\frac{1}{2}} b$$

This grammar can only generate strings of the form $a^n b^k$ (n as followed by m bs). Describe the parse P of $a^n b^k$ which maximizes $P_r(P|a^n b^k)$ and give the probability $P_r(P|a^n b^k)$ for this parse as a function of n and k.