CS628 DUE BY 11/23/04

Assignment #3

---

1. Provide a tight upper bound for the size of the accepting automaton for a string that contains a unique occurrence of a seed of weight w and span $s \geq w$.

2. Show that assuming the occurences of the seed in the mismatch string are independent yields an upper bound on the sensitivity.

3. Show that in the E-step of the Baum-Welch algorithm, the conditional expectation of the emissions count, $E_k(b)$, is given by

$$E_k(b) = \sum_j \frac{1}{p(\boldsymbol{x}^j)} \sum_{i:x_i^j=b} f_k^j(i) b_k^j(i).$$

For the next two problems you will need to download the data file linked from the course web-page. The file was generated by a two-states HMM corresponding to a fair/loaded coin flips. Your assignment should include your written code (Matlab is fine) and its output.

4. Using the Baum-Welch algorithm find the parameters of the HMM. Do that using the first 100, 500 and 1000 symbols of the sequence. Report how many flips at each step have an expected value which is in $[0.4, 0.6]$.

5. Do the same using the Viterbi training. At each step report how many flips have changed their assignment from the previous step.