CS114 (Spring 2015) Homework 4
Hidden Markov Model and POS Tagging

Due March 13, 2015

1 Hidden Markov Model

We use Hidden Markov Model to model the ice-cream sales and the weather that day. Our data are the ice-cream sales, and we are trying to infer the weather.

The HMM is fully specified as follows:

- Hidden states (weather) = X = {Cold, Hot}
- Observation (ice-cream sales) = O = {Good, Mediocre, Bad}
- The initial probability $\pi_C = 0.5$ and $\pi_H = 0.5$
- The transition probability matrix A

	Cold	Hot
Cold	0.6	0.4
Hot	0.5	0.5

  For example, $a_{CH} = P(X_{t+1} = H | X_t = C)$.

- The emission probability (or observation state probability) matrix B

	Good	Mediocre	Bad
Cold	0.2	0.3	0.5
Hot	0.6	0.3	0.1

  For example, $b_{CG} = P(O_t = G | X_t = C)$.

If you choose to do the computation by hand, you need to show all calculation steps. (Warning: it is quite tedious to do this by hand and calculator, but it is not absurd.) If you choose to write Python code to do the computation and print out the answers, turn in the code along with the answers. Given the HMM above, answer the following questions.

1. Given that the observation sequence is GMBG, enumerate all of the possible hidden state sequences and compute $P(X_1, X_2, X_3, X_4, O_1 = G, O_2 = M, O_3 = B, O_4 = G)$ (i.e. the score) for EACH of them.

2. In Forward algorithm, what does $\alpha_i(t)$ represent?

3. Run Forward algorithm to compute the sum of the scores of all possible hidden state sequences, which is equal to $P(O_1 = G, O_2 = M, O_3 = B, O_4 = G)$. Compute all of the relevant α values. (The answer should be equal to the sum of the answers in 1.)

4. In Viterbi algorithm, what does $\delta_i(t)$ represent?

5. Run Viterbi algorithm to compute the score of the best hidden state sequence. (The answer should be equal to the max of the answers in 1.)

2 POS tagging

The difficulties of POS tagging in English come from out-of-vocabulary words and words that can have multiple POS tags. Use Penn tagset to tag the following sentences. For each out-of-vocabulary word (e.g. repowentate, Vinken), describe some of the features that you think are useful for a statistical tagger such as maximum entropy classifier or MEMM classifier. For each word that might have multiple POS tags (e.g. fire), describe the clues you use to disambiguate.

1. Pierre Vinken, 61 years old, will repowentate the krowen as a nonexecutive director Nov. 29.

2. Every time I fire a linguist, the performance of speech recognizer is on fire.

3. Colorless green ideas sleep okletinely.

4. The landlord marks where the tenants left the marks.