Overview

- This homework is due on Moodle at 2:00 pm on Tuesday, February 3, 2015.
- This homework covers hidden Markov models and conditional random fields.

1 Hidden Markov Models, Sequence Prediction [Vighnesh Shiv, 35 points]

Complexity:

Suppose we have a hidden Markov model (HMM), and want to determine the highest-probability state sequence given an observation. Two ways to solve this problem are the naive algorithm and the Viterbi algorithm. The naive algorithm computes the probability of each possible state sequence and returns the sequence with the highest probability.

Question A [5 points]: What is the time complexity (big-O) of the naive algorithm?

Question B [5 points]: What is the time complexity (big-O) of the Viterbi algorithm?

Sequence Prediction:

Question C [10 points]: Implement these algorithms and submit neat commented code.:

- the Viterbi algorithm (finds the max-probability state sequence for a given HMM and observation);
- the forward algorithm (computes the probability a given HMM emits a given observation).

Question D [5 points]: The supplementary data folder contains 5 files titled sequenceprediction1.txt, sequenceprediction2.txt, ... sequenceprediction5.txt. Each file specifies an HMM. The first row contains two tab-delimited numbers: the number of states Y and the number of types of observations X. The X observations emit outputs $0, 1, \ldots, X - 1$.

The next Y rows of Y tab-delimited floating-point numbers describe the state transition matrix. Each row represents the current state, each column represents a state to transition to, and each entry represents the probability of that transition occurring. The next Y rows of X tab-delimited floating-point numbers describe the output emission matrix, encoded analogously to the state transition matrix. The file ends with 5 possible emissions from that HMM.

For each of these five HMMs:

- find the max-probability state sequence;
- find the probabilities of emitting the five sequences at the end of the corresponding file.

You may assume that the initial state is randomly selected along a uniform distribution.

Hidden Markov Model Training:

Ron is an avid music listener, and his genre preferences at any given time depend on his mood. Ron's possible moods are happy, mellow, sad, and angry. Ron experiences one mood per day (as humans are known to do) and chooses one of ten genres of music to listen to that day depending on his mood. Ron's roommate, who is known to take to odd hobbies, is interested in how Ron's mood affects his music selection, and thus collects data on Ron's mood and music selection for six years (2191 data points).

This data is contained in the supplementary file ron.txt. Each row contains two tab-delimited strings: Ron's mood and Ron's genre preference that day.

Question E [10 points]: Use a single M-step to train a Hidden Markov Model on the data in ron.txt. What are the learned state transition and output emission matrices?

2 Conditional Random Fields and HMMs [Vinny Augustine, 35 points]

Now we will train a conditional random field (CRF) for Ron's music interests. The CRF will use features of the form Y_t, Y_{t+1}. Under a CRF, the probability of observing $Y_{1:T}$ given corresponding $X_{1:T}$ is

$$P(Y_{1:T}|X_{1:T}) \propto \exp\left(\sum_{t=1}^{T}\langle\theta_n, f\left(X_t, Y_t\right)\rangle + \sum_{t=1}^{T-1}\langle\theta_e, f\left(Y_t, Y_{t+1}\right)\rangle\right)$$

Question F [30 points]: Train the CRF using gradient descent. Print out your model parameters. Does the CRF outperform the HMM? Under what conditions do you feel a CRF is better applicable than an HMM? Submit neat commented code. *Note:* Make assumptions when necessary but clearly state your assumptions.

Question G [2 points]: When the number of hidden states is unknown while training an HMM for a fixed observation set, if we want to increase the training data likelihood, we can do so by allowing more hidden states. True or false? Give an explanation.

Question H [3 points]: Prove that if a coefficient of the initial state or state transition probaiblity matrices of an HMM is initially 0, it will remain 0 until the end of the EM algorithm.