Machine Learning & Data Mining
Homework 4

Caltech CS/CNS/EE 155
January 26^{th}

This set is due 2:30pm, February 2^{th}, via Moodle. You are free to collaborate on all of the problems, subject to the collaboration policy stated in the syllabus.

1 Sequence Prediction

In this problem we will explore sequence prediction using Hidden Markov Models (HMM), as discussed in lecture.

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: What is the time complexity (big-O) of the naive algorithm?

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

Concepts

Question C: 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 D: 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.

Sequence Prediction

Please submit your code separately for *both* Questions E and F.

Question E: Implement:

i. The Viterbi algorithm.

ii. The Forward algorithm.

Question F: 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 $0, 1, ..., X-1$. 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:

i. Find the max-probability state sequence.

ii. 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.

HMM 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 Rons 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 G: 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 Naive Bayes

Question A: The Bayes Naive classifier selects the most likely classification V_{nb} given the attributes $a_1, a_2, \ldots, a_n$. This results in

$$V_{nb} = \arg\max_{v_j \in V} P(v_j) \prod P(a_i|v_j) \tag{1}$$

We generally estimate $P(a_i|v_j)$ using m-estimate:

$$P(a_i|v_j) = \frac{n_c + mp}{n + m}$$

where

$$\begin{aligned} n &= \text{the number of training examples for which } v = v_j \\ n_c &= \text{number of examples for which } v = v_j \text{ and } a = a_i \\ p &= \text{a priori estimate for } P(a_i|v_j) \\ m &= \text{the equivalent sample size} \end{aligned}$$

In this problem, let $m = 3$ and assume the uniform prior: $p = 1/(\text{number-of-attribute-values}) = \frac{1}{2}$ for all our attributes.

Consider the following data set, where Color, Type, and Origin are features, and Stolen? is either yes or no:

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	No
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	No
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Suppose you have a Red Domestic SUV. Given this dataset, calculate $P(\text{stolen}) \prod P(a_i|v_j)$ and $P(\text{not stolen}) \prod P(a_i|v_j)$ and the resulting V_{nb}. Would you classify your car as more likely to be stolen or not to be stolen?

Question B: If we were to use Naive Bayes with the data you used for the HMM problem, would it work well? If not, why? Hint: think about the assumptions Naive Bayes makes.