

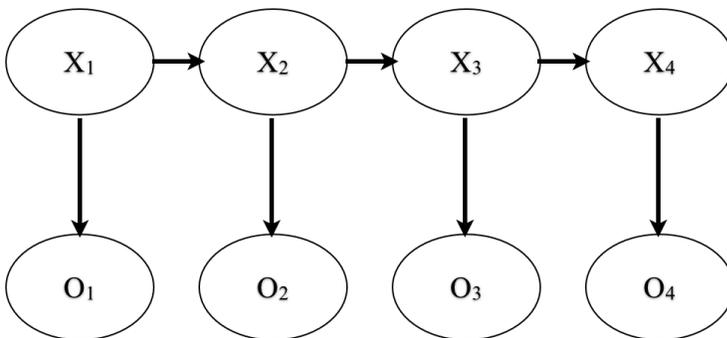
## 10-601 Machine Learning: Assignment 8

- The assignment is due at 3:00pm (beginning of class) on **Wednesday, April 2, 2008**.
- It will be worth **50 points**.
- Write your name at the top right-hand corner of each page submitted.
- Each student must hand in a hard-copy writeup. See the course webpage for collaboration policies.

### Q1: HMMs [25 pts]

You have been studying for midterms in Wean Hall for the past week and have not seen the light of day. The only knowledge of the weather you receive is based on whether or not the barista at the coffee stand arrives carrying an umbrella. Most of the time, an umbrella indicates it's raining. But sometimes the barista doesn't check the forecast and gets it wrong—thus, the observation is a noisy indication of the “true state” of the weather.

You can model this as an HMM, where each  $X$  is one day's weather (Rain = True or False), and each  $O$  is the observation of whether or not the barista has an umbrella. Each day's weather is dependent on the previous day's weather. So, you have the following HMM:



The CPT's are as follows:

- $P(O_i = True | X_i = True) = .9$
- $P(O_i = True | X_i = False) = .1$

- $P(X_{i+1} = True | X_i = True) = .8$
- $P(X_{i+1} = True | X_i = False) = .2$
- $P(X_1 = True) = 0.7$

1. Is  $X_3 \perp O_4 | X_2$ ?
2. What is  $P(X_3 = True | \mathbf{O} = \{False, True, True, False\})$ ?

## Q2: Support Vector Machines [15 pts]

We will use the following terms:

- $\mathbf{U}$  – a universe, of all integers from 1 to 100.
- $S_{\bar{x}}$  – a set that includes elements of  $\mathbf{U}$ .
- $\bar{x}$  – a binary vector of length 100, which represents the set  $S_{\bar{x}}$ . For instance, if the set were  $\{2,4\}$ , the vector representation would be  $[0,1,0,1,0,0,\dots,0]$ .
- $Pow(S_{\bar{x}})$  – the power set, or set of all subsets, of  $S_{\bar{x}}$ .  $Pow(S_{\bar{x}})$  has  $2^{|S_{\bar{x}}|}$  elements, for each subset of  $S_{\bar{x}}$ .
- $h(\bar{x})$  – a vector of size  $2^{100}$ , corresponding to  $Pow(S_{\bar{x}})$ .

Now, define  $f(\bar{x}_1, \bar{x}_2) = 2^{\bar{x}_1 \cdot \bar{x}_2}$ . Since  $\bar{x}_1 \cdot \bar{x}_2$  is the size of  $S_{\bar{x}_1} \cap S_{\bar{x}_2}$ , then  $f(\bar{x}_1, \bar{x}_2)$  is the size of  $Pow(S_{\bar{x}_1} \cap S_{\bar{x}_2})$ .

Is  $f$  a kernel function? If so, what is the Hilbert space?

## Q3: Collaborative Filtering [10 pts]

Suppose you were given data from a music community site (such as `last.fm`). Your data set includes data from 100,000 users. For each user you have some demographic information (age, sex, location, interests), and a log of every song they had listened to on their computer– including artist, title, album, and the date/time at which they listened to the song.

You want to build a system to recommend new music to users in this community. What collaborative filtering method presented in class would you use, and why would you choose that method over others?