Foundations of Algorithms and Machine Learning

(CS60020)

Spring Semester 2017

Assignment 7 Questions

1. A simple topic model: You are given a collection of documents, represented as "bags of words". In other words, each document is represented as a V-dimensional count vector X_d, where X_{dv} is the number of times word v has appeared in the d^{th} document. You want to find K "topics" in these documents. Imagine that each topic is a V-dimensional multinomial distribution ϕ_k, where ϕ_{kv} is the probability of word v appearing in the k^{th} topic.

(a) First, describe a generative model for the documents using the topics. Assume that each document comes from exactly one of the K topics. Explicitly indicate any additional parameter(s) that you may need in the generative process.
(b) Design an EM algorithm for estimating the parameters of the generative model. Specifically, derive the E and M steps of the algorithm.
(c) If you think that the EM algorithm does not work well, suggest an improvement.

2. Given a Hidden Markov Model with observations $X_1,\ldots,X_2$ and hidden states $Z_1,\ldots,Z_2$, design an efficient algorithm to compute $P(Z_t=k, Z_{t+1}=k' \vee X;\pi,A,\psi)$ for all $t=0,\ldots,T-1$. (Note that the EM algorithm for HMM requires this joint probability of adjacent states.).

3. You go to a shop to buy candy. The shopkeeper has N candies of various cost. You have to buy K candies such that the difference between the most expensive candy and the cheapest candy in your set of K candies is **minimum**.

Sample :
N = 7 [9, 100, 300, 200, 1000, 19, 29]
K = 3

we can select (9,19,29) as the best set because max(9,19,29) – min(9,19,29) = 20(lowest)

Design an efficient algorithm to find the set of K candies from N. Also report the time complexity.

4. Suppose you have been assigned the task of cutting a large metal sheet composed of A X B square pieces. You are only allowed to make horizontal and vertical cuts on the metal sheet. Each cut has a cost depending on whether it is a vertical or a horizontal cut. To break the metal sheet into squares of size 1 X 1, you would need to make (A-1) cuts horizontally and (B-1) cuts vertically. Each cut has a cost **c**. If you make a cut of cost ***c*** and it passes through ***n*** already-cut segments, the total cost of the cut is ***n x c***. ie. Initially the metal sheet is a single segment. If i make a single horizonal cut, there will be 2 segments. Now if i make a veritcal cut, it will go through these 2 segments. So the cost of vertical cut will be 2*C.

The cost of cutting the metal sheet into single squares is the sum of the cost of successive cuts used to cut the whole board into square metal pieces of size 1x1. Design an efficient algorithm to find the minimal cost of cutting the metal sheet into size of 1x1. Find the time complexity of the algorithm designed.

Example : Metal sheet of size 6 X 4.
cutting cost of row : [2 1 3 1 4]
cutting cost of column : [4 1 2]

The minimum cost is 42.