E0:270 Homework 3

April 11, 2016

Due: April 20, 2016

Problem 1

E-M algorithm for Bernoulli and Multinomial Mixture Models

- **Bernoulli Mixture Model**

  Consider a set of M binary variables $x_i \in \{0, 1\}$ where $i = 1, \ldots, M$, each of which is governed by a Bernoulli distribution with parameter μ_i so that $p(x; \mu) = \prod_{i=1}^{M} \mu_i^{x_i}(1 - \mu_i)^{(1-x_i)}$, where $x = \{x_1 \ldots x_M\}^T$, $\mu = \{\mu_1 \ldots \mu_M\}^T$. Consider a finite mixture of K such M-dimensional Bernoulli distributions: $p(x; \mu, \pi) = \sum_{k=1}^{K} \pi_k p(x; \mu_k)$, where $\mu = \{\mu_1 \ldots \mu_K\}$ with $\mu_i = \{\mu_{i1} \ldots \mu_{iM}\}^T$ and $\pi = \{\pi_1 \ldots \pi_K\}$, with $\pi_i \geq 0$ and $\sum_{i=1}^{K} \pi_i = 1$.

  (a) Derive the steps of the EM algorithm for finding the maximum likelihood estimates for this mixture distribution [**10 marks**].

  (b) Implement this algorithm on the dataset *MoBern.mat*, and estimate μ and π. $K = 5, M = 5$ [**10 marks**].

- **Multinomial Mixture Model**

  Now imagine the M variables to be categorical $x_i \in \{1, \ldots, L\}$ where $i = 1, \ldots, M$, each of which is governed by a Multinomial distribution with parameter $\mu_i = \{\mu_{ij}\}_{j=1}^{L}$ so that $p(x; \mu) \propto \prod_{i=1}^{M} \prod_{j=1}^{L} \mu_{ij}^{x_{ij}}$, where $x = \{x_1 \ldots x_M\}^T$, $\mu = \{\mu_1 \ldots \mu_M\}^T$. Consider a finite mixture of K such M-dimensional Multinomial distributions: $p(x; \mu, \pi) = \sum_{k=1}^{K} \pi_k p(x; \mu_k)$, where $\mu = \{\mu_1 \ldots \mu_K\}$ with $\mu_k = \{\mu_{kij}\}$, $i = 1 \ldots M$, $j = 1 \ldots L$; and $\pi = \{\pi_1 \ldots \pi_K\}$ as before.

  (a) Adapt the steps of the earlier EM algorithm for finding the maximum likelihood estimates for this multinomial mixture distribution [**5 marks**].

  (b) Implement this algorithm on the dataset *MoMult.mat*, and estimate μ and π. $K = 3, L = 3, M = 3$ [**5 marks**]

Problem 2

Here is a library for Hidden Markov Models.

Hidden Markov Model - Part 1 You are given a dataset X, in the form of an $N \times T \times 2$ matrix ($N = 200$ data sequences, each of length $T = 50$ and each instance of dimension 2), in `HMMSeq.mat`. Each sequence in the dataset was generated from an HMM with K^* components (where K^* is unknown for you), with each state having a Gaussian conditional.

1. Given the number of states K, use the function `mhmm_em.m` to learn the parameters of the model $HMM(K)$ from the data using the EM algorithm. Vary the number of states K from 1 to 6 to get 6 different models $\{HMM(K)\}_{K=1}^{6}$. **[5 marks]**

2. Using the function `viterbi_path.m`, and the models $HMM(1), \ldots, HMM(6)$, predict the most likely configuration of states $z_1 \ldots z_T$ for *only the first sequence* in $\mathbf{X}$. For each model, make two plots as follows: In the first plot, plot the two dimensional points $\{x_t\}_{t=1}^{T}$ using a particular color for a specific value of z_t. In the second plot, plot the sequence $\{(t, z_t)\}_{t=1}^{T}$, where z_t takes values from $\{1, ..K\}$, to observe the state transitions. **[10 marks]**

3. Using only $HMM(3)$, run the Viterbi algorithm on all the sequences one by one.

   (a) After processing each sequence, try to empirically estimate the state transition matrix from the most likely sequences of state transitions in the sequences processed so far (using the counts of different state transitions). Show any three of the estimated matrices **[2 marks]**

   (b) Compute the Frobenius-norm errors between the state transition matrix of $HMM(3)$ (learnt by Baum-Welch), and those by using the most likely state assignments. Plot this error against the number of sequences used in your estimation. **[3 marks]**

Hidden Markov Model - Part 2 You are given another dataset Y, in the form of an $N \times T \times 2$ matrix ($N = 200$ data sequences, each of length $T = 50$ and each instance of dimension 2), in `Seq.mat`. Each sequence in the dataset was generated using 3 Gaussians. However, you do not know whether the sequences were created using HMM, or the individual points in any sequence are sampled IID, as in a GMM.

(a) Discuss how you can decide which is the case. **[2 marks]**

(b) Implement it on the given data. Write your findings in the report. **[8 marks]**

Problem 3

Gibbs Sampling in Hidden Markov Models

- A discrete-output HMM is parametrized by its state transition matrix A and an emission matrix B. Each row of B represents one state, and is a multinomial distribution over the discrete output space. In other works, $A_{kl} = prob(Z_t = l | Z_{t-1} = k)$ and $B_{ki} = prob(X_t = i | Z_t = k)$, for any i, k, l, t, where X is the observed output and Z is the latent state variable.

  (a) Write down an algorithm based on Gibbs Sampling to infer these latent variables Z_t using X, A and B. **[5 marks]**

  (b) In the file $HMM-Gibbs.mat$, you are provided with (A, B) for an HMM, and a set X of $N = 100$ sequences, each of length $T = 50$, over the output space $\{1, 2, 3, 4, 5\}$. Implement your Gibbs Sampling-based algorithm, and infer the state variables in all the sequences. Report your results in a .mat file (3A.mat, to be uploaded). Compare the result with the most likely sequence found by Viterbi's algorithm (write in your report). **[10 marks]**

- **Mixture of HMMs** You are provided S sets of such HMM parameters $\{A_s, B_s\}$, and a set of N sequences, each of length T, over the output space. You do not know which sequence has been generated from which HMM. Set an unknown variable S_n with each sequence n, which takes a value in $\{1, 2, \ldots, S\}$. The prior distribution over S is π (known), i.e. $prob(S_n = s) = \pi_s$. Clearly, if $S_n = s$, then $prob(Z_t^n = l | Z_{t-1}^n = k) = A_{kl}^s$ and $prob(X_t^n = i | Z_t^n = k) = B_{ki}^s$.

  (a) Write down an algorithm based on Gibbs Sampling to infer the latent variables S_n and Z_t^n, using $\{A_k, B_k\}, \pi$ **[5 marks]**

(b) In the file $MHMM - Gibbs.mat$, you are provided 500 sequences of length 50, over the output space $\{1, 2, 3, 4, 5\}$. Implement the above algorithm, and infer the hidden state variables Z_t^n as well as the HMM variables S_n. Report your results in a .mat file (3B.mat, to be uploaded). [**10 marks**]

Problem 4

Bayesian estimation of HMM parameters

Suppose you do not know the matrices A and B (in case of Problem 3). One possibility is to estimate them using E-M algorithm. But another option is to consider them also as latent variables, in addition to Z. Write down the joint distribution of (X, Z, A, B) using suitable priors on A and B (hint: consider Dirichlet priors). Can you solve this inference problem using Gibbs Sampling? If so, write down the Gibbs Sampling steps. If not, discuss alternatives. [**10 marks**]

Non-credit: Read up Collapsed Gibbs Sampling