CS405 Homework 7

Course: Machine Learning(CS405) - Professor: Qi Hao

Question 1

Consider a density model given by a mixture distribution

$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|k)$

and suppose that we partition the vector $\mathbf{x}$ into two parts so that $\mathbf{x} = (\mathbf{x}_a, \mathbf{x}_b)$. Show that the conditional density $p(\mathbf{x}_b|\mathbf{x}_a)$ is itself a mixture distribution and find expressions for the mixing coefficients and for the component densities.

Question 2

Imagine a class where the probability that a student gets an "A" grade is P (A) =1/2, a "B" grade $P(B) = \mu$, a "C" grade $P(C) = 2\mu$, and a "D" grade $P(D) = 1/2 - 3\mu$. We are told that c students get a "C" and d students get a "D". We don't know how many students got exactly an "A" or exactly a "B". But we do know that h students got either an a or b. Therefore, a and b are unknown values where a + b = h. Our goal is to use expectation maximization to obtain a maximum likelihood estimate of μ.

(a) Expectation step: Compute the expected values of a and b given μ.

(b) Maximization step: Given the expected values of a and b, compute the maximum likelihood estimate of μ.

Hint. Compute the MLE of μ assuming unobserved variables are replaced by their expectation.

Question 3

Assume each data point $X_i \in \mathbb{R}^+ (i = 1 \ldots n)$ is drawn from the following process:

$Z_i \sim Multinomial(\pi_1, \pi_2, \ldots, \pi_K)$

$X_i \sim Gamma(2, \beta_{Z_i})$

The probability density function of $Gamma(2, \beta)$ is $P(X = x) = \beta^2 x e^{-\beta x}$

(a) Assume $K = 3$ and $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 4$. What's $P(Z = 1|X = 1)$?

(b) Describe the E-step: compute $P(Z = k|X = x)$ for each $X = x$. Write an equation for each value being computed.

Hint. $P(Z = 1|X = 1) = \frac{P(X=1|Z=1)P(Z=1)}{\sum\limits_{k=1} P(X=1|Z=k)P(Z=k)}$

Question 4

Verify the M-step equations (13.18) and (13.19) for the initial state probabilities and transition probability parameters of the hidden Markov model by maximization of the expected complete-data log likelihood function (13.17), using appropriate Lagrange multipliers to enforce the summation constraints on the components of π and $\mathbf{A}$.

Question 5

For a hidden Markov model having discrete observations governed by a multinomial distribution, show that the conditional distribution of the observations given the hidden variables is given by (13.22) and the corresponding M step equations are given by (13.23). Write down the analogous equations for the conditional distribution and the M step equations for the case of a hidden Markov with multiple binary output variables each of which is governed by a Bernoulli conditional distribution.

Hint. Refer to Section 2.1 and 2.2 for a discussion of the corresponding maximum likelihood solutions for i.i.d. data if required.

Question 6

Suppose we wish to train a hidden Markov model by maximum likelihood using data that comprises R independent sequences of observations, which we denote by $\mathbf{X}^{(r)}$ where $r = 1, \ldots, R$.

(a) Show that in the E step of the EM algorithm, we simply evaluate posterior probabilities for the latent variables by running the $\alpha$and β recursions independently for each of the sequences.

(b) Show that in the M step, the initial probability and transition probability parameters are re-estimated using modified forms of (13.18) and (13.19) given by

$$\pi_k = \frac{\sum_{r=1}^{R} \gamma(z_{1k}^{(r)})}{\sum_{r=1}^{R} \sum_{j=1}^{K} \gamma(z_{1j}^{(r)})}$$

$$A_{jk} = \frac{\sum_{r=1}^{R} \sum_{n=2}^{N} \xi(z_{n-1,j}^{(r)}, z_{n,k}^{(r)})}{\sum_{r=1}^{R} \sum_{l=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}^{(r)}, z_{n,l}^{(r)})}$$

where for notational convenience, we have assumed that the sequences are of the same length (the generalization to sequences of different lengths is straightforward).

(c) Show that the M-step equation for re-estimation of the means of Gaussian emission models is given by

$$\mu_k = \frac{\sum_{r=1}^{R} \sum_{n=1}^{N} \gamma(z_{nk}^{(r)}) \mathbf{x}_n^{(r)}}{\sum_{r=1}^{R} \sum_{n=1}^{N} \gamma(z_{nk}^{(r)})}$$

Note that the M-step equations for other emission model parameters and distributions take an analogous form.