Mixture Models and EM

1. **Gaussian**

   Consider a Gaussian mixture model in which the marginal distribution $p(z)$ for the latent variable is given by (9.10), and the conditional distribution $p(x|z)$ for the observed variable is given by (9.11). Show that the marginal distribution $p(x)$, obtained by summing $p(z)p(x|z)$ over all possible values of z, is a Gaussian mixture of the form (9.7). (Bishop 9.3)

2. **EM and Mixture Models**

   Show that if we maximize (9.40) with respect to μ_k while keeping the responsibilities $\gamma(znk)$ fixed, we obtain the closed form solution given by (9.17). (Bishop 9.8) (Hint use equation 2.43)

3. **Conditional Mixture Distribution, Bishop 9.10**

Consider the density model given by a mixture distribution

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x} \mid k)$$

and suppose that we partition the vector $\boldsymbol{x}$ into two parts, so that $\boldsymbol{x} = (\boldsymbol{x}_a, \boldsymbol{x}_b)$. Show that the conditional density $p(\boldsymbol{x}_b \mid \boldsymbol{x}_a)$ is itself a mixture distribution, where the components are the conditionals for component k (e.g., $p(\boldsymbol{x}_b \mid \boldsymbol{x}_a, k)$) .

4. **Mean of a Mixture Model, Bishop 9.19**

Consider a D-dimensional variable $\boldsymbol{x}$ each of whose components is itself a multinomial variable of degree M so that $\boldsymbol{x}$ is a binary vector with components x_{ij}m where $i \in [1, D], j \in [1, M]$, subject to the constraint that $\sum_j x_{ij} = 1$ for all i. Suppose that the distribution of these variables is described by a mixture of the discrete multinomial distributions considered in Section 2.2 so that

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x} \mid \mu_k)$$

where we have

$$p(\boldsymbol{x} \mid \mu_k) = \prod_{i=1}^{D} \prod_{j=1}^{M} p(x_{ij} = 1 \mid \mu_k)^{x_{ij}}$$

Given an observed data set $\{\boldsymbol{x}_n\}$, derive the E step of the EM algorithm for optimize the mixing coefficients and the component parameters of this distribution my ML.

5. Gaussian Mixture Models

Suppose we have a Gaussian mixture model where the covariance matrices Σ_k are all constrained to have a common value, call it Σ. Derive the EM condition for maximizing the likelihood function with respect to Σ under such a model.

6. **E and M**

Show that if we maximize the expected complete-data log likelihood function (9.55) for a mixture of Bernoulli distributions with respect to μ_k , we obtain the M step equation (9.59). (Bishop 9.15)

7. **E and M**

Show that maximization of the expected complete-data log likelihood function (9.62) for the Bayesian linear regression model leads to the M step re- estimation result (9.63) for α. (Bishop 9.20)

8. **E and M**

Consider a mixture distribution of the form

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

where the elements of x could be discrete or continuous or a combination of these. Denote the mean and covariance of $p(x|k)$ by μ_k and Σ_k, respectively. Show that the mean and covariance of the mixture distribution are given by (9.49) and (9.50) (Bishop 9.12)

9. **E and M**

Show that if we maximize the expected complete-data log-likelihood function given in eq. 9.55 for a mixture of Bernoulli's with respect to μ_k, we obtain the M-step equation 9.59.

10. **E and M**

Show that if we maximize the expected complete-data log-likelihood function given in eq. 9.55 for a mixture of Bernoulli's with respect to the mixing coefficients π_k, using a Lagrange multiplier to enforce to summation constraint (they must sum to 1), we obtain the M-step equation 9.60.

11. **Bernoulli Mixtures**

Consider the joint distribution of latent and observed variables for the Bernoulli distribution obtained by forming the product of the $p(\boldsymbol{x} \mid z, \mu)$ given by 9.52 and $p(z \mid \pi)$ given by 9.53. Show that if we marginalize this joint distribution with respect to z (i.e., sum over all possible choices for z), we obtain 9.59.