Practice Problem Set II

Gaussian Mixture Models, Expectation Maximization, Variational Bayes, Sampling

(From the text books by Kevin Murphy, and David Barber)

Exercise 11.2 EM for mixtures of Gaussians

Show that the M step for ML estimation of a mixture of Gaussians is given by

$$\boldsymbol{\mu}_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k} \quad (11.114)$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T}{r_k} = \frac{\sum_i r_{ik}\mathbf{x}_i\mathbf{x}_i^T - r_k\boldsymbol{\mu}_k\boldsymbol{\mu}_k^T}{r_k} \quad (11.115)$$

Exercise 11.3 EM for mixtures of Bernoullis

- Show that the M step for ML estimation of a mixture of Bernoullis is given by

$$\mu_{kj} = \frac{\sum_i r_{ik}x_{ij}}{\sum_i r_{ik}} \quad (11.116)$$

- Show that the M step for MAP estimation of a mixture of Bernoullis with a $\beta(\alpha, \beta)$ prior is given by

$$\mu_{kj} = \frac{(\sum_i r_{ik}x_{ij}) + \alpha - 1}{(\sum_i r_{ik}) + \alpha + \beta - 2} \quad (11.117)$$

Exercise 11.7 Manual calculation of the M step for a GMM

(Source: de Freitas.) In this question we consider clustering 1D data with a mixture of 2 Gaussians using the EM algorithm. You are given the 1-D data points $x = [1 \quad 10 \quad 20]$. Suppose the output of the E step is the following matrix:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0.4 & 0.6 \\ 0 & 1 \end{bmatrix} \quad (11.127)$$

where entry $r_{i,c}$ is the probability of obervation x_i belonging to cluster c (the responsibility of cluster c for data point i). You just have to compute the M step. You may state the equations for maximum likelihood estimates of these quantities (which you should know) without proof; you just have to apply the equations to this data set. You may leave your answer in fractional form. Show your work.

a. Write down the likelihood function you are trying to optimize.
b. After performing the M step for the mixing weights π_1, π_2, what are the new values?
c. After performing the M step for the means μ_1 and μ_2, what are the new values?

Exercise 201. *Derive the optimal EM update for fitting a mixture of Gaussians under the constraint that the covariances are diagonal.*

Exercise 202. *Consider a mixture of K isotropic Gaussians, each with the same covariance, $\mathbf{S}_i = \sigma^2\mathbf{I}$. In the limit $\sigma^2 \to 0$ show that the EM algorithm tends to the K-means clustering algorithm.*

Exercise 21.7 Forwards vs reverse KL divergence

(Source: Exercise 33.7 of (MacKay 2003).) Consider a factored approximation $q(x, y) = q(x)q(y)$ to a joint distribution $p(x, y)$. Show that to minimize the forwards KL $\mathbb{KL}(p||q)$ we should set $q(x) = p(x)$ and $q(y) = p(y)$, i.e., the optimal approximation is a product of marginals

Now consider the following joint distribution, where the rows represent y and the columns x.

	x			
	1	2	3	4
1	1/8	1/8	0	0
2	1/8	1/8	0	0
3	0	0	1/4	0
4	0	0	0	1/4

Show that the reverse KL $\mathbb{KL}(q||p)$ for this p has three distinct minima. Identify those minima and evaluate $\mathbb{KL}(q||p)$ at each of them. What is the value of $\mathbb{KL}(q||p)$ if we set $q(x, y) = p(x)p(y)$?

Exercise 265. *We wish to find a Gaussian approximation* $q(x) = \mathcal{N}(x|m, s^2)$ *to a distribution* $p(x)$. *Show that*

$$KL(p|q) = -\langle \log q(x) \rangle_{p(x)} + \text{const}. \qquad (28.11.29)$$

Write the KL divergence explicitly as a function of m *and* s^2 *and confirm the general result that the optimal* m *and* s^2 *that minimise* $KL(p|q)$ *are given by setting the mean and variance of* q *to those of* p.

The angles denote expectation.

Exercise 21.8 Derive the mean field Variational Bayes (VB) for Linear Regression models with suitable assumption about the true posterior.

Exercise 21.9 Derive the mean field VB for Gaussian Mixture Models. Derive the expression for ELBO.

11.6 ($\star\star$) **www** In this exercise, we show more carefully that rejection sampling does indeed draw samples from the desired distribution $p(\mathbf{z})$. Suppose the proposal distribution is $q(\mathbf{z})$ and show that the probability of a sample value $\mathbf{z}$ being accepted is given by $\widetilde{p}(\mathbf{z})/kq(\mathbf{z})$ where $\widetilde{p}$ is any unnormalized distribution that is proportional to $p(\mathbf{z})$, and the constant k is set to the smallest value that ensures $kq(\mathbf{z}) \geqslant \widetilde{p}(\mathbf{z})$ for all values of $\mathbf{z}$. Note that the probability of drawing a value $\mathbf{z}$ is given by the probability of drawing that value from $q(\mathbf{z})$ times the probability of accepting that value given that it has been drawn. Make use of this, along with the sum and product rules of probability, to write down the normalized form for the distribution over $\mathbf{z}$, and show that it equals $p(\mathbf{z})$.

Exercise 24.1 Gibbs sampling from a 2D Gaussian

Suppose $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (1, 1)$ and $\boldsymbol{\Sigma} = (1, -0.5; -0.5, 1)$. Derive the full conditionals $p(x_1|x_2)$ and $p(x_2|x_1)$. Implement the algorithm and plot the 1d marginals $p(x_1)$ and $p(x_2)$ as histograms. Superimpose a plot of the exact marginals.

Exercise 24.2 Derive Gibbs sampling steps for a Gaussian Mixture Model with k components.