T-61.5040 Oppivat mallit ja menetelmät
T-61.5040 Learning Models and Methods
Pajunen, Viitaniemi

Exercises 10, 30.3.2007

Problem 1.

Consider a mixture model

$$p(y_i|\theta,\lambda) = \sum_j \lambda_j N(y_i|\mu_j,\sigma_j^2)$$

and $L_{ij} = 1$ exactly when y_i is generated by the mixture component j. Otherwise $L_{ij} = 0$. We denote $\lambda_j = p(L_{ij} = 1)$ at all observations i. The parameter vector θ is $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2 \ldots)$.

i) Show that

$$p(y, L|\theta,\lambda) = \prod_{j=1}^{m} \prod_{i:L_{ij}=1} \lambda_j N(y_i|\mu_j,\sigma_j^2)$$

Hint: use the product rule on $p(y, L|\theta, \lambda)$.

ii) Compute the full conditional posteriors

$$p(\theta|y, L) \propto p(y|\theta, L)p(\theta|L)$$

and

$$p(L|y, \theta)$$

that are needed to implement the Gibbs sampler. Assume that the variances σ_j^2 are known.

Problem 2.

Consider the two-mixture model given in the lectures. Data is i.i.d. distributed as $p(y_i|\theta,\lambda) = \lambda_1 N(y_i|\mu_1,\sigma^2) + \lambda_2 N(y_i|\mu_2,\sigma^2)$, where the variance σ^2 is known. Here we have written $\theta = \{\mu_1, \mu_2, \sigma^2\}$ and $\lambda = \{\lambda_1, \lambda_2\}$ for the parameters of the model. Also, it holds that $\lambda_1 + \lambda_2 = 1$. Define variables L_{im} so that $L_{im} = 1$ when the data point y_i is generated by the m:th mixture component, and $L_{im} = 0$ otherwise.
Compute the Newton-Rhapson update $\mu_{m,new} = \mu_m - (\log p)'/(\log p)''$ for maximizing the log-likelihood $\log p(y|\theta,\lambda)$. When computing the derivatives, use the simplifying approximation of regarding $p(L_{im} = 1|\theta, y_i)$ as constant with respect to μ_m.

Problem 3.

In the lecture notes, the following function was seen to be a lower bound for the quantity that we want to maximize in the EM algorithm:

$$F(q, a) = E_q(\log p(s, a|y)) - E_q(\log q(s))$$

This is maximized alternatingly with respect to the distribution $q(s)$ and the parameters a. Show that

$$F(q, a) = -D(q||p(s|a, y)) + \log p(a|y)$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence between two distributions. Describe what happens at each maximization step. Show that the EM algorithm guarantees that the posterior $p(a|y)$ does not decrease. Above, s denotes the latent variables, $q(s)$ is the distribution of s, a denotes the parameters of the model, and y is the data.

Problem 4. (demo)

In the E-step of the EM algorithm, $F(q, a)$ is maximized by choosing the distribution $q(s)$ to equal $p(s|a_0, y)$. (Here a_0 denotes the current guess of the parameters). In the Normal mixture model, the function $F(q, a)$ can be shown to be

$$F(q, a) = \sum_m \sum_i \left[\log N(y_i|\mu_m, \Sigma_m) + \log \lambda_m\right] \tau_{im}$$

where λ_m is the mixture proportion (the weight of the m:th mixture component) and

$$\tau_{im} = p(L_{im} = 1|a_0, y_i).$$

Derive the M-step for updating λ_m and μ_m.