N. Bertschinger
M. Kaschube

Exercise Sheet 4
Discussed on Wed, May 16, 14:15

Exercise 1. *Consider a data set* $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)^T$ *in which the observations* $\{\mathbf{x}_n\}$ *are assumed to be drawn independently from a multivariate Gaussian distribution. Write down the log likelihood function* $\ln p(\mathbf{X}|\mu, \mathbf{\Sigma})$ *for this distribution. Estimate the mean* μ *by maximum likelihood. Interpret the result.*

2 points

Exercise 2. *Mixture of Gaussians: the conditional distribution of* $\mathbf{x}$ *given a particular value for the latent variable* $\mathbf{z}$ *is assumed to be a Gaussian, i.e.* $p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\mu_{\mathbf{k}}, \mathbf{\Sigma_k})$. *The probability* $p(z_k = 1) = \pi_k$ *is called mixing coefficient. Derive an expression for the conditional distribution of* $\mathbf{z}$ *given* $\mathbf{x}$, *i.e.* $p(z_k = 1|\mathbf{x})$ *in terms of* π_k *and* $\mathcal{N}(\mathbf{x}|\mu_{\mathbf{k}}, \mathbf{\Sigma_k})$. *Use the derived mathematical expression to interpret* $\gamma(z_k) \equiv p(z_k = 1|\mathbf{x})$ *(with words and by drawing a cartoon).*

2 points

Exercise 3. *a) Write down the log of the likelihood function of the Gaussian mixture model* $\ln p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma})$. *Explain why it is difficult to find the parameters for which this function has a maximum.*

1 point

b) Write down the conditions for μ_k *that must be satisfied at a maximum of the log likelihood function. Derive the expression*

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

with

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

end interpret the result.

2 points

c) Now maximize the log likelihood function with respect to the mixing coefficients π_k. To do so maximize

$$\ln p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma}) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

where λ is a Lagrange multiplier implementing the constraint that the π_k must sum to 1. Show that

$$\pi_k = \frac{N_k}{N}$$

and interpret the result.

2 points

d) Assuming Σ is fixed, why do these expression not constitute a closed-form solution for the parameters of the mixture model?

1 point