Machine Learning II

N. Bertschinger
M. Kaschube

Exercise Sheet 5
Discussed on Wed, May 23, 14:15

Exercise 1. *a) Obtain a relationship between K means and EM for Gaussian mixtures by considering a mixture model in which all components have covariance* $\epsilon\mathbf{I}$. *Show that*

$$p\left(\mathbf{x}|\mu_k, \Sigma_k\right) = \frac{1}{(2\pi\epsilon)^{1/2}} \exp\left\{-\frac{1}{2\epsilon}\|\mathbf{x}-\mu_k\|^2\right\}$$

1 point

b) Further, show

$$\gamma\left(z_{nk}\right) = \frac{\pi_k \exp\left\{-\|\mathbf{x}_n-\mu_k\|^2/2\epsilon\right\}}{\sum_j \pi_j \exp\left\{-\|\mathbf{x}_n-\mu_j\|^2/2\epsilon\right\}}$$

0 points

and discuss the limit case $\epsilon \to 0$, *specifically the link to the indicator variables*

$$r_{nk} = \begin{cases} 1 & \textit{if } k = \arg\min_j \|\mathbf{x}_n-\mu_j\|^2 \\ 0 & \textit{otherwise.} \end{cases}$$

1 point

c) Show further that in this limit, maximizing the expected complete-data log likelihood for this model, given by

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X},\mathbf{Z}|\mu,\Sigma,\pi)] = \sum_{n=1}^{N}\sum_{k=1}^{K}\gamma\left(z_{nk}\right)\left\{\ln\pi_k + \ln\mathcal{N}\left(\mathbf{x}_n|\mu_k,\Sigma_k\right)\right\}$$

is equivalent to minimizing the distortion measure

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|\mathbf{x}_n-\mu_k\|^2$$

for the K-means algorithm.

1 points

Exercise 2. *The data file 'two_clusters.txt" contains 100 data points drawn from a two-dimensional distribution. Fit a mixture of two circular Gaussian distributions to these data using EM. Do not allow the variance of either of the Gaussians to become smaller than a minimal value of 0.0001.*

Machine Learning II

N. Bertschinger
M. Kaschube

Exercise Sheet 5

Discussed on Wed, May 23, 14:15

2 points

Exercise 3. *Explore what happens to the fit of the mixture of Gaussians model from exercise 2 as the number of data points from each Gaussian is reduced and the number of potential Gaussians is increased. If you set the minimal variance given in exercise 1 to 0, a Gaussian distribution can settle around a single sample point and then have its variance shrink to 0. Why does this pathological behavior occur?*

1 points

Exercise 4. *Modify your code from exercise 2 to calculate function* $\mathcal{L}(q, \theta)$ *during each E and M step of EM. Check that* $\mathcal{L}$ *changes monotonically. Explicitly calculate the true log likelihood* $\ln p(\mathbf{X}, \theta)$ *of the data at the end of each M phase. Is it equal to* $\mathcal{L}$*?*

1 points

Exercise 5. *Modify the code in exercise 2 to fit a K-means model rather than a mixture of Gaussians. Can you see any practical difference in the solutions that arise?*

2 points