Intelligent Data Analysis

Martin Russell
School of Computer Science
Tuesday, 19 March 2019

Exercise sheet – week 9 – Gaussian Mixture Models (GMMs)

1. Let $X = \{x_1, x_2, x_3, \ldots, x_N\}$ be a set of real numbers.
   a. Show that the Maximum Likelihood estimate of the parameters m and v of a Gaussian probability density function for the set X are given by:

$$m = \frac{1}{N}\sum_{n=1}^{N} x_n \, , v = \frac{1}{N}\sum_{n=1}^{N} (x_n - m)^2$$

   [6 marks]

   b. Are these values of m and v a local or global maximum? Justify your answer.

   [4 marks]

2. A 2-dimensional Gaussian PDF g has mean $m = \begin{bmatrix}2\\2\end{bmatrix}$ and covariance matrix $C = \begin{bmatrix}13 & -5.2\\-5.2 & 7\end{bmatrix}$.

   The matrix C has eigenvalue decomposition $C = UDU^T$, where:

$$U = \begin{bmatrix} cos\left(\frac{\pi}{3}\right) & -sin\left(\frac{\pi}{3}\right) \\ sin\left(\frac{\pi}{3}\right) & cos\left(\frac{\pi}{3}\right) \end{bmatrix}, D = \begin{bmatrix}4 & 0\\0 & 16\end{bmatrix}.$$

   a. Sketch a 1-standard-deviation contour for g. **[4 marks].**

   b. Calculate $g\left(\begin{bmatrix}1.5\\1.5\end{bmatrix}\right)$. Show all of your calculations. **[4 marks].**

3. What are the similarities and differences between using an Mcomponent GMM to model a set of data points in N dimensional space compared to using a set of M centroids obtained through clustering? **[4 marks].**

4. Why is the E-M algorithm necessary? Why doesn't the simple maximum likelihood parameter estimation procedure from question 1 apply to an M-component Gaussian Mixture Model (GMM)? **[4 marks].**

5. Let $X = \{x_1, x_2, x_3, x_4\}$, where x_1= 1, x_2 = 7, x_3 = 5, x_4 = 4. Suppose that:
   - g_1 is a Gaussian PDF with mean m_1 = 2 and variance v_1 = 2, and
   - g_2 is a Gaussian PDF with mean m_2 = 3 and variance v_2 = 2, and
   - g is the Gaussian Mixture Model $g(x) = 0.3 \times g_1(x) + 0.7 \times g_2(x)$. Type equation here.

   a. Calculate the new values of the means m_1 and m_2 after the application of one iteration of the E-M algorithm with the samples X. **[8 marks].**

   b. Are the new values of means m_1 and m_2 guaranteed to correspond to a global or local maximum of the likelihood function? **[4 marks].**

[Total marks 38]