ECE641 Lab3 Report

Liming Chen (PUID: 0032119074, Mechanical Engineering)

November.22 2021

1 Introduction

N.A.

2 Parameter Estimation for Multivariate Gaussian Distributions

2.1 show that the density function of (10) with parameters θ of (11) forms an exponential family with natural sufficient statistics given in (12), (13), (14)

$$p(y,x|\theta) = \prod_{n=0}^{N-1}\prod_{k=0}^{K-1} \frac{\pi_k}{(2\pi)^{K/2}}|R_k|^{-1/2}\exp\left\{-\frac{1}{2}(y_n-\mu_k)^t R_k^{-1}(y_n-\mu_k)\right\} \tag{1}$$

so,

$$\begin{aligned}
\log p(y,x|\theta) &= \sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\delta(x_n-k)\left\{\log\left(\frac{1}{(2\pi)^{L/2}|R_k|^{1/2}}\right)-\frac{1}{2}(y_n-\mu_k)^t R_k^{-1}(y_n-\mu_k)\right\}\\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\delta(x_n-k)\Bigg\{\log\left(\frac{1}{(2\pi)^{L/2}|R_k|^{1/2}}\right)-\frac{1}{2}y_n^t R_k^{-1}y_n+\mu_k^t R_k^{-1}y_n-\frac{1}{2}\mu_k^t R_k^{-1}\mu_k+\log(\pi_k)\Bigg\}\\
&= \sum_{k=0}^{K-1}\Bigg[\sum_{n=1}^{N}\delta(x_n-k)\log(\pi_k)+\sum_{n=1}^{N}\delta(x_n-k)\log\left(\frac{1}{(2\pi)^{L/2}|R_k|^{1/2}}\right)-\\
&\qquad \frac{1}{2}\sum_{n=1}^{N}\delta(x_n-k)tr(y_n y_n^t R_k^{-1})+\sum_{n=1}^{N}\delta(x_n-k)\mu_k^t R_k^{-1}y_n-\frac{1}{2}\sum_{n=1}^{N}\delta(x_n-k)\mu_k^t R_k^{-1}\mu_k\Bigg]\\
&= \sum_{k=0}^{K-1}\Bigg[\sum_{n=1}^{N}\delta(x_n-k)\log(\pi_k)+\sum_{n=1}^{N}\delta(x_n-k)\log\left(\frac{1}{(2\pi)^{L/2}|R_k|^{1/2}}\right)-\\
&\qquad \frac{1}{2}tr\left(\sum_{n=1}^{N}\delta(x_n-k)y_n y_n^t R_k^{-1}\right)+\sum_{n=1}^{N}\delta(x_n-k)\mu_k^t R_k^{-1}y_n-\frac{1}{2}\sum_{n=1}^{N}\delta(x_n-k)\mu_k^t R_k^{-1}\mu_k\Bigg]
\end{aligned} \tag{2}$$

so, from equation 2, the natural sufficient statistics will be:

$$\begin{aligned}
N_k &= \sum_{n=0}^{N-1}\delta(x_n-k)\\
t_{1,k} &= \sum_{n=0}^{N-1}y_n\delta(x_n-k)\\
t_{2,k} &= \sum_{n=0}^{N-1}y_n y_n^t\delta(x_n-k)
\end{aligned} \tag{3}$$

2.2 Show that the ML estimate of θ given (Y,X) is formed by the expressions in (15), (16) and (17).

Substitute equation 3 to equation 2, we can get:

$$\log p(y,x|\theta) = \sum_{k=0}^{K-1}\left[N_k\log(\pi_k) + N_k\log\left(\frac{1}{(2\pi)^{L/2}|R_k|^{1/2}}\right) - \frac{1}{2}tr(t_{2,k}R_k^{-1}) + \mu_k^t R_k^{-1}t_{1,k} - \frac{1}{2}N_k\mu_k^t R_k^{-1}\mu_k\right] \quad (4)$$

So, we can estimate the parameter θ using Maximum Likelihood, for π_k we use Lagrange Multipliers to do the constraint optimization.
(1) For π_k, we first construct a function:

$$L(\theta) = \log p(y,x|\theta) + \lambda(\sum_{k=0}^{K-1}\pi_k - 1) \quad (5)$$

Then, calculate the derivative of equation 5 w.r.t. π_k and make it as 0:

$$\frac{\partial L(\theta)}{\partial \pi_k} = \frac{1}{\pi_k}N_k + \lambda = 0 \quad (6)$$

We can get:

$$\pi_k = -\frac{1}{\lambda}N_k \quad (7)$$

Due to the constraint:

$$\sum_{k=0}^{K-1}\pi_k = -\frac{1}{\lambda}\sum_{k=0}^{K-1}\sum_{n=0}^{N-1}\delta(x_n-k) = -\frac{1}{\lambda}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\delta(x_n-k) = -\frac{1}{\lambda}\sum_{n=0}^{N-1}1 = -\frac{N}{\lambda} = 1 \quad (8)$$

So:

$$\lambda = -N \quad (9)$$

and we get:

$$\hat{\pi_k} = \frac{N_k}{N} \quad (10)$$

(2) For μ_k, we calculate the derivative of equation 4 w.r.t. μ_k and make it as 0:

$$\frac{\partial \log p(y,x|\theta)}{\partial \mu_k} = R_k^{-1}t_{1,k} - N_k R_k^{-1}\mu_k = 0 \quad (11)$$

We can get:

$$\hat{\mu_k} = \frac{t_{1,k}}{N_k} \quad (12)$$

(3) For R_k, do it in a similar way as μ_k:

$$\frac{\partial \log p(y,x|\theta)}{\partial R_k^{-1}} = \frac{1}{2}N_k|R_k||R_k^{-1}|R_k - \frac{1}{2}t_{2,k} + t_{1,k}\mu_k^t - \frac{1}{2}N_k\mu_k^t\mu_k = 0 \quad (13)$$

substitute equation 12:

$$N_k\hat{R_k} - t_{2,k} + \frac{2t_{1,k}t_{1,k}^t}{N_k} - N_k\frac{t_{1,k}t_{1,k}^t}{N_k^2} = 0 \quad (14)$$

$$\hat{R_k} = \frac{t_{2,k}}{N_k} - \frac{t_{1,k}t_{1,k}^t}{N_k^2} \quad (15)$$

3 Parameter Estimation for Gaussian Mixture Distributions

3.1 Use the Matlab program *mk_data.m* to create 500 samples from a Gaussian mixture with M=2, K=3, π=[0.4,0.4,0.2].

N.A.

3.2 Print out a scatter plot of the samples generated in step.1. Circle and label each of the three clusters in the mixture distribution.

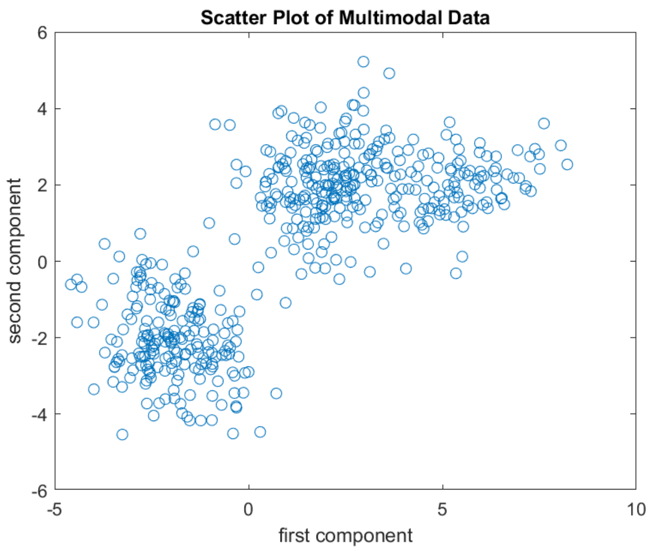

Figure 1: A scatter plot of the samples generated in step.1

3.3 Derive an explicit expression for $P\{X_n = k|Y = y, \hat{\theta}\}$sed in the E-step

From the Bayes' Rule:

$$P\{X|Y,\theta\} = \frac{P\{Y|X,\theta\}P\{X|\theta\}}{P\{Y|\theta\}} \tag{16}$$

Since $Y \sim N(\mu_{X_n}, R_{X_n})$, we know that:

$$P\{X = k|\theta\} = \pi_k \tag{17}$$

$$P\{Y|X = k, \theta\} = \frac{1}{(2\pi)^{p/2}}|R_k|^{-1/2} \exp\left\{-\frac{1}{2}(y-\mu_k)^t R_k^{-1}(y-\mu_k)\right\} \tag{18}$$

$$P\{Y|\theta\} = \sum_{k=0}^{K-1} \frac{\pi_k}{(2\pi)^{p/2}}|R_k|^{-1/2} \exp\left\{-\frac{1}{2}(y-\mu_k)^t R_k^{-1}(y-\mu_k)\right\} \tag{19}$$

Substitute the equation 17, 18 and 19 to equation 16, we can get:

$$P\{X_n = k|Y = y, \theta\} = \frac{\frac{1}{(2\pi)^{p/2}}|R_k|^{-1/2} \exp\left\{-\frac{1}{2}(y-\mu_k)^t R_k^{-1}(y-\mu_k)\right\}\pi_k}{\sum_{j=0}^{K-1} \frac{1}{(2\pi)^{p/2}}|R_j|^{-1/2} \exp\left\{-\frac{1}{2}(y-\mu_j)^t R_j^{-1}(y-\mu_j)\right\}\pi_j} \tag{20}$$

3.4 Implement the EM algorithm for computing the ML estimate of θ.

N.A. See the code.

3.5 Run 20 iterations of the EM algorithm, and print out the values of the estimated parameters. Do the indices k of the estimated and true parameters correspond?

Estimated parameters are:

$$\hat{\pi} = [0.4092, 0.3942, 0.1965], \tag{21}$$

$$\hat{\mu_0} = [1.9689, 2.0575]^t$$
$$\hat{\mu_1} = [-1.9689, -2.0893]^t \quad (22)$$
$$\hat{\mu_2} = [5.4083, 1.9470]^t,$$

and,

$$\hat{R_0} = \begin{bmatrix} 1.0924 & 0.2020 \\ 0.2020 & 1.0715 \end{bmatrix}$$
$$\hat{R_1} = \begin{bmatrix} 1.0351 & -0.2926 \\ -0.2926 & 1.1199 \end{bmatrix} \quad (23)$$
$$\hat{R_2} = \begin{bmatrix} 1.3284 & 0.3197 \\ 0.3197 & 0.5443 \end{bmatrix}$$

In this case, the indices k of the estimated and true parameters correspond. But actually it doesn't have to. If you run the random and algorithm several times, you will find they don't correspond some time.

3.6 Find the best correspondence of the true and estimated parameters, and print them out in a tabular form comparing their values.

The comparison of the estimated and true parameters is shown in Table. 1.

Table 1: The best correspondence

Clusters	Parameters	Estimated	True
1	π_1	0.4092	0.4
	μ_1	[1.9689, 2.0575]	[2, 2]
	R_1	$\begin{bmatrix} 1.0924 & 0.2020 \\ 0.2020 & 1.0715 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}$
2	π_2	0.3942	0.4
	μ_2	[-1.9689, -2.0893]	[-2, -2]
	R_2	$\begin{bmatrix} 1.0351 & -0.2926 \\ -0.2926 & 1.1199 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.1 \\ -0.1 & 1 \end{bmatrix}$
3	π_3	0.1965	0.2
	μ_3	[5.4083, 1.9470]	[5.5, 2]
	R_3	$\begin{bmatrix} 1.3284 & 0.3197 \\ 0.3197 & 0.5443 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}$

4 Order Identification for Gaussian Mixture Distributions

4.1 Use the data generated in problem 1 of the previous Section 3

N.A.

4.2 Implement the MDL order estimation method as described above using the initial value of K = 9.

N.A. See the code.

4.3 Implement a matlab subroutine that computes the MDL

N.A. see the code.

4.4 Plot the value of MDL versus each EM iteration

See Figure. 2.

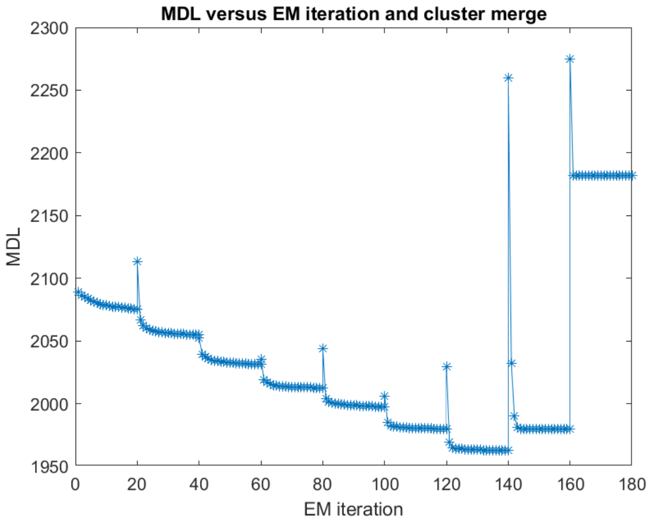

Figure 2: The value of MDL versus each EM iteration. It also includes the cluster merge.

4.5 Plot the MDL value versus K

See Figure. 3.

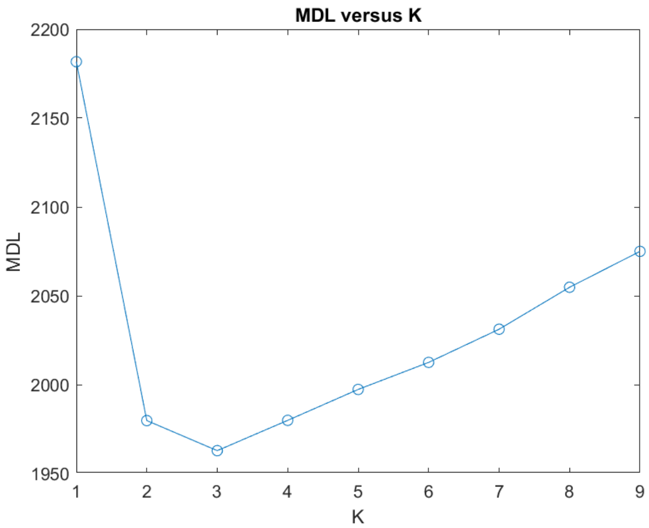

Figure 3: the MDL value versus K.

4.6 Does the estimated value of K correspond to the true value of K = 3?

Yes, from Figure. 3, we can clearly see that when K = 3, the MDL reach the minimum. So the estimated K will be 3 which is exactly same as the true value.

5 Appendix

The source code: `https://github.com/chenliming0422/ECE641-Lab3`