Writing Exercise

Bochen Wang, Tianyuan Liu

November 11, 2015

Exercise 14.2

1

The log-likelihood of data

$$\begin{aligned}
\prod_{i=1}^{N} g(x_i) &= \prod_{i=1}^{N}(\sum_{k=1}^{K} \pi_k g_k(x_i)) \\
l(\theta; Z) &= \log \prod_{i=1}^{N} g(x_i) \\
&= \sum_{i=1}^{N} \log(\sum_{k=1}^{K} \pi_k g_k(x_i))
\end{aligned}$$

2

Maximize the log-likelihood above using EM algorithm
Pretend we know the value of latent variable Δ for each point:

$$l(\theta; Z, \Delta) = \sum_{i=1}^{N} \Delta_{k,i} \log g_k(x_i; \mu_k, \sigma_k^2) + \sum_{i=1}^{N} \Delta_{k,i} \log \pi_k$$

E-step
For k = 1 to K, i = 1 to N

$$\begin{aligned}
T_{k,i} &= E(\Delta_{k,i} | \theta, Z) \\
&= \frac{\pi_k g_k(x_i; \mu_k, \sigma_k^2)}{\sum\limits_{k=1}^{K} \pi_k g_k(x_i; \mu_k, \sigma_k^2)}
\end{aligned}$$

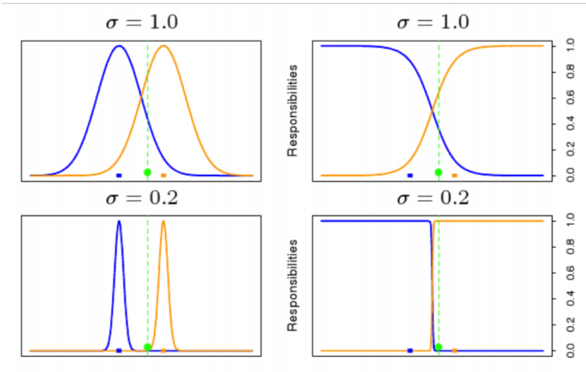
M-step
For k = 1 to K

$$\mu_k = \frac{\sum_{i=1}^{N} T_{k,i} x_i}{\sum_{i=1}^{N} T_{k,i}}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{N} T_{k,i}(x_i - \mu_k)}{\sum_{i=1}^{N} T_{k,i}}$$

$$\pi_k = \frac{\sum_{i=1}^{N} T_{k,i}}{N}$$

3



As the figure above shows, when $\sigma \to 0$, the responsibility curve is very steep. There is almost no soft assign interval. So the data point will no longer spreads to each mode, it will be assign to a specific mode. So that when $\sigma \to 0$, EM algorithm turns into K-Means.

Exercise 14.8

For the Procrustes transformation we want to evaluate

$$\underset{\mu,R}{arg\,min} \|X_2 - (X_1 R + \mathbf{1}\mu^T)\|_F$$

R is the rotation matrix, μ is the translation vector.
It has the same solution as

$$\underset{\mu,R}{arg\,min} \|X_2 - (X_1 R + \mathbf{1}\mu^T)\|_F^2$$

$$\because \|X\|_F^2 = trace(X^T X)$$

So the function we want to evaluate can be written as

$$\underset{\mu,R}{arg\,min} [trace((X_2 - X_1 R - \mathbf{1}\mu^T)^T (X_2 - X_1 R - \mathbf{1}\mu^T))]$$

$$
\begin{aligned}
(X_2 - X_1R - \mathbf{1}\mu^T)^T(X_2 - X_1R - \mathbf{1}\mu^T) &= ((X_2 - X_1R)^T - \mu\mathbf{1}^T)((X_2 - X_1R) - \mathbf{1}\mu^T) \\
&= (X_2 - X_1R)^T(X_2 - X_1R) \\
&- (X_2 - X_1R)^T\mathbf{1}\mu^T - \mu\mathbf{1}^T(X_2 - X_1R) \\
&+ \mu\mathbf{1}^T\mathbf{1}\mu^T \\
\because \mathbf{1}^T\mathbf{1} &= N \\
\mathbf{1}^T X_1 &= N\bar{x}_1^T \\
\mathbf{1}^T X_2 &= N\bar{x}_2^T
\end{aligned}
$$

Minimize the trace above by taking the μ derivative and set the derivative to zero.

$$
\begin{aligned}
-2(X_2 - X_1R)^T\mathbf{1} + 2N\mu &= 0 \\
\mu &= \frac{1}{N}(X_2^T - R^T X_1^T)\mathbf{1} \\
&= \bar{x}_2 - R^T\bar{x}_1
\end{aligned}
$$

With the μ we calculate, we can simplify the trace.

$$
\begin{aligned}
X_2 - X_1R - \mathbf{1}\mu^T &= X_2 - X_1R - \mathbf{1}(\bar{x}_2^T - \bar{x}_1^T R) \\
&= X_2 - \mathbf{1}\bar{x}_2^T - (X_1 - \mathbf{1}\bar{x}_1^T)R \\
&= \tilde{X}_2 - \tilde{X}_1 R
\end{aligned}
$$

Original problem transform into

$$
\underset{R}{arg\,min}[trace((\tilde{X}_2 - \tilde{X}_1R)^T(\tilde{X}_2 - \tilde{X}_1R))]
$$

To solve this, we have $RR^T = R^T R = I$, so we define a Lagrangean function as below.

$$
\begin{aligned}
\text{Let } E &= \tilde{X}_2 - \tilde{X}_1R \\
F &= trace(E^T E) + trace(L(R^T R - I)) \\
\frac{\partial F}{\partial R} &= -2\tilde{X}_1^T\tilde{X}_2 + 2\tilde{X}_1^T\tilde{X}_1R + R(L + L^T) \\
\text{Let } \frac{\partial F}{\partial R} &= 0
\end{aligned}
$$

We get the equation below.

$$
\frac{L + L^T}{2} = -R^T\tilde{X}_1^T\tilde{X}_2 + R^T\tilde{X}_1^T\tilde{X}_1R
$$

Because $\frac{L+L^T}{2}$ is symmetric and $R^T\tilde{X}_1^T\tilde{X}_1 R$ is symmetric too, $R^T\tilde{X}_1^T\tilde{X}_2$ must be symmetric.

$$\begin{aligned}
\text{Let } S &= \tilde{X}_1^T\tilde{X}_2 \\
\text{Has } R^T S &= S^T R \\
S &= RS^T R \\
\text{Also } S^T &= R^T S R^T \\
SS^T &= RS^T S R^T
\end{aligned}$$

Implement the SVD on S.

$$\begin{aligned}
S &= UDV^T \\
SS^T &= UDV^T VDU^T \\
&= UD^2U^T \\
S^T S &= VDU^T UDV^T \\
&= VD^2V^T \\
\because SS^T &= RS^T S R^T \\
\therefore UD^2U^T &= RVD^2V^T R^T \\
\therefore R &= UV^T
\end{aligned}$$

When we add the scaling parameter. We want to evaluate

$$\underset{\beta,R}{arg\,min}\|X_2 - \beta X_1 R\|_F$$

Familiar with the result above, we got $R = UV^T$.
Now we need to get β.

$$\begin{aligned}
\frac{\partial F}{\partial \beta} &= 2\beta \text{ trace}(R^T X_1^T X_1 R) - 2 \text{ trace}(X_2^T X_1 R) \\
\text{Let } \frac{\partial F}{\partial \beta} &= 0 \\
\beta &= \frac{trace(X_2^T X_1 R)}{trace(R^T X_1^T X_1 R)} \\
&= \frac{trace(X_2^T X_1 R)}{trace(X_1^T X_1 R R^T)} \\
&= \frac{trace(R^T X_1^T X_2)}{trace(X_1^T X_1)} \\
&= \frac{trace(VU^T UDV^T)}{\|X_1\|_F^2} \\
&= \frac{trace(U^T UDV^T V)}{\|X_1\|_F^2} \\
&= \frac{trace(D)}{\|X_1\|_F^2}
\end{aligned}$$

Exercise 14.11

In the multiscaling problem, we have the square distance matrix and we want to reconstruct the observation matrix X. In classical multiscaling problem we get the centered inner product matrix, and we want to reconstruct observation matrix X.
Let X denotes the original observation data matrix with $N \times p$ dimension.
Let X_c denotes the centered data matrix with subtracted column means.
We have

$$\begin{aligned} X_c &= (I - \frac{\mathbf{1}\mathbf{1}^T}{n})X \\ \text{Let } S &= X_c X_c^T \end{aligned}$$

So S is the centered inner product matrix with elements $< x_i - \tilde{x}, x_j - \tilde{x} >$.
Now we need to use square distance matrix d^2 to get S.

$$\begin{aligned} d_{i,j}^2 &= \|x_i - x_j\|^2 \\ &= \|(x_i - \tilde{x}) - (x_j - \tilde{x})\|^2 \\ &= \|x_i - \tilde{x}\|^2 + \|x_j - \tilde{x}\|^2 - 2 < x_i - \tilde{x}, x_j - \tilde{x} > \\ &= \|x_i - \tilde{x}\|^2 + \|x_j - \tilde{x}\|^2 - 2S_{i,j} \\ \therefore S &= -\frac{1}{2}(I - \frac{\mathbf{1}\mathbf{1}^T}{n})d^2(I - \frac{\mathbf{1}\mathbf{1}^T}{n}) \\ &= X_c X_c^T \end{aligned}$$

Now we do the eigendecomposition on matrix S.

$$\text{Let } S = E\Lambda E^T$$

Λ is a diagonal matrix contains all the eigenvalues of S. The column of E is the associated eigenvectors.

$$\begin{aligned} X_c X_c^T &= E\Lambda E^T \\ X_c &= E\Lambda^{\frac{1}{2}} \end{aligned}$$

If we only use the largest k eigenvalues and the associated eigenvectors. We get

$$\tilde{X}_c = E_k D_k$$

Apparently, the solution z_i is the rows of $\tilde{X}_c$. That is what we want to prove.