Expectation and Maximization

1. $S = \{\mathbf{x}_i \in R^d\}_{i=1}^N$. In this problem, the widely used expectation-maximization(EM) method is derived to partition high-dimensional points in S to K clusters, where $K << N$. Let $\mathbf{y}_k$ denote a center and $\boldsymbol{\xi}_i = (\xi_{i1}, ..., \xi_{iK})^T$ denote the exclusive membership of $\mathbf{x}_i$ to K clusters, where

$$\begin{aligned} \sum_{k=1}^{K} \xi_{ik} &= 1, \text{ for all } i, \\ \xi_{ik} &\in \{0, 1\}, \text{ for all } i, k. \end{aligned}$$

$\boldsymbol{\xi}_i$ exactly belongs the following set of standard unitary vectors

$$\Xi^K = \{\mathbf{e}_1^K, ..., \mathbf{e}_K^K\},$$

where $\mathbf{e}_k^K$ is a unitary vector with the kth bit one and the others zero. That $\boldsymbol{\xi}_i = \mathbf{e}_k^K$ means $\mathbf{x}_i$ belongs the kth cluster. The EM method is derived to minimize the following objective function

$$E(\{\xi_{ik}\}, \{\mathbf{y}_k\}) = \sum_{i=1}^{N}\sum_{k=1}^{K} \xi_{ik} ||\mathbf{x}_i - \mathbf{y}_k||^2.$$

Since E simultaneously contains discrete and continuous variables, it can not be optimized by the gradient descent method.

(a) For fixed $\{\mathbf{y}_k\}$, each membership vector $\boldsymbol{\xi}_i$ is considered as a random vector characterized by the following Boltzmann distribution,

$$\Pr(\boldsymbol{\xi}_i = \mathbf{e}_k^K) \propto \exp(-\beta ||\mathbf{x}_i - \mathbf{y}_k||^2)$$

where β denotes the inverse of a temperature-like parameter. For valid probabilities, we must have

$$\sum_{k=1}^{K} \Pr(\boldsymbol{\xi}_i = \mathbf{e}_k^K) = 1.$$

Express the exact form of $\Pr(\boldsymbol{\xi}_i = \mathbf{e}_k^K)$ according to the Boltzmann distribution and the above unitary condition.

(b) Let

$$\langle \boldsymbol{\xi}_i \rangle = (\langle \xi_{i1} \rangle, ..., \langle \xi_{iK} \rangle)^T,$$

where

$$\langle \xi_{ik} \rangle = \Pr(\boldsymbol{\xi}_i = \mathbf{e}_k^K).$$

Express the exact form of the expectation $\langle \boldsymbol{\xi}_i \rangle$.

(c) Solve the following equation to derive the updating rule for each $\mathbf{y}_k$,

$$\frac{\partial E(\{\langle \xi_{ik} \rangle\}, \{\mathbf{y}_k\})}{\partial \mathbf{y}_k} = 0.$$

(d) The solutions in the last two questions characterize the EM method for clustering analysis. Give a procedure to realize the derived EM method.

(e) Implement the EM method by the MATLAB language.

(f) Give examples to test your program.