IERG4300 Fall Tutorial 10

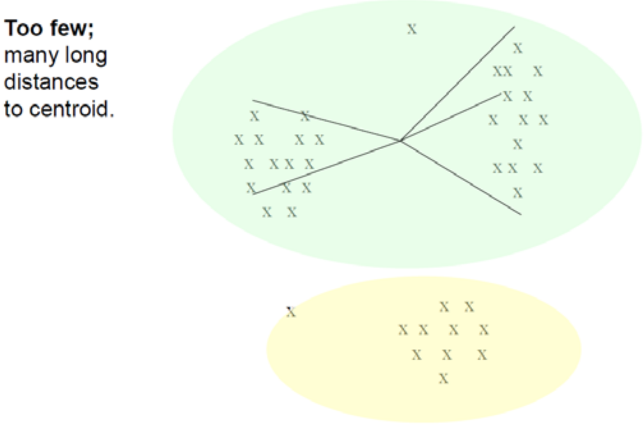
Hongyu DENG
Department of Information Engineering
The Chinese University of Hong Kong

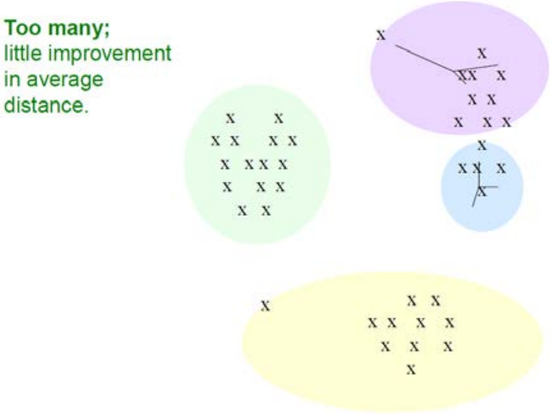
Outline

- k-means Clustering
- BMM
- Q&A

k-means clustering

- How to set parameter k?
  - We know that the value of k will affect the performance of clustering.

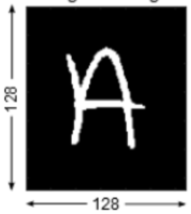




k-means clustering

- How to set parameter k?
  - Here is an example about a 26 letters dataset.
    - It includes 26 letters and every letter has uppercase and lowercase.
    - How to set the value of k?
    - Considering some letters have similar appearance of uppercase and lowercase, we can set the value of k between 26 to 52.

Original Image



(a)

k-means clustering

- How to set parameter k?
  - Need to determine “k” via domain knowledge or heuristics (as stated before).
  - For example, in some domains, such as array signal processing, we can utilize blind signal analysis to estimate the number of signal sources(k).
- Loss function

$$loss = \sum_{i=1}^{n} d(x_i, \mu_k)^2$$

k-means clustering

- Algorithm

Input:

- k: the number of clusters,
- D: a data set containing n objects.

Output: A set of k clusters.

Method:

(1) arbitrarily choose k objects from D as the initial cluster centers;
(2) **repeat**
(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4) update the cluster means, that is, calculate the mean value of the objects for each cluster;
(5) **until** no change;

Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.

BMM

- Bernoulli Mixture Model is more suitable in processing discrete data, especially for Binary form information.
    - First we introduce Bernoulli distribution
        - Two-dimensional Bernoulli distribution
        - Multi-dimensional Bernoulli distribution
        - Multivariate Bernoulli distribution
    - Then we introduce Bernoulli Mixture Model
        - Mixture Model for Two-dimensional Bernoulli distribution

BMM

- Bernoulli distribution
  - Two-dimensional Bernoulli distribution
    - The probability distribution function can be written as follow

$$P(\mathbf{x} \mid \mathbf{q}) = q_1^{x_1} q_2^{x_2} = q_1^{x_1}(1-q_1)^{(1-x_1)}$$

$Sample1:[0,1]$

$Sample2:[1,0]$

- Here, $\mathbf{x} = [x_1, x_2]$ is a 2-D Boolean vector, and $x_1 + x_2 = 1$

$$x_d \in \{0,1\}$$

- Besides, $\mathbf{q} = [q_1, q_2]$ is the Probability of $x_1 = 1$ and $x_2 = 1$

$$q_d \in [0,1]$$

BMM

- Bernoulli distribution
  - Multi-dimensional Bernoulli distribution
    - From 2-D to Multi-D
    - The probability distribution function can be written as follow

$$P(\mathbf{x} \mid \mathbf{q}) = \prod_{d=1}^{D} q_d^{x_d}$$

$$\downarrow$$

$$\mathbf{x} = [x_1, x_2, \cdots, x_D] \quad x_d \in \{0,1\} \quad \sum_{d=1}^{D} x_d = 1$$

$$\mathbf{q} = [q_1, q_2, \cdots, q_D] \quad q_d \in [0,1] \quad \sum_{d=1}^{D} q_d = 1$$

$Sample1 : [0,1,0,0]$

$Sample2 : [0,0,0,1]$

$\vdots$

$SampleN : [0,0,1,0]$

BMM

- Bernoulli distribution
  - Multivariate Bernoulli distribution
    - From Multi-D to Multivariate
    - The probability distribution function can be written as follow

$$P(\mathbf{x}|\mathbf{q}) = \prod_{d=1}^{D} q_d^{x_d}(1-q_d)^{(1-x_d)}$$

$$\mathbf{x} = [x_1, x_2, \cdots, x_D] \qquad x_d \in \{0,1\} \qquad \cancel{\sum_{d=1}^{D} x_d = 1}$$

$$\mathbf{q} = [q_1, q_2, \cdots, q_D] \qquad q_d \in [0,1] \qquad \cancel{\sum_{d=1}^{D} q_d = 1}$$

$Sample1 : [0,1,1,0]$

$Sample2 : [1,0,0,1]$

$\vdots$

$SampleN : [0,1,1,1]$

BMM

- Bernoulli Mixture Model
  - Basic background about Mixture Model
    - First, we define the general formula of the mixture model composed of **K** sub-distributions in the **D**-dimensional space

$$P(\mathbf{x} \mid \mathbf{Q}, \pi) = \sum_{k=1}^{K} \pi_k \boxed{P(\mathbf{x} \mid \mathbf{q}_k)}$$

It can be Bernoulli, Gaussian etc.

Here X is a dataset with N samples, each sample is independently identically distributed.

$$\boxed{\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]}$$

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_K]$$

$$\sum_{k=1}^{K} \pi_k = 1$$

We can consider it as weight.

BMM

- Bernoulli Mixture Model
  - Basic background about Mixture Model
    - First, we define the general formula of the mixture model composed of **K** sub-distributions in the **D**-dimensional space

$$P(\mathbf{x} \mid \mathbf{Q}, \pi) = \sum_{k=1}^{K} \pi_k P(\mathbf{x} \mid \mathbf{q}_k)$$

    - We always build the Maximum likelihood Estimation to estimate the parameter q. So the above equation can be rewritten as

$$L(\mathbf{X} \mid \mathbf{Q}, \boldsymbol{\pi}) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k P(\mathbf{x}_n \mid \mathbf{q}_k)$$

BMM

- Bernoulli Mixture Model
  - Basic background about Mixture Model
    - In order to maximize the likelihood function, we transfer it as a log-likelihood as follow

$$
\begin{aligned}
\log L(\mathbf{X} \mid \mathbf{Q}, \boldsymbol{\pi}) &= \log \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k P(x_n \mid q_k) \\
&= \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k P(x_n \mid q_k)
\end{aligned}
$$

BMM

- Bernoulli Mixture Model
  - Basic background about Mixture Model
    - Then we derive the posterior probability $\gamma(z_{nk})$ based on Bayes theorem

$$\boxed{\gamma(z_{nk})} = \boxed{P(z_k = 1 \mid \mathbf{x}_n)}$$

$$= \frac{P(\mathbf{x}_n \mid z_k = 1)P(z_k = 1)}{\sum_{k=1}^{K} P(\mathbf{x}_n \mid z_k = 1)P(z_k = 1)}$$

$$= \frac{\pi_k P(\mathbf{x}_n \mid \mathbf{q}_k)}{\sum_{k=1}^{K} \pi_k P(\mathbf{x}_n \mid \mathbf{q}_k)}$$

Probability

We have known the sample $\mathbf{x}_n$, and the probability of this sample comes from $k-th$ sub-distributions $z_k = 1$.

BMM

- Bernoulli Mixture Model
  - Mixture Model for Two-dimensional Bernoulli distribution
    - Taking derivative the above expression with respect to $q_{k,1}$

$$\frac{\partial}{\partial q_{k,1}}\left(\log L(\mathbf{X}\,|\,\mathbf{Q},\boldsymbol{\pi})\right)=\sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial q_{k,1}}\log\left(q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}\right)$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\left(\frac{x_{n,1}}{q_{k,1}}-\frac{1-x_{n,1}}{1-q_{k,1}}\right)$$

$$=0$$

- Next page is the detailed mathematical process

$$\frac{\partial}{\partial q_{k,1}}\left(\log L(\mathbf{X}\,|\,\mathbf{Q},\boldsymbol{\pi})\right)=\sum_{n=1}^{N}\frac{\partial}{\partial q_{k,1}}\log\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}$$

$$\boxed{=\frac{\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}\times\frac{\frac{\partial}{\partial q_{k,1}}\sum_{k=1}^{K}q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}}$$

Chain rule: $\frac{dw}{dx}=\frac{dw}{dy}\bullet\frac{dy}{dx}$

Hints: derivative of logarithmic function

$$\frac{d\log_\alpha|x|}{dx}=\frac{1}{x\ln\alpha}$$

$$=\frac{\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}\times\frac{\frac{\partial}{\partial q_{k,1}}q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial q_{k,1}}\log\left(q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}\right)$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial q_{k,1}}\left(x_{n,1}\log q_{k,1}+(1-x_{n,1})\log(1-q_{k,1})\right)$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\left(\frac{x_{n,1}}{q_{k,1}}-\frac{1-x_{n,1}}{1-q_{k,1}}\right)$$

$$=0$$

$$\frac{\partial}{\partial q_{k,1}}\left(\log L(\mathbf{X}\,|\,\mathbf{Q},\boldsymbol{\pi})\right)=\sum_{n=1}^{N}\frac{\partial}{\partial q_{k,1}}\log\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}$$

$$=\frac{\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}\times\frac{\boxed{\frac{\partial}{\partial q_{k,1}}\sum_{k=1}^{K}q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}}{q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}$$

$$=\frac{\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}\times\frac{\frac{\partial}{\partial q_{k,1}}q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}$$

Hints: All items are constant Except the $k-th$ item.

$$\frac{df(x_k)}{dx_k}=\frac{d\left(f(x_1)+f(x_2)+\cdots+f(x_k)+\cdots+f(x_K)\right)}{dx_k}$$
$$=\frac{df(x_1)}{dx_k}+\frac{df(x_2)}{dx_k}+\cdots+\frac{df(x_k)}{dx_k}+\cdots+\frac{df(x_K)}{dx_k}$$
$$=\underbrace{\frac{df(x_1)}{dx_k}+\frac{df(x_2)}{dx_k}+\cdots}_{0}+\frac{df(x_k)}{dx_k}+\underbrace{\cdots+\frac{df(x_K)}{dx_k}}_{0}$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial q_{k,1}}\log\left(q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}\right)$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial q_{k,1}}\left(x_{n,1}\log q_{k,1}+(1-x_{n,1})\log(1-q_{k,1})\right)$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\left(\frac{x_{n,1}}{q_{k,1}}-\frac{1-x_{n,1}}{1-q_{k,1}}\right)$$

$$=0$$

$$\frac{\partial}{\partial q_{k,1}}\left(\log L(\mathbf{X}\,|\,\mathbf{Q},\boldsymbol{\pi})\right)=\sum_{n=1}^{N}\frac{\partial}{\partial q_{k,1}}\log\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}$$

$$=\frac{\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}\times\frac{\frac{\partial}{\partial q_{k,1}}\sum_{k=1}^{K}q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}$$

$$=\frac{\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{\sum_{k=1}^{K}\pi_k q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}\times\boxed{\frac{\frac{\partial}{\partial q_{k,1}}q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}{q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}}}$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\boxed{\frac{\partial}{\partial q_{k,1}}\log\left(q_{k,1}^{x_{n,1}}(1-q_{k,1})^{(1-x_{n,1})}\right)}$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial q_{k,1}}\left(x_{n,1}\log q_{k,1}+(1-x_{n,1})\log(1-q_{k,1})\right)$$

$$=\sum_{n=1}^{N}\gamma(z_{nk})\left(\frac{x_{n,1}}{q_{k,1}}-\frac{1-x_{n,1}}{1-q_{k,1}}\right)$$

$$=0$$

Hints: derivative of logarithmic function

$$\frac{d\log_{\alpha}|x|}{dx}=\frac{1}{x\ln\alpha}$$

$$\Longleftarrow$$

BMM

- Bernoulli Mixture Model
  - Mixture Model for Two-dimensional Bernoulli distribution
    - Finally, we can obtain the result of $q_{k,1}$ and $q_{k,2}$

$$q_{k,1} = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_{n,1}}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$q_{k,2} = 1 - q_{k,1}$$

Ref:
[Ch9.4-MixturesofBernoulli.ppt (buffalo.edu)]
[chapter4.pdf]

Q&A for Homework 3