Ch9, Mixture Gaussian and EM

January 25, 2019

1 9.1. K-means Clustering

K-means clustering ( ) . 1. ($\boldsymbol{\mu}_k \in \mathbb{R}^D$, prototype vector) . 2. n ($\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$) . 3. k .
4. n '2 ' .
2 EM algorithm M(maximization) ,
3 E(expectation) .
.

1.0.1 9.1.1. Image segmentation and compression

skip

2 9.2. Mixtures of Gaussians

- 'Mixtures of Gaussian' K gaussian($\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$) mixing coefficient π_k .(2.3.9)

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad (9.7)$$

- mixing coefficient π_k 0 , 1 .

$$0 \leq \pi_k \leq 1 \tag{1}$$

$$\sum_{k=1}^{K} \pi_k = 1 \qquad (9.9) \tag{2}$$

- k 1 π_k 1-of-K K $\mathbf{z}$.
  (K $(z_1, \ldots, z_k, \ldots, z_K)$ 1 0.)
- mixing coefficient $\pi_1, \ldots, \pi_k, \ldots, \pi_K$ $\mathbf{z}$.

$$z_k \in \{0, 1\} \tag{3}$$

$$\sum_k z_k = 1 \tag{4}$$

$$p(z_k = 1) = \pi_k \tag{5}$$

- $\mathbf{z}$ '1-of-K' $p(\mathbf{z})$.

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k} \qquad (9.10)$$

- $p(\mathbf{x})$ 'mixtures of gaussian' k gaussian '$z_k = 1$' $p(\mathbf{x})$, $p(\mathbf{x} \mid z_k = 1)$.

$$p(\mathbf{x} \mid z_k = 1) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- $\mathbf{z}$ '1-of-K' $p(\mathbf{z})$.

$$p(\mathbf{x} \mid \mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} \qquad (9.11)$$

- $p(\mathbf{z})$ $p(\mathbf{x} \mid \mathbf{z})$ 'mixtures of gaussian', $p(\mathbf{x})$.

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z}) p(\mathbf{x} \mid \mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad (9.12)$$

- figure 9.4 $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{z}) p(\mathbf{x} \mid \mathbf{z})$ graphical model .
- , $p(\mathbf{x})$ $p(\mathbf{x}, \mathbf{z})$, .
- '$\mathbf{x}$ $\mathbf{z}$ $\gamma(z_k)$.

$$\gamma(z_k) \equiv p(z_k = 1 \mid \mathbf{x}) = \frac{\pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \qquad (9.13)$$

- π_k $z_k = 1$ '' , $p(z_k = 1 \mid \mathbf{x})$ '' .
- $\gamma(z_k)$ **responsibility** .

2.0.1 9.2.1. Maximum likelihood

- feature D n $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$.($N \times D$ $\mathbf{X}$)
- , N 'mixtures of gaussian' .@@@ (figure 9.6 graphical model . figure 9.4 .) N likelihood .
- likehood for the incomplete data set $\{\mathbf{X}\}$

$$\begin{aligned} p(\mathbf{X} \mid \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{n=1}^{N} p(\mathbf{x}_n \mid \boldsymbol{\pi}_n, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) && (6) \\ &= \prod_{n=1}^{N} \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{X}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} && (7) \\ & && (8) \end{aligned}$$

- log likehood for the incomplete data set $\{\mathbf{X}\}$, eq 9.14

$$\ln p(\mathbf{X} \mid \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n) \tag{9}$$

$$= \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{X}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \qquad (9.14) \tag{10}$$

$$\tag{11}$$

- **X** likelihood $(\{\pi_1, \ldots, \pi_N\}, \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k\}, \{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K\}$.
- 'gaussian mixture' maximum likelihood (singularity) . (p433, p434)
- $\ln \sum$. Gaussian mixture .

2.0.2 9.2.2. EM for Gaussian mixtures

- (9.14) , $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{\boldsymbol{k}}, \pi_k$ 0 .

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \qquad (9.17) \tag{12}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathsf{T}} \qquad (9.19) \tag{13}$$

$$\pi_k = \frac{N_k}{N} \qquad (9.22) \tag{14}$$

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \tag{15}$$

$$\gamma(z_k) \equiv p(z_k = 1 \mid \mathbf{x}) = \frac{\pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \qquad (9.13) \tag{16}$$

- .
  π_k $\boldsymbol{\mu}_k$ $\boldsymbol{\Sigma}_k$
  .
  .

  .
  EM .

-

1. $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{\boldsymbol{k}}, \pi_k$

- (E) $\gamma(z_k)$, figure 9.8(b) $(\mathbf{x_n})$ $\gamma(z_n k)$, $\gamma(z_1)$ / $\gamma(z_2)$
- (M) $\gamma(z_n k)$ (9.17), (9.19), (9.22) $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{\boldsymbol{k}}, \pi_k$
- '2' (likelihood(9.14))

3 9.3. An Alternative View of EM

- $\mathbf{X} \in \mathbb{R}^{N\times D}$ $\mathbf{Z} \in \mathbb{R}^{N\times K}$ $\mathbf{X}$ log likelihood .

$$\ln p(\mathbf{X} \mid \boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) \right\} \qquad (9.29)$$

- $\ln \sum$ $\ln p(\mathbf{X} \mid \boldsymbol{\theta})$.
  $\ln p(\mathbf{X} \mid \boldsymbol{\theta})$ $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big]$. ($\mathbf{Z} \sim p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}})$)
  $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big]$ $\ln p(\mathbf{X} \mid \boldsymbol{\theta})$, 9.4 .

- $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big]$ $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$.

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})$$

- 'E ' $\boldsymbol{\theta}^{\text{old}}$ $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$
  'M ' $\boldsymbol{\theta}^{\text{old}}$ $\boldsymbol{\theta}^{\text{new}}$.

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \qquad (17)$$

$$\boldsymbol{\theta}^{\text{old}} \leftarrow \boldsymbol{\theta}^{\text{new}} \qquad (18)$$

$$(19)$$

EM $\boldsymbol{\theta}$.

- EL MAP(maximum posterior) ,
  $\boldsymbol{\theta}$ $p(\boldsymbol{\theta})$,
  $\boldsymbol{\theta}$.
  (E $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$, M .)

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) + \ln p(\boldsymbol{\theta})$$

$p(\boldsymbol{\theta})$ MLE (singularity) .?

3.0.1 9.3.1. Gaussian mixtures revisited

- $\mathbf{X}$ 'incomplete data set' $\{\mathbf{X}\}$
  $(\mathbf{x}_n)$ $(\mathbf{z}_n)$ $\mathbf{Z}$ 'complete data set' $\{\mathbf{X}, \mathbf{Z}\}$.

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k} \qquad (9.10) \qquad (20)$$

$$p(\mathbf{x} \mid \mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} \qquad (9.11) \qquad (21)$$

- complete data set $\{\mathbf{X}, \mathbf{Z}\}$ log likelihood .

$$p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{nk}} \qquad (9.35) \tag{22}$$

$$\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \} \qquad (9.36) \tag{23}$$

$$\tag{24}$$

$\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})$ ln gaussian .
$\ln p(\mathbf{X} \mid \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ (9.14) $\ln \sum \mathcal{N}$.

$$\ln p(\mathbf{X} \mid \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n) \tag{25}$$

$$= \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{X}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \qquad (9.14) \tag{26}$$

$$\tag{27}$$

- $\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})$ MLE .
  $\mathbf{Z}$.
  $\mathbf{Z}$ $p(\mathbf{Z})$
  $\mathbf{X}$ $\mathbf{Z}$ $p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta})$ $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big]$ $\boldsymbol{\theta}$.
  ($\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big]$ $\ln p(\mathbf{X} \mid \boldsymbol{\theta})$, 9.4 .)
- gaussian mixture $\mathbf{X}$ $\mathbf{Z}$ $p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta})$ (9.38) ,
  n k gaussian z_{nk} responsibility $\gamma(z_{nk})$.

$$\mathbb{E}[z_{nk}] = \gamma(z_{nk}) \qquad (9.39)$$

  gaussian mixture $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big]$.
  ($\boldsymbol{\theta}$ $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$, $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$. $\boldsymbol{\theta}^{\text{old}}$ $\gamma(z_{nk})$ $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$.)

$$\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\big] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \} \qquad (9.40)$$

-
  gaussian mixture maximum likelihood
  $\ln p(\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})$ $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\big]$ $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$ EM .

1. $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$

- (E) $\gamma(z_{nk})$
- (M) $\mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\big]$ $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$

- 2

3.0.2 9.3.2. Relation to K-means

skip

3.0.3 9.3.3. Mixtures of Bernoulli distributions

skip

3.0.4 9.3.4. EM for Bayesian linear regression

skip

4 9.4. The EM Algorithm in General

- EM . .

`EM` .

$$\ln p(\mathbf{X} \mid \boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) \right\} \qquad (9.29)$$

- $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{X}$ $\mathbf{Z}$ $\boldsymbol{\theta}$
  likelihood(9.69) log-likelihood(9.29) $\boldsymbol{\theta}$, EM .

$$p(\mathbf{X} \mid \boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) \qquad (9.29) \tag{28}$$

$$\ln p(\mathbf{X} \mid \boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) \right\} \qquad (9.69) \tag{29}$$

- $\ln p(\mathbf{X} \mid \boldsymbol{\theta})$.

$$\ln p(\mathbf{X} \mid \boldsymbol{\theta}) = \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q \| p) \qquad (9.70) \tag{30}$$

$$\mathcal{L}(q, \boldsymbol{\theta}) = \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})}{q(\mathbf{Z})} \right\} \qquad (9.71) \tag{31}$$

$$\mathrm{KL}(q \| p) = -\sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}))}{q(\mathbf{Z})} \right\} \qquad (9.72) \tag{32}$$

  $\mathcal{L}(q, \boldsymbol{\theta})$ $q(\mathbf{Z})$ functional.? (Appendix D)
  .
  KL 0 $\ln p(\mathbf{X} \mid \boldsymbol{\theta}) \geq \mathcal{L}(q, \boldsymbol{\theta})$.
  $\mathcal{L}(q, \boldsymbol{\theta})$ $\ln p(\mathbf{X} \mid \boldsymbol{\theta})$ lower bound.

- EM ...
  'E '(figure 9.12) $p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})$ $\boldsymbol{\theta}$ $\boldsymbol{\theta}^{\text{old}}$ $\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$ $q(\mathbf{Z})$.
  $\ln p(\mathbf{X} \mid \boldsymbol{\theta}^{\text{old}})$ $q(\mathbf{Z})$ $q(\mathbf{Z})$ $\ln p(\mathbf{X} \mid \boldsymbol{\theta}^{\text{old}})$.(figure 9.11)
  $\mathrm{KL}(q \| p)$ $q(\mathbf{Z})$ $q(\mathbf{Z})$ $\mathrm{KL}(q \| p)$ $\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$.
  $\mathrm{KL}(q \| p)$ $q(\mathbf{Z})$ $p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}}))$.
  $\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$.

$$\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}} \ln p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \tag{33}$$

$$= \mathbb{E}_{\mathbf{Z}}\big[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta})\big] + \text{const} \tag{34}$$

$$= \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) + \text{const} \qquad (9.74) \tag{35}$$

$(-\sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}} \ln p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{\text{old}})$ q .) .
'M '(figure 9.13) $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ $\boldsymbol{\theta}$, $\ln \mathcal{N}$.