复习

李子龙
上海交通大学
计算机科学与工程系
logcreative@outlook.com

1 聚类

1.1 k-means

Algorithm 1: k-means

初始化均值点 $\mu_1, \cdots, \mu_k$;
repeat
  foreach $x_n \in X$ **do**
    foreach k **do**

$$r_{nk} = \begin{cases} 1, & \text{if } k = \arg\min_j \|x_n - \mu_j\|^2 \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

  foreach μ_n **do**

$$\mu_i \leftarrow \frac{\sum_n r_{nk} x_n}{\sum_n b_{nk}} \tag{1.2}$$

until μ_i分配不变;

1.2 GMM

1.3 GDA

Bernoulli

Algorithm 2: GMM

初始化均值矩阵 $\boldsymbol{\mu}_k$，协方差矩阵 $\boldsymbol{\Sigma}_k$，混合参数 π_k，初始化对数似然值;

repeat

  for $n \leftarrow 1$ *to* N **do**

    for $k \leftarrow 1$ *to* K **do**

$$\gamma_{nk}^{(t)} \leftarrow \frac{\pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{1.3}$$

  for $k \leftarrow 1$ *to* K **do**

$$N_k = \sum_{n=1}^{N} \gamma_{nk} \tag{1.4}$$

$$\boldsymbol{\mu}_k^{(t+1)} \leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}^{(t)} \boldsymbol{x}_n \tag{1.5}$$

$$\boldsymbol{\Sigma}_k^{(t+1)} \leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}^{(t)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(t+1)})(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(t+1)})^T \tag{1.6}$$

$$\pi_k \leftarrow \frac{N_k}{N} \tag{1.7}$$

until $\ln p(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$没有明显变化;

return $\boldsymbol{\mu}, \boldsymbol{\Sigma}$

$$\begin{aligned} y &\sim \text{Bernoulli}(\phi) \\ x|y = 0 &\sim \mathcal{N}(\mu_0, \Sigma) \\ x|y = 1 &\sim \mathcal{N}(\mu_1, \Sigma) \end{aligned}$$

$$\begin{aligned} p(y) &= \phi^y (1-\phi)^{1-y} \\ p(x|y=0) &= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^\top \Sigma^{-1}(x-\mu_0)\right) \\ p(x|y=1) &= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^\top \Sigma^{-1}(x-\mu_1)\right) \end{aligned}$$

$$\log \int p(x|y)p(y)dy \tag{1.8}$$

$$\phi = \frac{1}{n}\sum_{i=1}^{n} |y_i = 1|$$
$$\mu_0 = \frac{\sum_{i=1}^{n} |y_i = 0| x_i}{\sum_{i=1}^{n} |y_i = 0|}$$
$$\mu_1 = \frac{\sum_{i=1}^{n} |y_i = 1| x_i}{\sum_{i=1}^{n} |y_i = 1|}$$
$$\Sigma = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_{y_i})(x_i - \mu_{y_i})^\top$$

1.4 EM

$$p(x, z) = p(x|z)p(z) \tag{1.9}$$
$$p(x) = \sum_z p(x, z) = \sum_z p(x|z)p(z) \tag{1.10}$$

Algorithm 3: EM

1 初始化参数 θ^{old};

2 E 步：评估 $p(Z|X, \theta^{\text{old}})$;

3 M 步：得到新的参数

$$\theta^{\text{new}} = \arg\max_\theta Q(\theta, \theta^{\text{old}})$$
$$Q(\theta, \theta^{\text{old}}) = \sum_Z p(Z|X, \theta^{\text{old}}) \ln p(X, Z|\theta)$$

查看对数似然值是否收敛，不收敛返回步骤 2。

$$\begin{tikzcd}
y \arrow[d, bend left, "{q(x|y) = q(x|y, \theta)}"] & q(y) = q(y|\theta) \\
x \arrow[u, bend left, "{p(y|x)}"] & p(x) = \delta(x - x_N)
\end{tikzcd}$$

Jenson's Inequation Method

$$\begin{aligned}
\log(P(x|\theta)) &= \log\left(\sum_y P(x,y|\theta)\right) \\
&= \log\left(\sum_y q(y)\frac{P(x,y|\theta)}{q(y)}\right) \\
&\geq E_q\left[\log\frac{P(x,y|\theta)}{q(y)}\right] \\
&\geq E_q\left[\log\frac{P(y|x,\theta)P(x|\theta)}{q(y)}\right] \\
&\geq E_q[\log(P(x|\theta))] - E_q\left[\log\frac{q(y)}{P(y|x,\theta))})\right] \\
&\geq E_q[\log(P(x|\theta))] - KL(q(y)\|P(y|x,\theta)) \\
&\geq \boxed{\log(P(x|\theta)) - KL(q(y)\|P(y|x,\theta))}
\end{aligned}$$

$$\begin{aligned}
\log(P(x|\theta)) &\geq E_q\left[\log\frac{P(x,y|\theta)}{q(y)}\right] \\
&\geq E_q\left[\log(P(x,y|\theta))\right] - E_q[\log(q(y))] \\
&\geq E_q\left[\log(P(x,y|\theta))\right] + H(q(y))
\end{aligned}$$

KL Matching

$$p(x) = \delta(x - \mathbf{x}_N) \tag{1.11}$$

$$\max_\theta q(x|\theta) = \max_\theta \int q(x|y)q(y)dy$$

$$p(x) \leftrightarrow q(x|\theta) \tag{1.12}$$

$$\begin{aligned}
\min_\theta KL(p||q) &= \min_\theta \int p(x)\log\frac{p(x)}{q(x|\theta)} \\
&= \min_\theta \int p(x)\log p(x) - \int p(x)\log q(x|\theta) \\
&= \min_\theta \int \frac{1}{N}\sum_{t=1}^{N}\delta(x-x_t)\log q(x|\theta) \\
&= \frac{1}{N}\sum_{t=1}^{N}\log q(x_t|\theta) \qquad\qquad \int \delta(x)f(x) = f(0)
\end{aligned}$$

$$p(x)p(y|x) \leftrightarrow q(x|y,\theta)q(y|\theta) \tag{1.13}$$

$$
\begin{aligned}
\max F(p(y|x),\theta) &= \int p(x)p(y|x)\log\frac{q(x|y,\theta)q(y|\theta)}{p(y|x)} \\
&= \int p(x)p(y|x)\log\frac{q(x|y,\theta)q(y|\theta)}{q(x|\theta)}\frac{q(x|\theta)}{p(y|x)} \\
&= \int p(x)p(y|x)\left[\log q(x|\theta) - \log\frac{p(y|x)}{q(y|x,\theta)}\right] \\
&= \frac{1}{N}\sum_{t=1}^{N}\log q(x_t|\theta) - \int p(x)\cdot KL(p(y|x)||q(y|x,\theta)) \\
&\leq \log q(x|\theta)@[p(y|x) = q(y|x,\theta)]
\end{aligned}
$$

fix $p(y|x)$, $F = \int p(y|x)\log[q(x|y,\theta)q(y|\theta)]$

E Step – fix $\theta : p(y|x) = q(y|x,\theta)$

M Step – fix $p(y|x)$,

$$
F = \int p(y|x)\log[q(x|y,\theta)q(y|\theta)]) - \int p(y|x)^{\text{old}}\log p(y|x)
$$

$\max_\theta F(P(y|x)^{\text{old}},\theta)$

2 贝叶斯学习

Bayes rule:

$$
P(\Theta|X) = \frac{P(X|\Theta)P(\Theta)}{P(X)} \tag{2.1}
$$

Bayes learning - Maximum A Posteriori (MAP)

引入先验知识。

$$
\max_\Theta \log p(\Theta|X) = \max_\Theta(\log p(X|\Theta) + \log p(\Theta)) \tag{2.2}
$$

例子：

$$
\begin{aligned}
p(x|\Theta) &= G(x|\mu, \Sigma) \\
p(\mu) &= G(\mu|\mu_0, \sigma_0^2) \\
p(\mu|x) &\propto p(x|\mu, \Sigma)p(\mu) \\
&= \prod_{t=1}^{N} p(x_t|\mu, \Sigma)p(\mu) && \text{(i.i.d.)} \\
\log p(\mu|x) &\propto \sum_{t=1}^{N} \log G(x_t|\mu, \Sigma) + \log G(\mu|\mu_0, \sigma_0^2 I) \\
&= \sum_{t=1}^{N} \left\{ -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_t - \mu)^\top \Sigma^{-1}(x_t - \mu) \right\} \\
&\quad - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\sigma_0^2 I| - \frac{1}{2}(\mu - \mu_0)^\top \sigma_0^{-2} I(\mu - \mu_0)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log p(\mu|x)}{\partial \mu} &= \sum_{t=1}^{N} \left\{ \frac{1}{2} \times 2\Sigma^{-1}(x_t - \mu)(-1) \right\} - \frac{1}{2} \times 2\sigma_0^{-2}(\mu - \mu_0) \\
&= \Sigma^{-1}(N\bar{x} - N\mu) - \sigma_0^{-2}\mu + \sigma_0^{-2}\mu_0 = 0 \\
\mu &= (N\Sigma^{-1} + \sigma_0^{-2} I)^{-1}(N\Sigma^{-1}\bar{x} + \sigma_0^{-2}\mu_0) \\
&= (\sigma_0^2\Sigma^{-1} + \frac{1}{N} I)^{-1}(\sigma_0^2\Sigma^{-1}\bar{x} + \frac{1}{N}\mu_0)
\end{aligned}
$$

For a fixed σ_0^2, $N \to 0$, $\mu \to (\sigma_0^2\Sigma^{-1})^{-1}(\sigma_0^2\Sigma^{-1})\bar{x} = \bar{x}$

For a fixed $N \ll +\infty, \sigma_0^2 \to 0, \mu \to (\frac{1}{N} I)^{-1}(\frac{1}{N}\mu_0) = \mu_0$

$$
-2\log p(\mu|X) \propto \sum_t \|x_t - \mu\|^2 + \|\mu - \mu_0\|^2 \tag{2.3}
$$

$\max_k q(x|k) = \int q(x|\theta)q(\theta)d\theta$ marginal likelihood

ML $\log q(x|\theta)$

BL $\log q(x|\theta)q(\theta) = \log q(x|\theta) + \log q(\theta)$ 后项为正则化项

2.1 VAE

$$
\log P(X) - \mathcal{D}[Q(z|x)\|P(z|X)] = E_{z\sim Q}[\log P(X|z)] - \mathcal{D}[Q(z|X)\|P(z)] \tag{2.4}
$$

$$
\begin{aligned}
\log P(x) &= \int_z q(z|x)\log P(x)dz \\
&= \int_z q(z|x)\log \frac{P(z,x)}{P(z|x)}dz \\
&= \int_z q(z|x)\log \left(\frac{P(z,x)}{q(z|x)}\frac{q(z|x)}{P(z|x)}\right)dz \\
&= \int_z q(z|x)\log \frac{P(z,x)}{q(z|x)}dz + \int_z q(z|x)\log \frac{q(z|x)}{P(z|x)}dz \\
&= \int_z q(z|x)\log \frac{P(z,x)}{q(z|x)}dz + KL(q(z|x)\|P(z|x)) \\
&= \int_z q(z|x)\log \frac{P(x|z)P(z)}{q(z|x)}dz + KL(q(z|x)\|P(z|x)) \\
&= \int_z q(z|x)\log P(x|z)dz - \int_z q(z|x)\log \frac{q(z|x)}{P(x|z)}dz + KL(q(z|x)\|P(z|x)) \\
&= \int_z q(z|x)\log P(x|z)dz - KL(q(z|x)\|P(x|z)) + KL(q(z|x)\|P(z|x))
\end{aligned}
$$

Reparameterization

$$z \sim \mathcal{N}(z|\mu,\sigma^2) \tag{2.5}$$

$$z = \mu + \sigma \cdot \epsilon \tag{2.6}$$

$$\epsilon \sim \mathcal{N}(\epsilon|0,1) \tag{2.7}$$

虽然VAE比普通的AE模型训练出来的效果要好很多，但是训练过VAE模型的人都知道，它生成出来的图片相对GANs那种直接利用对抗学习的方式会比较模糊，这是由于它是通过直接计算生成图片和原始图片之间的均方误差，所以得到的是一张“平均图像”。

Conditional VAE given y_i

2.2 GAN

$$\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{2.8}$$

$$
\begin{aligned}
C(G) &= E_{x\sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(x)}{p_{\text{data}}(x)+p_g(x)}\right] + E_{x\sim p_g}\left[\log \frac{p_g(x)}{p_{\text{data}}(x)+p_g(x)}\right] \\
&= -\log(4) + KL\left(p_{\text{data}}\middle\|\frac{p_{\text{data}}+p_g}{2}\right) + KL\left(p_g\middle\|\frac{p_{\text{data}}+p_g}{2}\right) \\
&= -\log(4) + 2\cdot JSD(p_{\text{data}}\|p_g)
\end{aligned}
$$

Algorithm 4: 特征值分解 PCA

Input: 数据集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}, \mathbf{x}_t \in \mathbb{R}^{n\times 1}$

Output: 主成分 $\mathbf{w}$

计算平均值 $\mu \leftarrow \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$;

foreach $i \leftarrow 1$ *to* N **do**

  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mu$;

计算散度矩阵 $C \leftarrow XX^T$;

特征值分解求 C 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 与对应的特征向量 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$;

选取最大的特征值对应的特征向量与数据的乘积即为主成分 $\mathbf{w} \leftarrow \mathbf{v_1}^T\mathbf{X}$;

return $\mathbf{w}$;

Algorithm 5: 奇异值分解

Input: 数据集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}, \mathbf{x}_t \in \mathbb{R}^{n\times 1}$

Output: 主成分 $\mathbf{w}$

计算平均值 $\mu \leftarrow \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$;

foreach $i \leftarrow 1$ *to* N **do**

  $\mathbf{x}_i \leftarrow \mathbf{x}_i - \mu$;

奇异值分解 $\mathbf{X} = \mathbf{U\Sigma V}^T$;

两边同乘 $\mathbf{U}^T$，$\mathbf{U}^T\mathbf{X} = \mathbf{\Sigma V}^T$ 得到压缩数据;

选取 $\mathbf{\Sigma V}^T$ 中最大的那一个奇异值（习惯上应为左上角的值）对应的向量（一般为第一行）即为主成分 $\mathbf{w}$;

return $\mathbf{w}$;

3 数据降维

3.1 PCA

$$
\begin{tikzcd}
y \arrow[d, bend left, "q(x|y) = G(x|Ay+\mu, \Sigma_e)"] & q(y) = G(y|0, \Sigma_y) \\
x & x = Ay + \mu + e
\end{tikzcd}
$$

$$
\int q(x,y)dy = q(x|\theta) = G(x|\mu_x, \Sigma_x)
$$

$$
\begin{aligned}
\mu_x &= E[x] = AE[y] + \mu + E[e] = \mu \qquad\qquad \Sigma_y = I \\
\Sigma_x &= \mathrm{cov}(x) \\
&= E[(x-\mu)(x-\mu)^\top] \\
&= E[(Ay+e)(Ay+e)^\top] \\
&= E[Ayy^\top A + Aye^\top + e(Ay)^\top + ee^\top] \\
&= E[Ayy^\top A + ee^\top] \\
&= A\Sigma_y A^\top + \Sigma_e
\end{aligned}
$$

$$
\frac{\sum_{i=1}^{d'} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq t \tag{3.1}
$$

3.2 FA

$$
\begin{tikzcd}
y \arrow[d, bend left, "q(x|y) = G(x|Ay+\mu, \Sigma_e)"] & q(y) = G(y|0, \Sigma_y) \\
x \arrow[u, bend left, "p(y|x) = Wx"] & x = Ay + \mu + e
\end{tikzcd}
$$

E-Step:

$$
\begin{aligned}
q(x,y) &= \frac{1}{()} e^{-\frac{1}{2}(x-Ay-\mu)^\top \Sigma_x^{-1}(x-Ay-\mu)} \frac{1}{()} e^{-\frac{1}{2}y^\top I^{-1} y} \\
&\propto \exp\left\{-\frac{1}{2}\left[y^\top A^\top \Sigma_e^{-1} Ay - 2y^\top A^\top \Sigma_e^{-1}(x-\mu) + (x-\mu)^\top \Sigma_e^{-1}(x-\mu)\right] - \frac{1}{2} y^\top I^{-1} y\right\} \\
&= \exp\left\{\frac{1}{2}\left[y^\top (A^\top \Sigma_e^{-1} A + I^{-1}) y - 2y^\top A^\top \Sigma_e^{-1}(x-\mu)\right] + \cdots\right\}
\end{aligned}
$$

$$
\begin{aligned}
p(y|x) = q(y|x,\theta) &= \frac{q(x|y)q(y)}{q(x|\theta)} \\
&= \frac{q(x,y)}{q(x|\theta)} \\
&= G(y|\mu_{y|x}, \Sigma_{y|x}) \\
&\propto \exp\left\{-\frac{1}{2}(y-\mu_{y|x})^\top \Sigma_{y|x}^{-1}(y-\mu_{y|x})\right\} \\
&= \exp\left\{-\frac{1}{2} y^\top \Sigma_{y|x}^{-1} y - 2y^T \Sigma_{y|x}^{-1} \mu_{y|x} + \mu_{y|x}^{-1} \mu_{y|x}\right\}
\end{aligned}
$$

$$\Rightarrow \begin{cases} \Sigma_{y|x}^{-1} = A^\top \Sigma_e^{-1} A + I^{-1} \\ \Sigma_{y|x}^{-1}\mu_{y|x} = A^\top \Sigma_e^{-1}(x-\mu) \end{cases} \quad \text{归一化 } x \text{ 则 } y \text{ 系数一致}$$

$$\Rightarrow \begin{cases} \Sigma_{y|x} = (A^\top \Sigma_e^{-1} A + I)^{-1} \\ \mu_{y|x} = (A^\top \Sigma_e^{-1} A + I)^{-1} A^\top \Sigma_e^{-1}(x-\mu) = (A^\top A + \sigma_e^2 I)^{-1} A^\top (x-\mu) \end{cases}$$

$$\Rightarrow W = (A^\top A + \sigma_e^2 I)^{-1} A^\top \approx A^{-1}\text{伪逆}$$

M-Step: $\max_\Theta Q(p^{\text{old}}(y|x), \Theta])$

$$Q = \int p^{\text{old}}(y|x) \cdot \ln[G(y|0,I)G(x|Ay+\mu, \sigma^2 I)]dy$$

$$A_{\text{new}} = \left(\sum_{t=1}^N x_t (E[y|x_t])^\top\right)\left(\sum_{t=1}^N E[yy^\top|x_t]\right)^{-1}$$

$$\sigma_{\text{new}}^2 = \frac{1}{Nd}\text{Tr}\left\{\sum_{t=1}^N (x_t x_t^\top - A_{\text{new}} E[y|x_t] x_t^\top)\right\}$$

$$q(x|\theta) = G(x|\mu, \Sigma_x) = G(x|AA^\top + \sigma^2 I)$$

$$\text{i.i.d.}\{x_t\}_{t=1}^N, \max_\theta \prod_{t=1}^N q(x_t|\theta) \Leftrightarrow \max_\theta \sum_{t=1}^N \log q(x_t|\theta)$$

$$\Rightarrow \begin{cases} \mu = \frac{1}{N}\sum_{t=1}^N x_t \\ \Sigma_x = UDU^\top = \frac{1}{N}\sum_t (x_t-\mu)(x_t-\mu)^\top = AA^\top + \sigma^2 I \end{cases}$$

Bias-variance decompostion

$$x = f(y) + \epsilon$$

$$E\left[\left(x - \hat{f}(y)\right)^2\right] = E[x^2 + \hat{f}^2 - 2x\hat{f}]$$

$$= \text{Var}[\epsilon] + \text{Var}[\hat{f}] + \text{Bias}^2[\hat{f}]$$

Model Selection

$$J_{\text{AIC}} = \ln p(X_N|\hat{\Theta}_k) - d_k$$

$$J_{\text{BIC}} = \ln p(X_N|\hat{\Theta}_k) - \frac{1}{2} d_k \ln N$$

3.3 VFA (Variational FA)

$p(x)p(y|x,\theta)p(\theta|x) \leftrightarrow q(x|y,\theta)q(y|\theta)q(\theta)$

$$
\begin{aligned}
F &= \int p(x)p(y|x,\theta)p(\theta|x)\log\frac{q(x|y,\theta)q(y|\theta)q(\theta)}{p(y|x,\theta)p(\theta|x)} \\
&= \int p(x)p(y|x,\theta)p(\theta|x)\log\frac{q(x|y,\theta)q(y|\theta)}{q(x|\theta)}\frac{q(x|\theta)q(\theta)}{p(y|x,\theta)p(\theta|x)} \\
&= \int p(x)p(y|x,\theta)p(\theta|x)\log\frac{q(y|x,\theta)}{p(y|x,\theta)}\frac{q(x|\theta)q(\theta)}{q(x|k)p(\theta|x)}q(x|k) \\
&= \int p(x)p(y|x,\theta)p(\theta|x)\left(\log\frac{q(y|x,\theta)}{p(y|x,\theta)}+\log\frac{q(\theta|x)}{p(\theta|x)}+\log q(x|k)\right) \quad \text{VBEM}
\end{aligned}
$$

3.4 ICA

$$x = As \tag{3.2}$$

Algorithm 6: FastICA

1. Choose an initial weight vector $\mathbf{w}$;
2. Let $\mathbf{w}^+ = E\{\mathbf{w}g(\mathbf{w}^\top\mathbf{x})\} - E\{g'(\mathbf{w}^\top\mathbf{x})\}\mathbf{w}$;
3. Let $\mathbf{w} = \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$;
4. If not converged, go back to 2.

4 SVM

基础的支持向量机（Support Vector Machine, SVM）是一种二分类模型，是一种定义在特征空间上的间隔最大的分类器模型。对于本文的三分类问题，将会采用一对一策略（one-vs-one scheme），分解为三个二分类问题，组合这些分类器进行求解。

本文使用核函数方法映射到另一个特征空间进行分类，采用软间隔方法允许一定的误差、配合 L_2 正则化降低过拟合风险**?**：

$$
\begin{aligned}
\min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\
\text{subject to } & y_i(\mathbf{w}^\top\mathbf{x}_i + b) \geq 1 - \xi_i \\
& \xi_i \geq 0, \quad i = 1, 2, \cdots, m
\end{aligned} \tag{4.1}
$$

其中 C 为正则化系数，ξ_i 为“松弛变量”（slack variables），(4.1) 的对偶问题为

$$
\begin{aligned}
\min_{\alpha} \quad & \frac{1}{2}\alpha^\top\mathbf{Q}\alpha - \mathbf{e}^\top\alpha \\
\text{subject to } & y^\top\alpha = 0 \\
& 0 \leq \alpha_i \leq C, \quad i = 1, \cdots, n
\end{aligned} \tag{4.2}
$$

其中

$$\mathbf{Q} = (Q_{ij})_{n\times n} = (y_i y_j \kappa(\mathbf{x}_i, \mathbf{x}_j))_{n\times n} \tag{4.3}$$

这里 $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top\phi(\mathbf{x}_j)$ 为核函数。

表 1: 仅分类

	k-means	RPCL
E-Step	$b_i^t \leftarrow \begin{cases} 1 & \text{if } \lVert x^t - m_i \rVert = \min_j \lVert x^t - m_j \rVert \\ 0 & \text{otherwise} \end{cases}$	$p_{j,t} = \begin{cases} 1 & \text{if } j = \arg\min_j \epsilon_t(\theta_j) \\ -\gamma & \text{if } j = \arg\min_{j \neq c} \epsilon_t(\theta_j) \\ 0 & \text{otherwise} \end{cases}$
M-Step	$\mu_i \leftarrow \dfrac{\sum_n r_{nk} x_n}{\sum_n b_{nk}}$	$m_j^{\text{new}} \leftarrow m_j^{\text{old}} + \eta p_{j,t}(x_t - m_j^{\text{old}})$

表 2: 含有概率模型

	GMM	EM
E-Step	$\gamma_{nk}^{(t)} \leftarrow \dfrac{\pi_k \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$ $N_k = \sum_{n=1}^{N} \gamma_{nk}$	$p(Z \mid X, \theta^{\text{old}})$
M-Step	$\boldsymbol{\mu}_k^{(t+1)} \leftarrow \dfrac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}^{(t)} \boldsymbol{x}_n$ $\boldsymbol{\Sigma}_k^{(t+1)} \leftarrow \dfrac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk}^{(t)} (\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(t+1)})(\boldsymbol{x}_n - \boldsymbol{\mu}_k^{(t+1)})^T$ $\pi_k \leftarrow \dfrac{N_k}{N}$	$\theta^{\text{new}} = \arg\max_\theta Q(\theta, \theta^{\text{old}})$ $Q(\theta, \theta^{\text{old}}) = \sum_Z p(Z \mid X, \theta^{\text{old}}) \ln p(X, Z \mid \theta)$

表 3: 两个学派

MLE	MAP
$\theta_{\text{MLE}} = \arg\max_\theta \prod_{i=1}^{n} p\left(y^{(i)} \mid x^{(i)}, \theta\right)$	$\theta_{\text{MAP}} = \arg\max_\theta \prod_{i=1}^{n} p\left(y^{(i)} \mid x^{(i)}, \theta\right) p(\theta)$

表 4: 数据降维

PCA	FA	ICA
特征值分解 $XX^\top w = \lambda w$	假设数据 $x_t = Ay_t + \mu + e_t$	假设 $x = As$
或者奇异值分解 $X = UDV^\top \Rightarrow U^\top X = DV^\top$	EM 拟合参数	FastICA 寻找 $s = Wx$