Gaussian Mixture Model Note

FanmingL

2018 年 11 月 30 日

1 高斯混合模型

1.1 高斯混合模型的数学形式

$$\mathcal{N}(\mathbf{x}|\mu,\Sigma)=\frac{1}{(2\pi)^{d/2}det(\Sigma)^{\frac{1}{2}}}\exp(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)) \tag{1}$$

式1为高斯分布的概率密度函数，其中 μ 为变量的均值，Σ 为协方差矩阵，d 为向量维数，可以证明协方差矩阵是半正定的矩阵，所以协方差矩阵必然是实对称的。

假设某个分布由 K 个高斯分布构成，每次输出，都分两步，第一步从这 K 个高斯分布中随机选择一个，然后以此高斯分布输出。第 k 个高斯分布被选择到的概率为 π_k，其均值和方差分别为 μ_k,Σ_k，为方便表示，引入一隐变量 $\mathbf{z}$，若 $\mathbf{z}$ 的第 k 维，$z_k=1$ 表示选择了第 k 个高斯分布作为输出的分布，那么显然 $\mathbf{z}$ 的 K 维中有且只有一维为 1。于是我们可以写出 z_k 的分布如下。

$$p(z_k=1)=\pi_k \tag{2}$$

$\mathbf{z}$ 的分布如下

$$p(\mathbf{z})=\prod_{k=1}^{K}(\pi_k)^{z_k} \tag{3}$$

我们称 $p(\mathbf{z})$ 为先验概率。

当我们已知 $\mathbf{z}$ 时，即我们已知哪个高斯模型被选中时，条件概率密度分布可写为

$$p(\mathbf{x}|\mathbf{z})=\prod_{k=1}^{K}\mathcal{N}(\mathbf{x}|\mu_k,\Sigma_k)^{z_k} \tag{4}$$

我们称 $p(\mathbf{x}|\mathbf{z})$ 为似然概率。于是我们可以用全概率公式写出

$$p(\mathbf{x})=\sum_{k=1}^{K}p(z_k=1)\mathcal{N}(\mathbf{x}|\mu_k,\Sigma_k)=\sum_{k=1}^{K}\pi_k\mathcal{N}(\mathbf{x}|\mu_k,\Sigma_k) \tag{5}$$

我们称 $p(\mathbf{x})$ 为证据。当我们有了证据、似然、先验，我们可以用贝叶斯公式，得到后验概率为

$$\begin{aligned}p(z_k=1|\mathbf{x})&=\frac{p(z_k=1)p(\mathbf{x}|z_k=1)}{p(\mathbf{x})}\\&=\frac{\pi_k\mathcal{N}(\mathbf{x}|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(\mathbf{x}|\mu_j,\Sigma_j)}\end{aligned} \tag{6}$$

我们称 $p(z_k=1|\mathbf{x})$ 为后验概率密度分布，即当我们有了 $\mathbf{x}$ 的样本之后，对 z_k 概率密度分布的估计。

1.2 EM 算法

实际上我们对参数 π_k,μ_k,Σ_k 都一无所知，当我们拿到数据时，应当如何去估计这三者的值呢？自然是极大似然估计了。假设我们有 N 个样本，那么似然函数可以写为

$$\begin{aligned}L(\pi,\mu,\Sigma)&=\prod_{n=1}^{N}p(\mathbf{x}_n;\pi,\mu,\Sigma)\\&=\prod_{n=1}^{N}\left(\sum_{k=1}^{K}\pi_k\mathcal{N}(\mathbf{x}|\mu_k,\Sigma_k)\right)\end{aligned} \tag{7}$$

两边取对数，得到对数似然函数

$$
\begin{aligned}
l(\pi,\mu,\Sigma) &= \ln(L(\pi,\mu,\Sigma)) \\
&= \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)\right)
\end{aligned} \tag{8}
$$

将 l 对 μ_k 求偏导得

$$
\begin{aligned}
\frac{\partial l}{\partial \mu_k} &= \sum_{n=1}^{N} \frac{\pi_k \frac{\partial \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)}{\partial \mu_k}}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n;\mu_j,\Sigma_j)} \\
&= \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k) \frac{\partial(-1/2(\mathbf{x}_n-\mu_k)^T \Sigma_k^{-1}(\mathbf{x}_n-\mu_k))}{\partial \mu_k}}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n;\mu_j,\Sigma_j)} \\
&= \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k) \Sigma_k^{-1}(\mathbf{x}_n-\mu_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n;\mu_j,\Sigma_j)} \\
&= \sum_{n=1}^{N} \left[\frac{\pi_k \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n;\mu_j,\Sigma_j)} \Sigma_k^{-1}(\mathbf{x}_n-\mu_k)\right]
\end{aligned} \tag{9}
$$

上式，中括号的左半部分恰是后验概率分布，即式6，为方便表示，令

$$
\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n;\mu_j,\Sigma_j)} \tag{10}
$$

所以 l 对 μ_k 的偏导数为

$$
\begin{aligned}
\frac{\partial l}{\partial \mu_k} &= \sum_{n=1}^{N} \left[\gamma(z_{nk}) \Sigma_k^{-1}(\mathbf{x}_n-\mu_k)\right] \\
&= \sum_{n=1}^{N} \left[\gamma(z_{nk}) \Sigma_k^{-1}\mathbf{x}_n\right] - \sum_{n=1}^{N} \left[\gamma(z_{nk}) \Sigma_k^{-1}\mu_k\right] \\
&= \sum_{n=1}^{N} \left[\gamma(z_{nk}) \Sigma_k^{-1}\mathbf{x}_n\right] - \Sigma_k^{-1}\mu_k \sum_{n=1}^{N} \gamma(z_{nk})
\end{aligned} \tag{11}
$$

为方便表示令

$$
N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \tag{12}
$$

再令 $\frac{\partial l}{\partial \mu_k} = 0$ 得

$$
\begin{gathered}
\sum_{n=1}^{N} \left[\gamma(z_{nk}) \Sigma_k^{-1}\mathbf{x}_n\right] = \Sigma_k^{-1}\mu_k N_k \\
\sum_{n=1}^{N} \left[\gamma(z_{nk})\mathbf{x}_n\right] = \mu_k N_k
\end{gathered} \tag{13}
$$

即

$$
\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})\mathbf{x}_n \tag{14}
$$

至此求出了 μ_k 的表示式即式14。

下面计算 π_k 的极大似然估计，由于 π_k 有如下约束

$$\sum_{k}^{K} \pi_k = 1 \tag{15}$$

故引入拉格朗日乘子 λ，将对数似然估计写为

$$l'(\pi, \mu, \Sigma) = \sum_{n=1}^{N} \left[\ln \left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k) \right) \right] + \lambda(\sum_{j=1}^{K} \pi_j - 1) \tag{16}$$

l' 对 π_k 求偏导得

$$\frac{\partial l'}{\partial \pi_k} = \sum_{n=1}^{N} \frac{\mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n; \mu_j, \Sigma_j)} + \lambda \tag{17}$$

令其为零，即

$$\sum_{n=1}^{N} \frac{\mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n; \mu_j, \Sigma_j)} + \lambda = 0 \tag{18}$$

两边乘上 π_k 得

$$\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n; \mu_j, \Sigma_j)} + \lambda \pi_k = 0 \tag{19}$$

将每个 k 叠加，得到

$$\begin{aligned} \sum_{n=1}^{N} \frac{\sum_{j=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n; \mu_j, \Sigma_j)} + \lambda \sum_{j=1}^{K} \pi_j &= 0 \\ \sum_{n=1}^{N} 1 + \lambda \sum_{j=1}^{K} \pi_j &= 0 \\ N &= -\lambda \sum_{j=1}^{K} \pi_j \end{aligned} \tag{20}$$

将约束条件代入上式，得

$$\lambda = -N \tag{21}$$

将式21代入至式19中得

$$\begin{aligned} \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n; \mu_j, \Sigma_j)} - N\pi_k &= 0 \\ \sum_{n=1}^{N} \gamma(z_{nk}) &= N\pi_k \\ N_k &= N\pi_k \\ \pi_k &= \frac{N_k}{N} \end{aligned} \tag{22}$$

于是我们得到 π_k 的表示式

$$\pi_k = \frac{N_k}{N} \tag{23}$$

下面推导 Σ_k 的极大似然估计，同样是令对数似然函数8对 Σ_k 求偏导数。

$$\frac{\partial l}{\partial \Sigma_k} = \sum_{n=1}^{N} \frac{\pi_k \frac{\partial \mathcal{N}(\mathbf{x}_n; \mu_k, \Sigma_k)}{\partial \Sigma_k}}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n; \mu_j, \Sigma_j)} \tag{24}$$

下面单独求 $\frac{\partial \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)}{\partial \Sigma_k}$ 的值

$$
\begin{aligned}
\frac{\partial \mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)}{\partial \Sigma_k} &= \frac{\partial \frac{1}{(2\pi)^{d/2}\sqrt{det(\Sigma_k)}}\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))}{\partial \Sigma_k} \\
&= \frac{1}{(2\pi)^{d/2}}\frac{\partial det(\Sigma_k)^{-\frac{1}{2}}\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))}{\partial \Sigma_k} \\
&= \frac{1}{(2\pi)^{d/2}}[\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))\frac{\partial det(\Sigma_k)^{-\frac{1}{2}}}{\partial \Sigma_k} \\
&+ det(\Sigma_k)^{-\frac{1}{2}}\frac{\partial \exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))}{\partial \Sigma_k}] \\
&= \frac{1}{(2\pi)^{d/2}}[\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))\left(-\frac{1}{2}det(\Sigma_k)^{-\frac{3}{2}}det(\Sigma_k)\Sigma_k^{-1}\right) \\
&- \frac{1}{2}det(\Sigma_k)^{-\frac{1}{2}}\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))\frac{\partial((\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))}{\partial \Sigma_k}] \\
&= \frac{1}{(2\pi)^{d/2}}\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))[\left(-\frac{1}{2}det(\Sigma_k)^{-\frac{1}{2}}\Sigma_k^{-1}\right) \\
&- \frac{1}{2}det(\Sigma_k)^{-\frac{1}{2}}\left(-\Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}\right)] \\
&= -\frac{1}{2}det(\Sigma_k)^{-\frac{1}{2}}\frac{1}{(2\pi)^{d/2}}\exp(-\frac{1}{2}(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}_n-\mu_k))[\Sigma_k^{-1} \\
&- \Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}] \\
&= -\frac{1}{2}\mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)\left[\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}\right]
\end{aligned} \tag{25}
$$

所以

$$
\begin{aligned}
\frac{\partial l}{\partial \Sigma_k} &= -\frac{1}{2}\sum_{n=1}^{N}\left[\frac{\pi_k\mathcal{N}(\mathbf{x}_n;\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(\mathbf{x}_n;\mu_j,\Sigma_j)}\left(\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}\right)\right] \\
&= -\frac{1}{2}\sum_{n=1}^{N}\left[\gamma(z_{nk})\left(\Sigma_k^{-1}-\Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}\right)\right] \\
&= -\frac{1}{2}\left[\sum_{n=1}^{N}\left(\gamma(z_{nk})\Sigma_k^{-1}\right)-\sum_{n=1}^{N}\left(\gamma(z_{nk})\Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}\right)\right]
\end{aligned} \tag{26}
$$

令偏导为零，得

$$
\begin{aligned}
\sum_{n=1}^{N}\left(\gamma(z_{nk})\Sigma_k^{-1}\right) &= \sum_{n=1}^{N}\left(\gamma(z_{nk})\Sigma_k^{-1}(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\Sigma_k^{-1}\right) \\
\Sigma_k\sum_{n=1}^{N}\gamma(z_{nk}) &= \sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T \\
\Sigma_kN_k &= \sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T \\
\Sigma_k &= \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T
\end{aligned} \tag{27}
$$

于是我们得到 Σ_k 的表示式

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T \tag{28}$$

至此我们用 $\gamma(z_{nk})$ 表示出了 μ_k, Σ_k, π_k 的极大似然估计，分别为式14，式28，式23。但是有个问题，后验概率 $\gamma(z_{nk})$ 由式10给出，它本身就与 μ_k, Σ_k, π_k 有很复杂的关系。虽然不可以直接得到 μ_k, Σ_k, π_k 的闭式解，但是，这也暗示了我们，可以利用迭代的方式，去逼近最优解。即 EM 算法，高斯混合模型的 EM 算法可以表示为算法1。

Algorithm 1 对于高斯混合模型的 EM 算法

```
初始化每个高斯混合模型的均值 μ_k，协方差矩阵 Σ_k，以及对应的 π_k
while 算法没有收敛 do
  for n ∈ [1, N] do
    for k ∈ [1, K] do
      E 步：计算 γ(z_nk) = π_k N(x_n; μ_k, Σ_k) / Σ_{j=1}^{K} π_j N(x_n; μ_j, Σ_j)
    end for
  end for
  for k ∈ [1, K] do
    M 步计算 N_k：计算 N_k = Σ_{n=1}^{N} γ(z_nk)
    M 步计算 μ_k：计算 μ_k^new = (1/N_k) Σ_{n=1}^{N} γ(z_nk) x_n
    M 步计算 Σ_k：计算 Σ_k^new = (1/N_k) Σ_{n=1}^{N} γ(z_nk)(x_n − μ_k^new)(x_n − μ_k^new)^T
    M 步计算 π_k：计算 π_k^new = N_k / N
  end for
  计算对数似然概率 l(π, μ, Σ) = Σ_{n=1}^{N} ln( Σ_{k=1}^{K} π_k N(x_n; μ_k, Σ_k) )
  if l(π, μ, Σ) 变化很小时 then
    算法收敛，迭代结束
  end if
end while
```

1.3 在线 EM 算法

实际上 EM 算法的 M 步，可以用梯度上升法求解，为方便，将之前得到的对数似然函数对 μ_k, Σ_k, π_k 的偏导数写在下方。

$$
\begin{aligned}
\frac{\partial l}{\partial \mu_k} &= \sum_{n=1}^{N} \left[\gamma(z_{nk})\Sigma_k^{-1}(\mathbf{x}_n - \mu_k)\right] \\
&\approx N\mathbb{E}_{z_k=1|\mathbf{x}_n}\left[\Sigma_k^{-1}(\mathbf{x}_n - \mu_k)\right] \\
\frac{\partial l}{\partial \Sigma_k} &= -\frac{1}{2}\sum_{n=1}^{N}\left[\gamma(z_{nk})\left(\Sigma_k^{-1} - \Sigma_k^{-1}(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T\Sigma_k^{-1}\right)\right] \\
&\approx -\frac{N}{2}\mathbb{E}_{z_k=1|\mathbf{x}_n}\left[\Sigma_k^{-1} - \Sigma_k^{-1}(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T\Sigma_k^{-1}\right] \\
\frac{\partial l'}{\partial \pi_k} &= \frac{1}{\pi_k}\sum_{n=1}^{N}\gamma(z_{nk}) - N \\
&\approx N\left(\mathbb{E}_{z_k=1|\mathbf{x}_n}\left[\frac{1}{\pi_k}\right] - 1\right)
\end{aligned} \tag{29}
$$

所以根据梯度上升法，三者的迭代公式为

$$
\begin{aligned}
\mu_k^{new} &= \mu_k + \alpha N\mathbb{E}_{z_k=1|\mathbf{x}_n}\left[\Sigma_k^{-1}(\mathbf{x}_n - \mu_k)\right] \\
\Sigma_k^{new} &= \Sigma_k - \alpha\frac{N}{2}\mathbb{E}_{z_k=1|\mathbf{x}_n}\left[\Sigma_k^{-1} - \Sigma_k^{-1}(\mathbf{x}_n - \mu_k^{new})(\mathbf{x}_n - \mu_k^{new})^T\Sigma_k^{-1}\right] \\
\pi_k^{new} &= \pi_k + \alpha N\left(\mathbb{E}_{z_k=1|\mathbf{x}_n}\left[\frac{1}{\pi_k}\right] - 1\right)
\end{aligned} \tag{30}
$$

这三个迭代公式还是要使用到期望操作，这个操作去要用到所有的数据点，不妨用当前值取近似期望。

$$
\begin{aligned}
\mu_k^{new} &= \mu_k + \alpha N\gamma(z_{nk})\left[\Sigma_k^{-1}(\mathbf{x}_n - \mu_k)\right] \\
\Sigma_k^{new} &= \Sigma_k - \alpha\frac{N}{2}\gamma(z_{nk})\left[\Sigma_k^{-1} - \Sigma_k^{-1}(\mathbf{x}_n - \mu_k^{new})(\mathbf{x}_n - \mu_k^{new})^T\Sigma_k^{-1}\right] \\
\pi_k^{new} &= \pi_k + \alpha N\left(\gamma(z_{nk})\left[\frac{1}{\pi_k}\right] - 1\right)
\end{aligned} \tag{31}
$$

整理得

$$
\begin{aligned}
\mu_k^{new} &= (I - \alpha N\gamma(z_{nk})\Sigma_k^{-1})\mu_k + \alpha N\gamma(z_{nk})\Sigma_k^{-1}\mathbf{x}_n \\
\Sigma_k^{new} &= (I - \alpha\frac{N}{2}\gamma(z_{nk})\Sigma_k^{-2})\Sigma_k + \alpha\frac{N}{2}\gamma(z_{nk})\Sigma^{-1}(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T\Sigma_k^{-1} \\
\pi_k^{new} &= (1 - \frac{\alpha N}{\pi_k})\pi_k + \frac{\alpha N\gamma(z_{nk})}{\pi_k}
\end{aligned} \tag{32}
$$

在之后的应用中，我们常假设 Σ_k 为对角阵，于是 Σ 得更新公式化为，

$$
\Sigma_k^{new} = (I - \alpha\frac{N}{2}\gamma(z_{nk})\Sigma_k^{-2})\Sigma_k + \alpha\frac{N}{2}\gamma(z_{nk})\Sigma^{-2}diag((\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T) \tag{33}
$$

其中 $diag(*)$ 表示取 $*$ 的对角元素构成对角矩阵。实际上三式都能够写成这样的形式，即

$$
\begin{aligned}
\mu_k^{new} &= (I - \rho_\mu)\mu_k + \rho_\mu\mathbf{x}_n \\
\Sigma_k^{new} &= (I - \rho_\Sigma)\Sigma_k + \rho_\Sigma diag\left((\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T\right) \\
\pi_k^{new} &= (1 - \rho_\pi)\pi_k + \rho_\pi\gamma(z_{nk})
\end{aligned} \tag{34}
$$

式34便是在线更新 μ_k, Σ_k, π_k 的迭代公式，可见迭代过程只与当前的观测值有关，学习速度与 $\rho_\mu, \rho_\Sigma, \rho_\pi$ 有关。

1.4 基于高斯混合模型的背景建模

对于个视频，每个像素的位置有 RGB 三个通道，可以把每个像素看成是维数为 3 的随机过程，我们可以将该过程用高斯混合模型去拟合，高斯分量个数 K 一般取 3 至 5，均假设协方差矩阵为对角矩阵，即不同通道的数据完全不相关。即我们对每个像素都建立一个高斯混合模型。

1.4.1 初始化 GMM

我们用视频的第一幅图片去初始化每个像素点的高斯混合模型，具体方法是：对所有高斯分量，其协方差都取同一个合适的值，均值均设为零，权重 π_k 均设为 0，再任取一高斯分量，令其均值等于该像素点的值，权重设为 1。

1.4.2 更新参数

更新参数的步骤，形式如式34所示，只是对于学习率 $\rho_\mu, \rho_\Sigma, \rho_\pi$ 的设定有所不同。算法描述如算法2。

算法2的简单理解就是，首先判断当前像素点是否属于高斯混合模型的某个分量。如果不属于任何一个，则把权重最低的高斯分量删去，用当前像素点初始化新的高斯分量；而如果属于多个，则取权重最大的那个作为匹配到的高斯分量。对匹配到的高斯分量更新 μ_k, Σ_k, π_k，更新公式为

$$\begin{aligned}
\pi_k^{new} &= (1-\alpha)\pi_k + \alpha \\
\mu_k^{new} &= (1-\rho)\mu_k + \rho \mathbf{p}_i \\
\Sigma_k^{new} &= (I-\rho)\Sigma_k + \rho diag\left((\mathbf{x}_n-\mu_k)(\mathbf{x}_n-\mu_k)^T\right)
\end{aligned} \tag{35}$$

其中 α 为学习率，ρ 定义为

$$\rho = \alpha \mathcal{N}(\mathbf{p}_i; \mu_k, \Sigma_k) \tag{36}$$

对于没有匹配到的高斯分量，只更新权重 π_k。

$$\pi_k^{new} = (1-\alpha)\pi_k \tag{37}$$

1.4.3 总结

若直接用 EM 算法对背景进行建模，计算量会非常大，而这里给出的算法，则只需要遍历一边图像就可以更新 GMM 的参数，可谓是非常高效，同样在实际操作中，也显示出了非常好的效果。但是，缺点也是有的，我们必须要设定一个 K 值，而且 GMM 的初始化也对收敛过程有很大的影响。在之后的十几年中，有很多新的方法都是针对这两个问题提出的。

Algorithm 2 对背景基于 GMM 建模的方法

初始化每个像素点的高斯混合模型
初始化学习率 α
while true **do**
  读取一张图像 $\mathbf{P}$
  for 每个像素 $\mathbf{p}_i \in \mathbf{P}$ **do**
    $index = -1, weight = 0$
    for $k \in [1, K]$ **do**
      当前像素对应的 GMM 的参数分别为 $\mu_{ik}, \Sigma_{ik}, \pi_{ik}$
      if 每个通道的像素值 $|p_{ic} - \mu_{ik}| < 2.5\Sigma_{ikcc}$ **then**
        if $\pi_{ik} > weight$ **then**
          $index = k$
          $weight = \pi_{ik}$
        end if
      end if
    end for
    if $index = -1$ **then**
      判断该像素点为前景
      删掉最小权重 π_{ik} 对应的高斯分量
      用当前像素值 $\mathbf{p}_i$ 初始化新的高斯分量
    else
      判断该像素为背景
      for $k \in [1, K]$ **do**
        $\pi_{ik}^{new} = (1-\alpha)\pi_{ik} + \alpha(M_{ik})$
        其中 M_{ik} 仅在 $k = index$ 时为 1，其余为 0
      end for
      令 $k = index$
      $\rho = \alpha\mathcal{N}(\mathbf{p}_i; \mu_k, \Sigma_k)$
      更新 $\mu_k^{new} = (1-\rho)\mu_k + \rho\mathbf{p}_i$
      更新 $\Sigma_k^{new} = (I-\rho)\Sigma_k + \rho diag\left((\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T\right)$
    end if
  end for
end while