EM算法与混合高斯模型

【参考资料】

吴恩达 CS229

周志华《机器学习》

[详解EM算法与混合高斯模型(Gaussian mixture model, GMM)]

1. 混合高斯模型

首先回顾一下一维高斯分布的概率密度函数：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

其中μ和σ^2分别为均值和方差。

多维变量$X=(x_1,x_2,\ldots,x_n)$的联合概率密度函数为：

$$f(X)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(X-u)^T\Sigma^{-1}(X-u)\right]$$

其中，n为变量维数，$u=(u_1,u_2,\ldots,u_n)$为均值，Σ为协方差矩阵。

混合高斯模型（Gaussian mixture model，GMM）就是多个高斯分布的合成。假设组成GMM的高斯分布有K个，则GMM的概率密度函数如下：

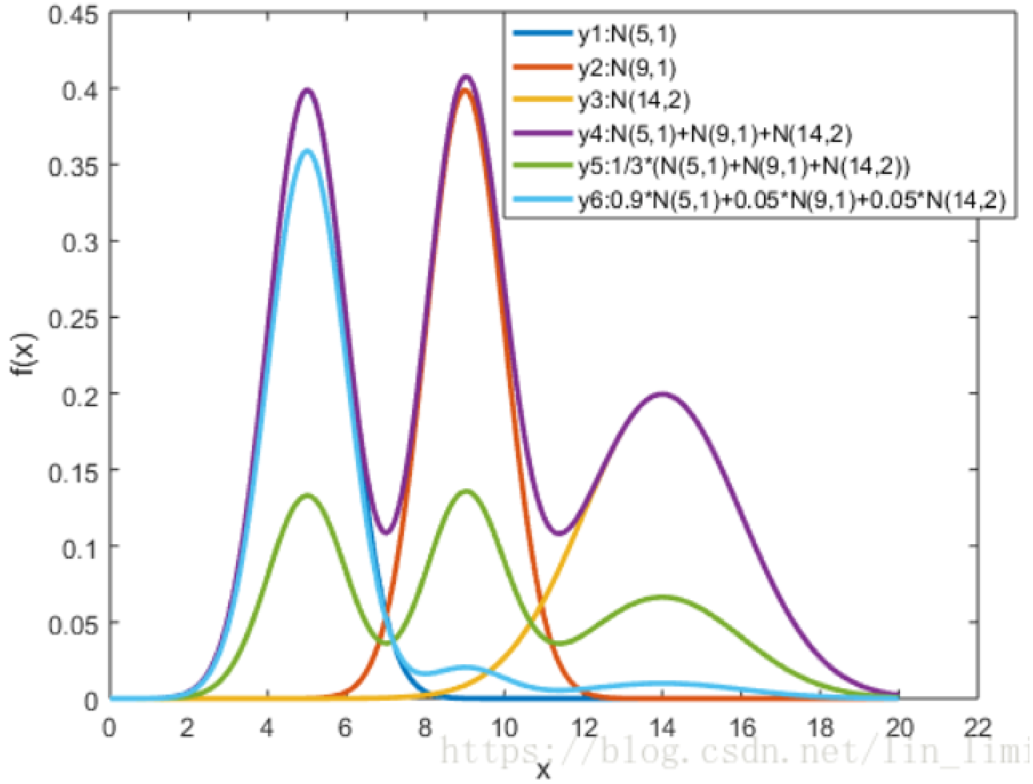
$$p(x)=\sum_{j=1}^{K}p(j)p(x|j)=\sum_{j=1}^{K}\phi_j N(x|u_j,\Sigma_j)$$

其中$p(x|j)=N(x|u_j,\Sigma_j)$是第j个高斯分布的概率密度函数，可以看成是选定第j个高斯分布后，该高斯分布产生x的概率；$p(j)=\phi_j$是第j个高斯分布的权重，也可以视为选择第j个高斯分布的概率，并且满足$\sum_{j=1}^{K}\phi_j=1$。

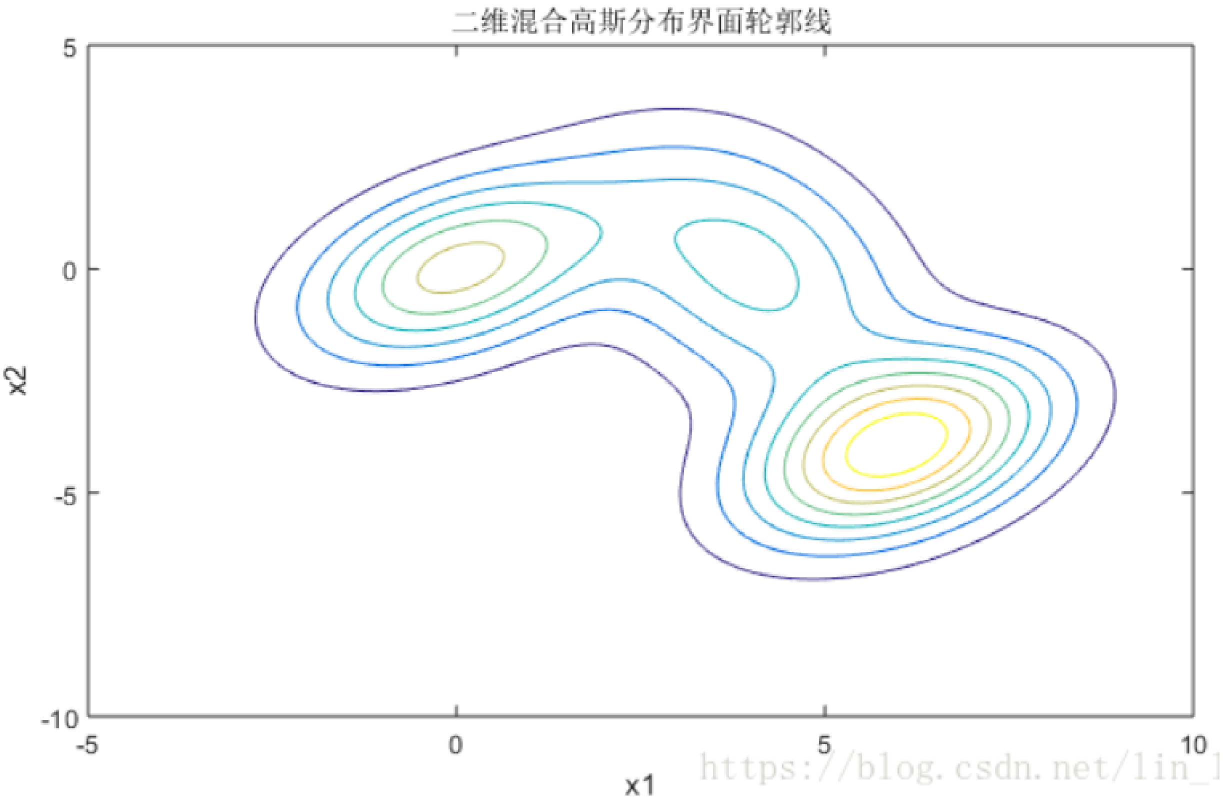
所以，混合高斯模型并不是什么新奇的东西，它的本质就是融合几个单高斯模型，来使得模型更加复杂，从而产生更复杂的样本。理论上，如果某个混合高斯模型融合的高斯模型个数足够多，它们之间的权重设定得足够合理，这个混合模型可以拟合任意分布的样本。

GMM的图例如下

一维的情形：



二维空间中三个高斯分布的混合：



2. EM算法估计参数

GMM的参数有各个高斯成分对应的均值μ与协方差Σ，以及它们的权重ϕ。在估计GMM模型参数时的一个问题是，我们无法确定观测到的样本x具体是由哪一个高斯成分产生的。换句话说，“哪一个高斯分布”是一个隐变量，我们把它用z来表示。这时候，就需要用到EM算法来估计GMM的参数。

对于特定的样本i，我们假设z服从一个参数为ϕ的多项式分布，并且$\phi_j \geq 0$，$\sum_{j=1}^{k} \phi_j = 1$，其中$j$表示单个高斯成分的序号。于是有：

$$p\left(z^{(i)}=j\right)=\phi_j$$

表示选择第j个高斯分布的概率是ϕ_j，并且我们有：

$$x^{(i)} \left| z^{(i)}=j \sim \mathcal{N}\left(\mu_j, \Sigma_j\right)\right.$$

即$x^{(i)}$是由这第j个高斯分布产生的。所以，总的似然函数为：

$$\begin{aligned}
\ell(\phi, \mu, \Sigma) &= \sum_{i=1}^{m} \log p\left(x^{(i)} ; \phi, \mu, \Sigma\right) \\
&= \sum_{i=1}^{m} \log \sum_{z^{(i)}=1}^{k} p\left(x^{(i)} | z^{(i)} ; \mu, \Sigma\right) p\left(z^{(i)} ; \phi\right)
\end{aligned}$$

所以，我们需要估计的参数是ϕ_j，μ_j，Σ_j。接下来就是调用EM算法的步骤，首先初始化一组ϕ_j，μ_j，Σ_j。然后在**E-Step**中，我们利用贝叶斯公式，计算$z^{(i)}$的后验概率：

$$\begin{aligned}
w_j^{(i)} &= Q_i\left(z^{(i)}=j\right)=P\left(z^{(i)}=j | x^{(i)} ; \phi, \mu, \Sigma\right) \\
&= \frac{p\left(x^{(i)} | z^{(i)}=j ; \mu, \Sigma\right) p\left(z^{(i)}=j ; \phi\right)}{\sum_{l=1}^{k} p\left(x^{(i)} | z^{(i)}=l ; \mu, \Sigma\right) p\left(z^{(i)}=l ; \phi\right)}
\end{aligned}$$

$w_j^{(i)}$就代表了我们对$z^{(i)}$的选择的一种“猜测”，并且因为计算的是概率，所以这是一种“soft”的、不确定的猜测。

接下来，利用我们对$z^{(i)}$的猜测，在**M-Step**中更新参数，也就是说，我们要最大化下面的似然函数：

$$\begin{aligned}
&\sum_{i=1}^{m} \sum_{z} Q_i\left(z^{(i)}\right) \log \frac{p\left(x^{(i)}, z^{(i)} ; \phi, \mu, \Sigma\right)}{Q_i\left(z^{(i)}\right)} \\
&\quad= \sum_{i=1}^{m} \sum_{j=1}^{k} Q_i\left(z^{(i)}=j\right) \log \frac{p\left(x^{(i)} | z^{(i)}=j ; \mu, \Sigma\right) p\left(z^{(i)}=j ; \phi\right)}{Q_i\left(z^{(i)}=j\right)} \\
&\quad= \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \log \frac{\frac{1}{(2\pi)^{n/2}\left|\Sigma_j\right|^{1/2}} \exp\left(-\frac{1}{2}\left(x^{(i)}-\mu_j\right)^T \Sigma_j^{-1}\left(x^{(i)}-\mu_j\right)\right) \cdot \phi_j}{w_j^{(i)}}
\end{aligned}$$

为了更新参数μ_j，我们把上面的似然函数对μ_j求导：

$$
\begin{aligned}
&\nabla_{\mu_j} \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \log \frac{\frac{1}{(2\pi)^{n/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}\left(x^{(i)}-\mu_j\right)^T \Sigma_j^{-1}\left(x^{(i)}-\mu_j\right)\right) \cdot \phi_j}{w_j^{(i)}} \\
&\quad = -\nabla_{\mu_j} \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \frac{1}{2}\left(x^{(i)}-\mu_j\right)^T \Sigma_j^{-1}\left(x^{(i)}-\mu_j\right) \\
&\quad = \sum_{i=1}^{m} w_j^{(i)}\left(\Sigma_j^{-1} x^{(i)}-\Sigma_j^{-1} \mu_j\right)
\end{aligned}
$$

令其等于0，可得

$$
\mu_j := \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}
$$

即各混合成分的均值可通过样本加权平均来估计，样本权重是属于该成分的后验概率。

接下来推导ϕ_j的更新规则，在似然函数中，与ϕ_j有关的项为

$$
\sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \log \phi_j
$$

但是，由于ϕ_j是一个概率，所以它还需要满足$\sum_{j=1}^{k} \phi_j = 1$的约束条件，因此构造拉格朗日算子为

$$
\mathcal{L}(\phi) = \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \log \phi_j + \beta\left(\sum_{j=1}^{k} \phi_j - 1\right)
$$

其中β是拉格朗日乘子。对上式求导，可得

$$
\frac{\partial}{\partial \phi_j} \mathcal{L}(\phi) = \sum_{i=1}^{m} \frac{w_j^{(i)}}{\phi_j} + \beta
$$

将其置零，可得

$$
\phi_j = \frac{\sum_{i=1}^{m} w_j^{(i)}}{-\beta}
$$

利用约束$\sum_j \phi_j = 1$，我们可以将上式改写为：

$$
-\beta = \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} = \sum_{i=1}^{m} 1 = m
$$

这里利用了$\sum_j w_j^{(i)} = 1$。

于是ϕ_j的更新就可以表示为

$$\phi_j := \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)}$$

即每个高斯成分的混合系数由样本属于该成分的平均后验概率确定。

同理，可以得到Σ_j的更新规则。

各个参数的更新规则总结如下：

$$\begin{aligned}
\phi_j &:= \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)} \\
\mu_j &:= \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}} \\
\Sigma_j &:= \frac{\sum_{i=1}^{m} w_j^{(i)} \left(x^{(i)} - \mu_j\right)\left(x^{(i)} - \mu_j\right)^T}{\sum_{i=1}^{m} w_j^{(i)}}
\end{aligned}$$