GMM model:

由K个 Gaussian 分布组成，每个称为一个 component，线性加成

一个 GMM 的概率密度函数

$$p(x)=\sum_{k=1}^{K}\pi_k\cdot N(x|\mu_k,\Sigma_k)$$

$\pi=[\pi_1,\dots\pi_K]$，$\sum_i^K\pi_i=1$

(μ_k,Σ_k) 是第k个 component 的 mean var

GMM 去生成数

① 选一个 component，每个 component 被选中的概率为 π_k.

② 从这个 component 中选一个点.. $(x_i\sim N(\mu_k,\Sigma_k))$

GMM 做 clustering：假设 data 是从 GMM 生成出来的，那ln 反推 GMM 的概率分布就可. GMM 的k个 component 对应了 K 个 clusters.

Goal：given $x_1,\dots x_n$, learn $\pi_k, \mu_k \Sigma_k$

如何参数估计：MLE.

$$\max p(x_1,\dots x_n|\pi,\mu,\Sigma)$$

$$=\max_{\pi,\mu,\Sigma}\prod_{i=1}^{n}p(x_i|\pi,\mu,\Sigma)$$

$$=\max_{\pi,\mu,\Sigma}\prod_{i=1}^{n}\sum_{k=1}^{K}p(x_i,k|\pi,\mu,\Sigma)$$

取 log：即 $\max\sum_{i=1}^{n}\ln\sum_{k=1}^{K}\pi_k\cdot N(x_i|\mu_k,\Sigma_k)$

求解：

$$\mathcal{L}=\sum_{i=1}^{n}\sum_{k=1}^{K}\phi_i(k)\{\ln\pi_k+\ln N(x_i|\mu_k,\Sigma_k)\}+C$$

①. 估计 data 由每个 component 生成的概率（不是 component 选中概率）

对 x_i，它由第k个 component 生成的概率为 $\phi_i(k)=\dfrac{\pi_k\cdot N(x_i|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j\cdot N(x_i|\mu_j,\Sigma_j)}$

后验prob

$$N(x|\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{D}{2}}}\cdot\frac{1}{|\Sigma|^{\frac{1}{2}}}\cdot e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

② 通过极大似然估计，令导数 =0，$\Rightarrow$

$$\mu_k=\frac{1}{n_k}\sum_{i=1}^{n}\phi_i(k)\cdot x_i$$

$$\Sigma_k=\frac{1}{n_k}\sum_{i=1}^{n}\phi_i(k)\cdot(x_i-\mu_k)(x_i-\mu_k)^T$$

其中 $n_k=\sum_{i=1}^{n}\phi_i(k)$

$$\pi_k=\frac{n_k}{n}$$

循环直到 $\mathcal{L}$ 收敛

GMM和 Bayes classifier

GMM feels like k-class Bayes classifier.

label of x_i is label $(x_i) = \arg\max_k \pi_k \cdot \underline{N(x_i \mid \mu_k, \Sigma_k)}$

估计x_i由第k个component生成的后验prob.

class prior

class-condition density function

更新: $\pi_k = \frac{n_k}{n}$ $\mu_k = \frac{\sum_{i=1}^{n} \phi_i(k) \cdot x_i}{n_k}$ $\Sigma_k = \frac{1}{n_k} \cdot \sum_{i=1}^{n} \phi_i(k) \cdot (x_i - \mu_k)(x_i - \mu_k)^T$

与前面极相似! 但这里$\phi_i(k)$是知道的! 第i个点的类别.

Until: 似然函数收敛.