人工智能之机器学习

EM算法

上海育创网络科技股份有限公司

主讲人：刘老师(GerryLiu)

课程要求

- 课上课下“九字”真言
  - 认真听，**善摘录，勤思考**
  - **多温故，乐实践**，再发散
- 四不原则
  - 不懒散惰性，不迟到早退
  - 不请假旷课，不拖延作业
- 一点注意事项
  - 违反“四不原则”，不推荐就业

课程内容

- 最大似然估计
- K-means算法
- EM算法
- GMM算法

最大似然估计(MLE)回顾
Maximum Likelihood Estimation

- **MLE就是利用已知的样本结果，反推最有可能(最大概率)导致这样结果的参数值的计算过程。**直白来讲，就是给定了一定的数据，假定知道数据是从某种分布中随机抽取出来的，但是不知道这个分布具体的参数值，即“模型已定，参数未知”，MLE就可以用来估计模型的参数。**MLE的目标是找出一组参数(模型中的参数)，使得模型产出观察数据的概率最大。**

$$\arg\max_{\theta} p(X;\theta)$$

最大似然估计(MLE)回顾
Maximum Likelihood Estimation

- 假定盒子中有黑白两种球，数目未知，黑白球比例也未知，现只知道随机的十次有放回的抽样情况，求各个盒子中抽取出白球的概率?

盒子编号	1	2	3	4	5	6	7	8	9	10
1	黑	黑	黑	黑	黑	黑	黑	黑	黑	黑
2	黑	黑	白	白	黑	黑	黑	白	黑	黑
3	黑	白	黑	黑	白	黑	白	黑	白	白
4	白	黑	白	白	黑	白	黑	白	白	白
5	白	白	白	白	白	白	白	白	白	白

最大似然估计(MLE)回顾
Maximum Likelihood Estimation

- MLE求解过程：
  - 编写似然函数(即联合概率函数)<似然函数：在样本固定的情况下，样本出现的概率与参数θ之间的函数>；
  - 对似然函数取对数，并整理；(一般都进行)
  - 求导数；
  - 解似然方程。

$$L(X;\theta)\rightarrow l(\theta)=\ln(L(X;\theta))\rightarrow\frac{\partial l}{\partial\theta}$$

$$\arg\max_{p} P(x_i;p)=\arg\max_{p}\ln(P(x_i;p))$$

最大似然估计(MLE)回顾
Maximum Likelihood Estimation

- 盒子中只有黑球和白球，假定白球的比例为p，那么黑球的比例为1-p。因为采取的是有放回的随机抽取，那么每次抽取出来的球的颜色服从同一独立分布情况，即每次抽取之间是独立互不影响的。

$$L(X;p)=P(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10};p)=\prod_{i=1}^{10}P(x_i;p)$$

$$l(p)=\ln(L(X;p))=\sum_{i=1}^{10}P(x_i;p)$$

$$\frac{\partial l}{\partial p}=0\rightarrow p=?$$

最大似然估计(MLE)回顾
Maximum Likelihood Estimation

IBEIFENG.COM

- 盒子1中抽取出白球的概率：

$$L(X;p)=(1-p)^{10} \quad l(p)=10\ln(1-p) \xrightarrow{0\le p\le 1} p=0$$

- 盒子2中抽取出白球的概率：

$$L(X;p)=p^3(1-p)^7 \quad l(p)=3\ln p+7\ln(1-p) \qquad \frac{\partial l}{\partial p}=\frac{3}{p}-\frac{7}{1-p}$$

$$\xrightarrow{\frac{\partial l}{\partial p}=0} p=0.3$$

- 盒子3中抽取出白球的概率：

$$L(X;p)=p^5(1-p)^5 \quad l(p)=5\ln p+5\ln(1-p) \qquad \frac{\partial l}{\partial p}=\frac{5}{p}-\frac{5}{1-p}$$

$$\xrightarrow{\frac{\partial l}{\partial p}=0} p=0.5$$

最大似然估计(MLE)回顾
Maximum Likelihood Estimation

- 盒子4中抽取出白球的概率：

$$L(X;p)=p^7(1-p)^3 \quad l(p)=7\ln p+3\ln(1-p) \quad \frac{\partial l}{\partial p}=\frac{7}{p}-\frac{3}{1-p}$$

$$\xrightarrow{\frac{\partial l}{\partial p}=0} p=0.7$$

- 盒子5中抽取出白球的概率：

$$L(X;p)=p^{10} \quad l(p)=10\ln p \quad \xrightarrow{0\le p\le 1} p=1$$

盒子编号	白球	黑球
1	0	1
2	0.3	0.7
3	0.5	0.5
4	0.7	0.3
5	1	0

贝叶斯算法估计

- 贝叶斯算法估计是一种从先验概率和样本分布情况来计算后验概率的一种方式。
- 贝叶斯算法中的常见概念：
  - P(A)是事件A的先验概率或者边缘概率；
  - P(A|B)是已知B发生后A发生的条件概率，也称为A的后验概率；
  - P(B|A)是已知A发生后B发生的条件概率，也称为B的后验概率；
  - P(B)是事件B的先验概率或者边缘概率。

$$P(AB)=P(A)P(B|A)=P(B)P(A|B) \Longrightarrow P(A|B)=\frac{P(A)P(B|A)}{P(B)}$$

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

贝叶斯算法估计

- 现在有五个盒子，假定每个盒子中都有黑白两种球，并且黑白球的比例如下；现已知从这五个盒子中的任意一个盒子中有放回的抽取两个球，且均为白球，问这两个球是从哪个盒子中抽取出来的?

盒子编号	白球(p)	黑球(q)
1	0	1
2	0.3	0.7
3	0.5	0.5
4	0.7	0.3
5	1	0

贝叶斯算法估计

- 使用MLE(最大似然估计)，结论是从第五个盒子抽取的球：

$$L(X;p)=p^2 \xrightarrow[p]{\arg\max L(X;p)} p=1$$

- 使用贝叶斯算法估计，结论是从第五个盒子抽取的球：

假定抽出白球为事件B，从第i个盒子中抽取为事件A_i。

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{P(B)}=\frac{0.2*0*0}{0.366}=0$$

$$P(A_2|B)=0.049 \quad P(A_3|B)=0.137$$

$$P(A_4|B)=0.268 \quad P(A_5|B)=0.546$$

盒子编号	白球(p)	黑球(q)
1	0	1
2	0.3	0.7
3	0.5	0.5
4	0.7	0.3
5	1	0

	1	2	3	4	5
P(A)	0.2	0.2	0.2	0.2	0.2

贝叶斯算法估计

- 现在不是从五个盒子中任选一个盒子进行抽取，而是按照一定的概率选择对应的盒子，概率如下。假定抽出白球为事件B，从第i个盒子中抽取为事件A_i。结论是从第四个盒子抽取的球。

盒子编号	白球(p)	黑球(q)
1	0	1
2	0.3	0.7
3	0.5	0.5
4	0.7	0.3
5	1	0

	1	2	3	4	5
P(A)	0.1	0.2	0.2	0.4	0.1

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{P(B)}=\frac{0.1*0*0}{0.364}=0$$

$$P(A_2|B)=0.049 \qquad P(A_3|B)=0.137$$

$$P(A_4|B)=0.538 \qquad P(A_5|B)=0.275$$

最大后验概率估计(MAP)
Maximum a posteriori estimation

- MAP和MLE一样，都是通过样本估计参数θ的值；在MLE中，是使似然函数P(x|θ)最大的时候参数θ的值，MLE中假设先验概率是一个等值的；而在MAP中，则是求θ使P(x|θ)P(θ)的值最大，这也就是要求θ值不仅仅是让似然函数最大，同时要求θ本身出现的先验概率也得比较大。
- 可以认为MAP是贝叶斯算法的一种应用

$$P(\theta'|X)=\frac{P(\theta')P(X|\theta')}{P(X)} \quad \to \quad \arg\max_{\theta'} P(\theta'|X) \to \arg\max_{\theta'} P(\theta')P(X|\theta')$$

K-means算法回顾

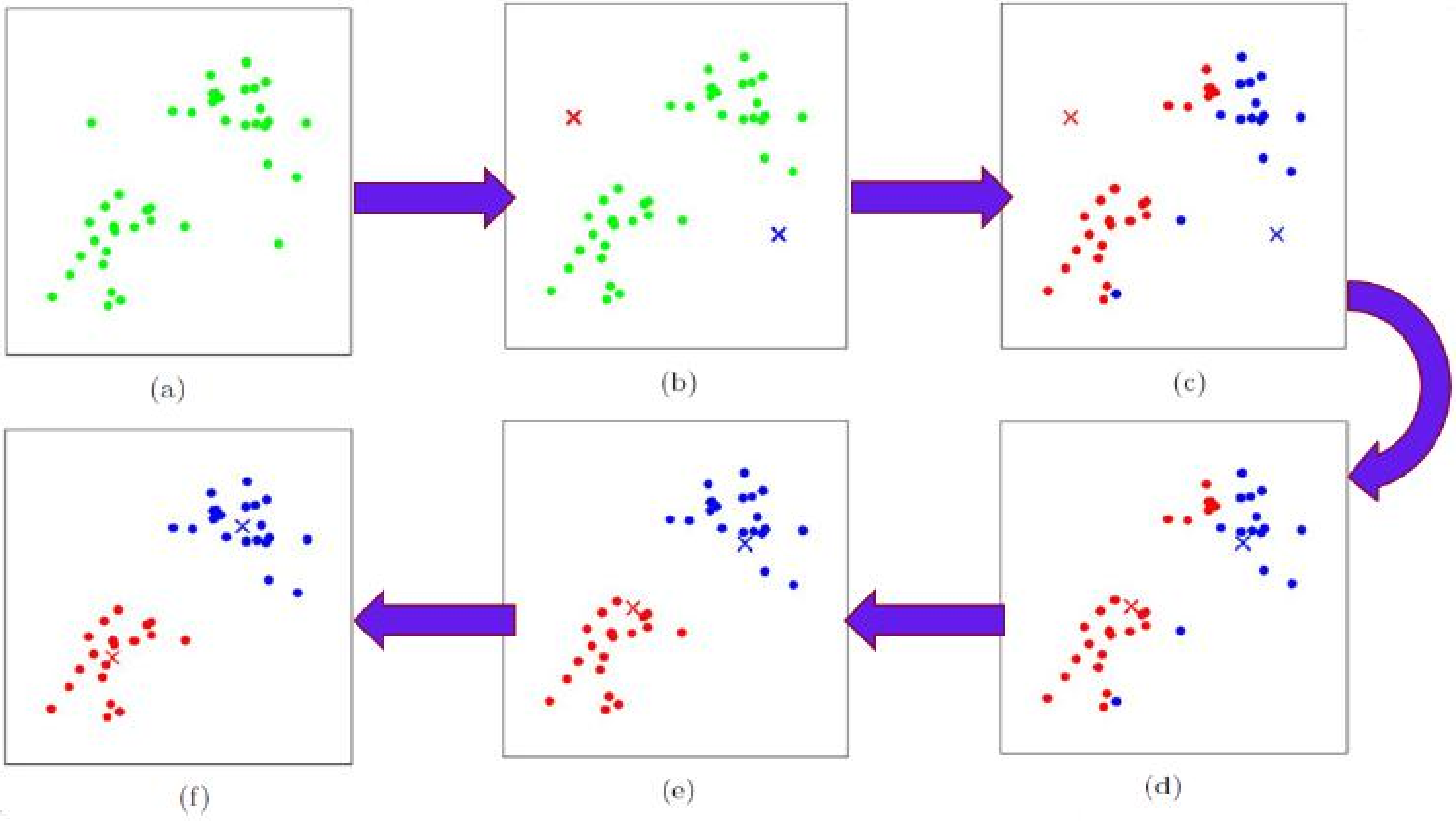
- K-means算法，也称为k-均值聚类算法，是一种非常广泛使用的聚类算法之一。
- 假定输入样本为S=$x_1,x_2,x_3,\ldots,x_m$,则算法步骤为：
  - 选择初始的k个簇中心点$\mu_1\mu_2\ldots\mu_k$；
  - 将样本X_i标记为距离簇中心最近的簇: $label_i$；
  - 迭代处理所有样本数据，计算出各个样本点所属的对应簇；
  - 更新簇中心点坐标：μ_i；
  - 重复上述三个操作，直到算法收敛

$$label_i = \arg\min_{1\le j\le k} \left\| x_i - \mu_j \right\|$$

$$\mu_i = \frac{1}{|C_i|}\sum_{j\in c_i} x_j$$

- 算法收敛条件：迭代次数/簇中心变化率/MSE/MAE。
- 备注：每次找到的只是当前情况下的最优解，每次迭代都会改变数据的分布情况

K-means算法回顾



K-means算法回顾

$$J(k,\mu)=\frac{1}{m}\sum_{i=1}^{m}\left\|x^{(i)}-\mu_{k^{(i)}}\right\|^{2}$$

$$k,\mu=\arg\min_{k,\mu} J(k,\mu)$$

EM算法

- EM算法(Expectation Maximization Algorithm, 最大期望算法)是一种迭代类型的算法，是一种在概率模型中寻找参数最大似然估计或者最大后验估计的算法，其中概率模型依赖于无法观测的隐藏变量。
- EM算法流程：
  - 初始化分布参数/模型参数
  - 重复下列两个操作直到收敛：
    - E步骤：估计隐藏变量的概率分布期望函数；
    - M步骤：根据期望函数重新估计分布参数。

EM算法原理

- 给定的m个训练样本$\{x^{(1)},x^{(2)},\ldots,x^{(m)}\}$，样本间独立，找出样本的模型参数θ，极大化模型分布的对数似然函数如下：

$$\theta = \arg\max_{\theta} \sum_{i=1}^{m} \log\left(P\left(x^{(i)};\theta\right)\right)$$

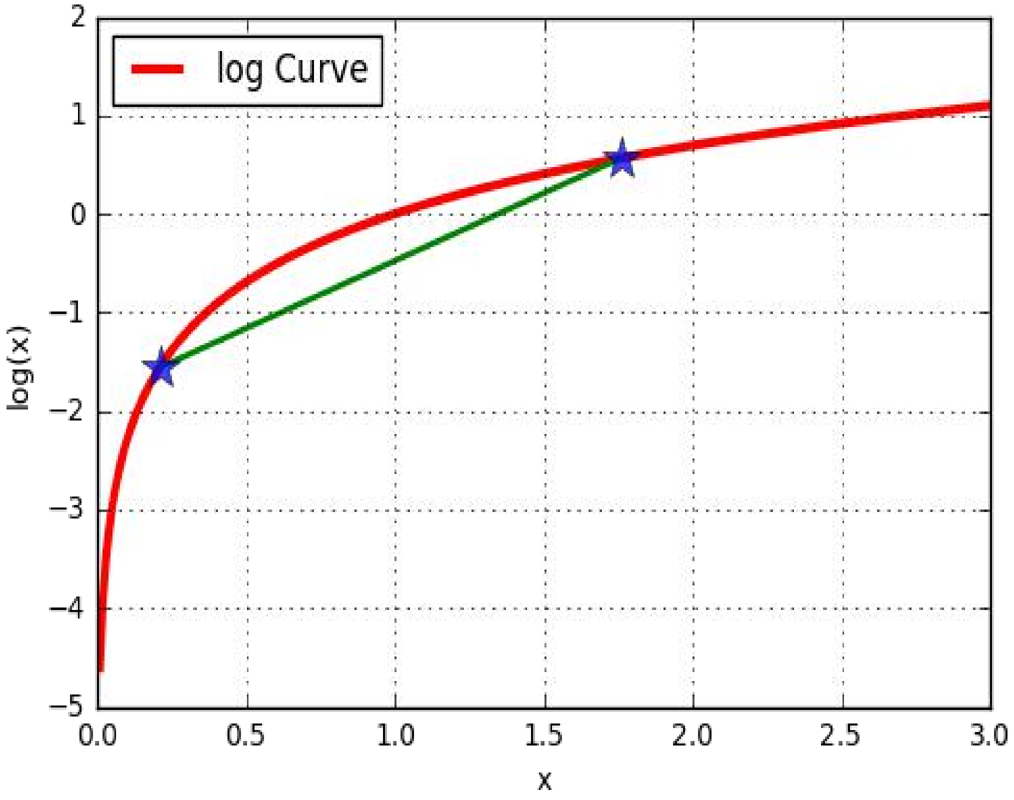
- 假定样本数据中存在隐含数据$z=\{z^{(1)},z^{(2)},\ldots,z^{(k)}\}$，此时极大化模型分布的对数似然函数如下：

$$\begin{aligned}
\theta &= \arg\max_{\theta} \sum_{i=1}^{m} \log\left(P\left(x^{(i)};\theta\right)\right) \\
&= \arg\max_{\theta} \sum_{i=1}^{m} \log\left(\sum_{z^{(i)}} P\left(z^{(i)}\right) P\left(x^{(i)} \middle| z^{(i)};\theta\right)\right) \\
&= \arg\max_{\theta} \sum_{i=1}^{m} \log\left(\sum_{z^{(i)}} P\left(x^{(i)},z^{(i)};\theta\right)\right)
\end{aligned}$$

EM算法原理

$$\sum_z Q(z;\theta)=1$$

- 令z的分布为Q(z;θ)，并且Q(z;θ)≥0；那么有如下公式：



$$l(\theta)=\sum_{i=1}^{m}\log\sum_z p(x,z;\theta)$$

$$=\sum_{i=1}^{m}\log\sum_z Q(z;\theta^{old})\cdot\frac{p(x,z;\theta)}{Q(z;\theta^{old})}$$

$$=\sum_{i=1}^{m}\log\left(E_Q\left(\frac{p(x,z;\theta)}{Q(z;\theta^{old})}\right)\right)$$

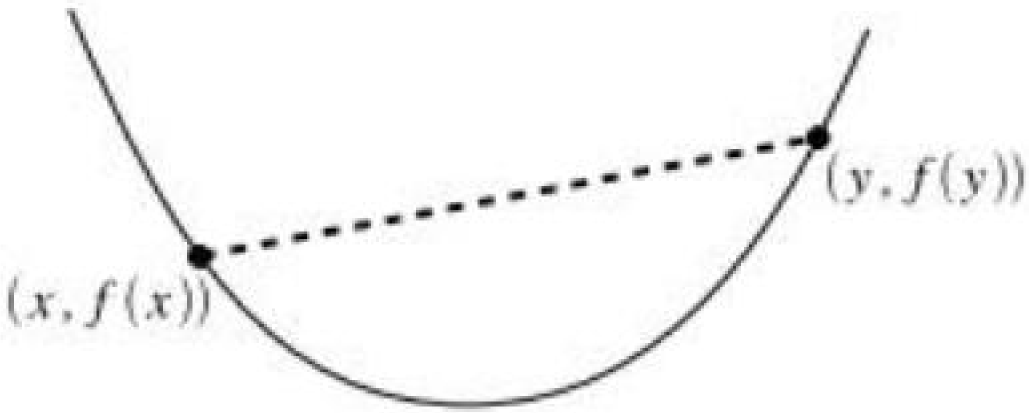
$$\ge\sum_{i=1}^{m}E_Q\left(\log\left(\frac{p(x,z;\theta)}{Q(z;\theta^{old})}\right)\right)$$

$$=\sum_{i=1}^{m}\sum_z Q(z;\theta^{old})\log\left(\frac{p(x,z;\theta)}{Q(z;\theta^{old})}\right)$$

Jensen不等式

- 如果函数f为凸函数，那么存在下列公式：

$$f(\theta x+(1-\theta)y)\le \theta f(x)+(1-\theta)f(y)$$



- 若$\theta_1,\ldots,\theta_k \geq 0$，$\theta_1+\ldots+\theta_k=1$;则

$$f(\theta_1 x_1+\ldots+\theta_k x_k)\le \theta_1 f(x_1)+\ldots+\theta_k f(x_k)$$

$$f(E(x))\le E(f(x))$$

EM算法原理

$$l(\theta) \geq \sum_{i=1}^{m}\sum_{z} Q(z;\theta^{old})\log\left(\frac{p(x,z;\theta)}{Q(z;\theta^{old})}\right)$$

- 根据Jensen不等式的特性，当下列式子的值为常数的时候，l(θ)函数才能取等号。

$$\frac{p(x,z;\theta^{old})}{Q(z;\theta^{old})} = c, \quad \forall x, \forall z$$

$$\sum_{z} Q(z;\theta^{old}) = 1$$

$$Q(z,\theta^{old}) = \frac{p(x,z;\theta^{old})}{c} = \frac{p(x,z;\theta^{old})}{c \cdot \sum_{z^i} Q(z^i;\theta^{old})}$$

$$= \frac{p(x,z;\theta^{old})}{\sum_{z^i} c \cdot Q(z^i;\theta^{old})} = \frac{p(x,z;\theta^{old})}{\sum_{z^i} p(x,z^i;\theta^{old})} = \frac{p(x,z;\theta^{old})}{p(x;\theta^{old})} = p(z \mid x;\theta^{old})$$

EM算法原理

北风网 IBEIFENG.COM

$$\theta=\arg\max_{\theta} l(\theta)=\arg\max_{\theta}\sum_{i=1}^{m}\sum_{z}Q\left(z;\theta^{\text{old}}\right)\log\left(\frac{p(x,z;\theta)}{Q\left(z;\theta^{old}\right)}\right)$$

$$=\arg\max_{\theta}\sum_{i=1}^{m}\sum_{z}p\left(z\mid x;\theta^{old}\right)\log\left(\frac{p(x,z;\theta)}{p\left(z\mid x;\theta^{old}\right)}\right)$$

$$\cong\arg\max_{\theta}\sum_{i=1}^{m}\sum_{z}p\left(z\mid x;\theta^{old}\right)\log\left(p(x,z;\theta)\right)$$

EM算法流程

- 条件：样本数据$x=\{x^1,x^2,\dots,x^m\}$，联合分布$p(x,z;\theta)$，条件分布$p(z|x;\theta)$，最大迭代次数J
  - 1. 随机初始化模型参数θ的初始值θ^0
  - 2. 开始EM算法的迭代处理：
    - E步：计算联合分布的条件概率期望

$$Q^j=p\left(z\middle|x;\theta^j\right) \qquad l(\theta)=\sum_{i=1}^{m}\sum_{z}Q^j\log\left(p(x,z;\theta)\right)$$

    - M步：极大化L函数，得到θ^{j+1}

$$\theta^{j+1}=\arg\max_{\theta} l(\theta)$$

    - 如果θ^{j+1}已经收敛，则算法结束，输出最终的模型参数θ，否则继续迭代处理

EM算法直观案例

- 假设现有两个装有不定数量黑球、白球的盒子，随机从盒子中抽取出一个白球的概率分布为p_1和p_2；为了估计这两个概率，每次选择一个盒子，有放回的连续随机抽取5个球，记录如下：

盒子编号	1	2	3	4	5	统计
1	白	白	黑	白	黑	3白-2黑
2	黑	黑	白	白	黑	2白-3黑
1	白	黑	黑	黑	黑	1白-4黑
2	白	黑	白	黑	白	3白-2黑
1	黑	白	黑	白	黑	2白-3黑

EM算法直观案例

- 使用MLE最大似然估计：

$$l(p_1)=\log\left(p_1^6(1-p_1)^9\right)=6\log p_1+9\log(1-p_1)$$

$$\frac{\partial l(p_1)}{\partial p_1}=\frac{6}{p_1}-\frac{9}{1-p_1}\xrightarrow{\text{令}\frac{\partial l(p_1)}{\partial p_1}=0}p_1=0.4$$

$$l(p_2)=\log\left(p_2^5(1-p_2)^5\right)=5\log p_2+5\log(1-p_2)$$

$$\frac{\partial l(p_2)}{\partial p_2}=\frac{5}{p_2}-\frac{5}{1-p_2}\xrightarrow{\text{令}\frac{\partial l(p_2)}{\partial p_2}=0}p_2=0.5$$

EM算法直观案例

- 如果现在不知道具体的盒子编号，但是同样还是为了求解p_1和p_2的值，这个时候就相当于多了一个隐藏变量z, z表示的是每次抽取的时候选择的盒子编号，比如z1就表示第一次抽取的时候选择的是盒子1还是盒子2。

盒子编号	1	2	3	4	5	统计
z1	白	白	黑	白	黑	3白-2黑
z2	黑	黑	白	白	黑	2白-3黑
z3	白	黑	黑	黑	黑	1白-4黑
z4	白	黑	白	黑	白	3白-2黑
z5	黑	白	黑	白	黑	2白-3黑

EM算法直观案例

- 随机初始一个概率值：p_1=0.1和p_2=0.9；然后使用最大似然估计计算每轮操作中从两个盒子中抽取的最大概率。然后计算出来的z值，重新使用极大似然估计法则估计概率值。

$$L(z_1=1|x;p_1)=0.1^3*0.9^2=0.00081$$

$$L(z_1=2|x;p_2)=0.9^3*0.1^2=0.00729$$

$$p_1=\frac{1}{3};p_2=0.6$$

轮数	盒子1	盒子2
1	0.00081	**0.00729**
2	**0.00729**	0.00081
3	**0.06561**	0.00009
4	0.00081	**0.00729**
5	**0.00729**	0.00081

轮数	盒子1	盒子2
1	0.01565	**0.03456**
2	**0.0313**	0.02304
3	**0.0626**	0.01536
4	0.01565	**0.03456**
5	**0.0313**	0.02304

EM算法直观案例

- 使用最大似然概率法则估计z和p的值，但是在这个过程中，只使用一个最有可能的值。如果考虑所有的z值，然后对每一组z值都估计一个概率p，那么这个时候估计出来的概率可能会更好，可以用期望的方式来简化这个操作。

轮数	盒子1	盒子2
1	0.00081	0.00729
2	0.00729	0.00081
3	0.06561	0.00009
4	0.00081	0.00729
5	0.00729	0.00081

轮数	盒子1概率	盒子2概率
1	0.1	0.9
2	0.9	0.1
3	0.999	0.001
4	0.1	0.9
5	0.9	0.1

EM算法直观案例

- 以p_1估计为例，计算如下：

$$l(p_1)=\log\left(p_1^{5.199}(1-p_1)^{9.796}\right)=5.199\log p_1+9.796\log(1-p_1)$$

$$\frac{\partial l(p_1)}{\partial p_1}=\frac{5.199}{p_1}-\frac{9.796}{1-p_1}\xrightarrow[\partial p_1]{\text{令}\frac{\partial l(p_1)}{\partial p_1}=0}p_1=0.347$$

轮数	盒子1概率	盒子2概率
1	0.1	0.9
2	0.9	0.1
3	0.999	0.001
4	0.1	0.9
5	0.9	0.1

轮数	盒子1白球	盒子1黑球
1	0.3	0.2
2	1.8	2.7
3	0.999	3.996
4	0.3	0.2
5	1.8	2.7
总计	5.199	9.796

EM算法直观案例

- 计算出p_1和p_2的概率值后，再次计算从每个盒子中抽取的概率如下：

$$p_1 = 0.347 \qquad p_2 = 0.58$$

轮数	盒子1	盒子2
1	0.0178	0.0344
2	0.0335	0.0249
3	0.063	0.018
4	0.0178	0.0344
5	0.0335	0.0249

轮数	盒子1概率	盒子2概率
1	0.34	0.66
2	0.57	0.43
3	0.78	0.22
4	0.34	0.66
5	0.57	0.43

EM算法直观案例

- 再次计算概率值如下：

$$p_1 = 0.392 \qquad p_2 = 0.492$$

轮数	盒子1白球	盒子1黑球
1	1.02	0.68
2	1.14	1.71
3	0.78	3.12
4	1.02	0.68
5	1.14	1.71
总计	5.1	7.9

轮数	盒子2白球	盒子2黑球
1	1.98	1.32
2	0.86	1.29
3	0.22	0.88
4	1.98	1.32
5	0.86	1.29
总计	5.9	6.1

EM算法收敛证明

- EM算法的收敛性只要我们能够证明对数似然函数的值在迭代的过程中是增加的即可

$$\sum_{i=1}^{m}\log\left(p\left(x^{i};\theta^{j+1}\right)\right)\geq\sum_{i=1}^{m}\log\left(p\left(x^{i};\theta^{j}\right)\right)$$

EM算法收敛证明

$$L\left(\theta,\theta^{j}\right)=\sum_{i=1}^{m}\sum_{z}p\left(z\middle|x^{i};\theta^{j}\right)\log p\left(x^{i},z;\theta\right)$$

$$H\left(\theta,\theta^{j}\right)=\sum_{i=1}^{m}\sum_{z}p\left(z\middle|x^{i};\theta^{j}\right)\log p\left(z\middle|x^{i};\theta\right)$$

$$L\left(\theta,\theta^{j}\right)-H\left(\theta,\theta^{j}\right)=\sum_{i=1}^{m}\log p\left(x^{i};\theta\right)$$

$$\left[L\left(\theta^{j+1},\theta^{j}\right)-L\left(\theta^{j},\theta^{j}\right)\right]-\left[H\left(\theta^{j+1},\theta^{j}\right)-H\left(\theta^{j},\theta^{j}\right)\right]$$

$$=\sum_{i=1}^{m}\log\ p\left(x^{i};\theta^{j+1}\right)-\sum_{i=1}^{m}\log\ p\left(x^{i};\theta^{j}\right)$$

EM算法收敛证明

$$L\left(\theta^{j+1},\theta^{j}\right)-L\left(\theta^{j},\theta^{j}\right)\geq 0$$

$$H\left(\theta^{j+1},\theta^{j}\right)-H\left(\theta^{j},\theta^{j}\right)=\sum_{i=1}^{m}\sum_{z}p\left(z\middle|x^{i};\theta^{j}\right)\log\frac{p\left(z\middle|x^{i};\theta^{j+1}\right)}{p\left(z\middle|x^{i};\theta^{j}\right)}$$

$$\leq\sum_{i=1}^{m}\log\left(\sum_{z}p\left(z\middle|x^{i};\theta^{j}\right)\cdot\frac{p\left(z\middle|x^{i};\theta^{j+1}\right)}{p\left(z\middle|x^{i};\theta^{j}\right)}\right)=0$$

$$\sum_{i=1}^{m}\log\ p\left(x^{i};\theta^{j+1}\right)-\sum_{i=1}^{m}\log\ p\left(x^{i};\theta^{j}\right)\geq 0$$

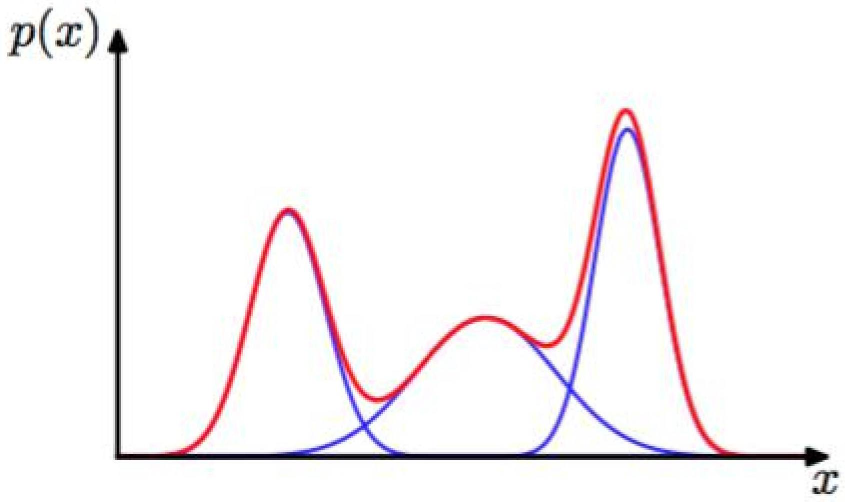
问题

- 随机选择1000名用户，测量用户的身高；若样本中存在男性和女性，身高分别服从高斯分布$N(\mu_1,\sigma_1)$和$N(\mu_2,\sigma_2)$的分布，试估计参数:$\mu_1,\sigma_1,\mu_2,\sigma_2$;
  - 如果明确的知道样本的情况(即男性和女性数据是分开的)，那么我们使用极大似然估计来估计这个参数值。
  - 如果样本是混合而成的，不能明确的区分开，那么就没法直接使用极大似然估计来进行参数的估计啦。

GMM

- GMM(Gaussian Mixture Model, 高斯混合模型)是指该算法由多个高斯模型线性叠加混合而成。每个高斯模型称之为component。GMM算法描述的是数据的本身存在的一种分布。
- GMM算法常用于聚类应用中，component的个数就可以认为是类别的数量。
- 假定GMM由k个Gaussian分布线性叠加而成，那么概率密度函数如下：

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



GMM

- 对数似然函数

$$l(\pi,\mu,\Sigma)=\sum_{i=1}^{N}\log\left(\sum_{k=1}^{K}\pi_k p\left(x^i;\mu_k,\Sigma_k\right)\right)$$

GMM-EM算法求解
E step

- w表示的是第i个样本属于第j个簇/类别的概率值

$$w_j^{(i)} = Q_i\left(z^{(i)} = j\right) = p\left(z^{(i)} = j \middle| x^{(i)}; \pi, \mu, \Sigma\right)$$

GMM-EM算法求解
M step

$$l(\pi,\mu,\Sigma)=\sum_{i=1}^{m}\sum_{z^{(i)}}Q_i\left(z^{(i)}\right)\log\left(\frac{p\left(x^{(i)},z^{(i)};\pi,\mu,\Sigma\right)}{Q_i\left(z^{(i)}\right)}\right)$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{k}Q_i\left(z^{(i)}=j\right)\log\frac{p\left(x^{(i)}\middle|z^{(i)}=j;\mu,\Sigma\right)\cdot p\left(z^{(i)}=j;\pi\right)}{Q_i\left(z^{(i)}=j\right)}$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{k}w_j^{(i)}\log\frac{\left(\frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_j\right|^{\frac{1}{2}}}e^{-\frac{1}{2}\left(x^{(i)}-\mu_j\right)^T\Sigma_j^{-1}\left(x^{(i)}-\mu_j\right)}\right)\cdot\pi_j}{w_j^{(i)}}$$

GMM-EM算法求解
对均值求偏导

$$l(\pi,\mu,\Sigma)=\sum_{i=1}^{m}\sum_{j=1}^{k}w_j^{(i)}\left(-\frac{1}{2}\left(x^{(i)}-\mu_j\right)^T\Sigma_j^{-1}\left(x^{(i)}-\mu_j\right)\right)+c$$

$$\frac{\partial l}{\partial \mu_l}=\frac{\partial}{\partial \mu_l}-\frac{1}{2}\sum_{i=1}^{m}w_l^{(i)}\left(x^{(i)T}\Sigma_l^{-1}x^{(i)}-x^{(i)T}\Sigma_l^{-1}\mu_l-\mu_l^T\Sigma_l^{-1}x^{(i)}+\mu_l^T\Sigma_l^{-1}\mu_l\right)$$

$$=\frac{1}{2}\sum_{i=1}^{m}w_l^{(i)}\left(\left(x^{(i)T}\Sigma_l^{-1}\right)^T+\Sigma_l^{-1}x^{(i)}-\left(\left(\Sigma_l^{-1}\right)^T+\Sigma_l^{-1}\right)\mu_l\right)$$

$$=\sum_{i=1}^{m}w_l^{(i)}\left(\Sigma_l^{-1}x^{(i)}-\Sigma_l^{-1}\mu_l\right)$$

GMM-EM算法求解
对均值求偏导

$$\frac{\partial l}{\partial \mu_l} = \sum_{i=1}^{m} w_l^{(i)} \left(\Sigma_l^{-1} x^{(i)} - \Sigma_l^{-1} \mu_l \right)$$

$$\xrightarrow{\text{令} \frac{\partial l}{\partial \mu_l} = 0} \mu_l = \frac{\sum_{i=1}^{m} w_l^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_l^{(i)}}$$

GMM-EM算法求解
对方差求偏导

$$l(\pi,\mu,\Sigma)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{k}w_j^{(i)}\left(\log\Sigma_j^{-1}-\left(x^{(i)}-\mu_j\right)^T\Sigma_j^{-1}\left(x^{(i)}-\mu_j\right)\right)+c$$

$$\frac{\partial l}{\partial\Sigma_l}=-\frac{1}{2}\sum_{i=1}^{m}w_l^{(i)}\left(\Sigma_l-\left(x^{(i)}-\mu_j\right)\left(x^{(i)}-\mu_j\right)^T\right)$$

$$\xrightarrow{\text{令}\frac{\partial l}{\partial\Sigma_l}=0}\Sigma_l=\frac{\sum_{i=1}^{m}w_l^{(i)}\left(x^{(i)}-\mu_l\right)\left(x^{(i)}-\mu_l\right)^T}{\sum_{i=1}^{m}w_l^{(i)}}$$

GMM-EM算法求解
对概率使用拉格朗日乘子法求解

$$l(\pi,\mu,\Sigma)=\sum_{i=1}^{m}\sum_{j=1}^{k}w_j^{(i)}\log\pi_j+c \quad s.t:\sum_{j=1}^{k}\pi_j=1$$

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

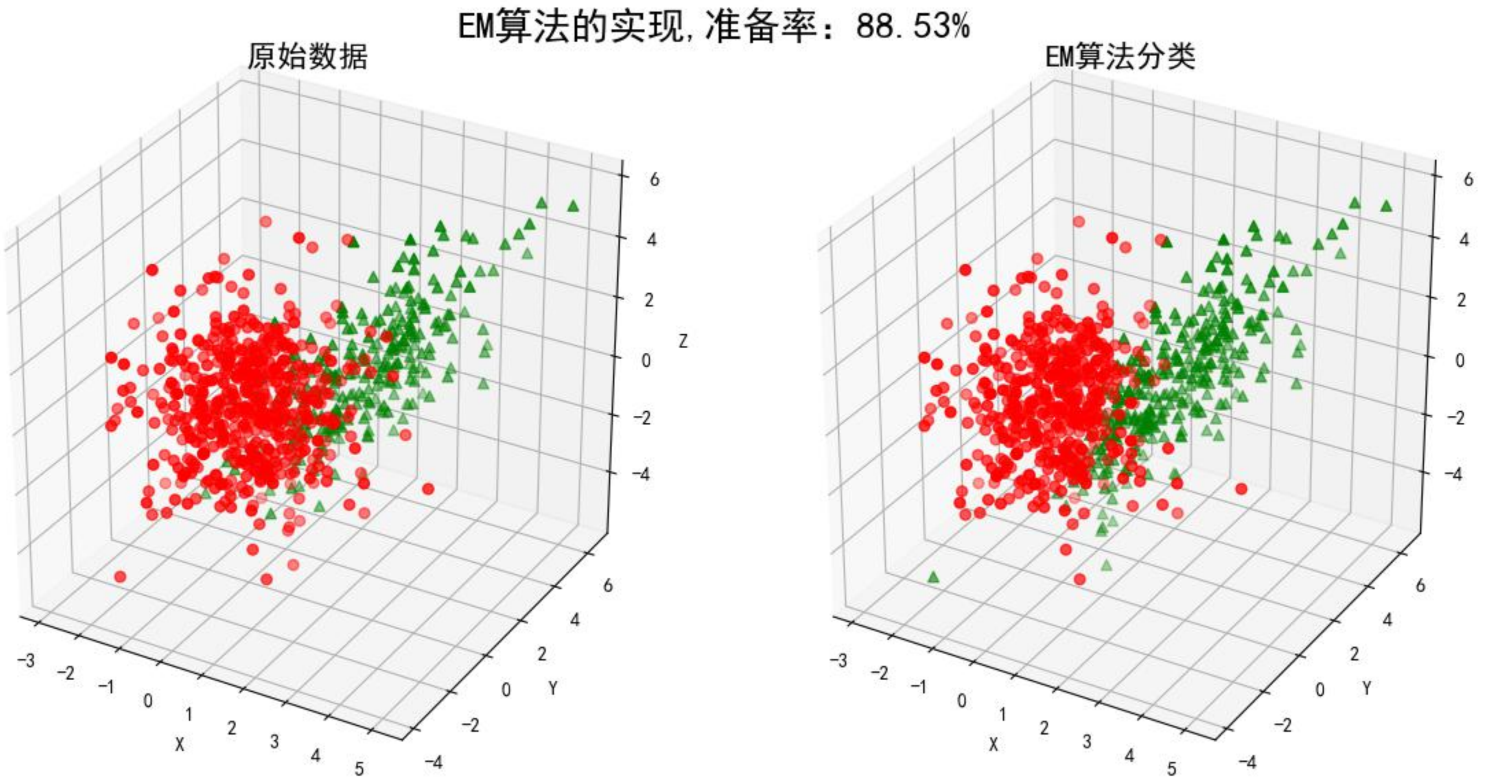
$$\frac{\partial L}{\partial\pi_l}=\sum_{i=1}^{m}\frac{w_l^{(i)}}{\pi_l}+\beta \xrightarrow{\text{令}\frac{\partial L}{\partial\pi_l}=0}$$

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

$$\pi_l=\frac{1}{m}\sum_{i=1}^{m}w_l^{(i)}$$

案例一：EM分类初识

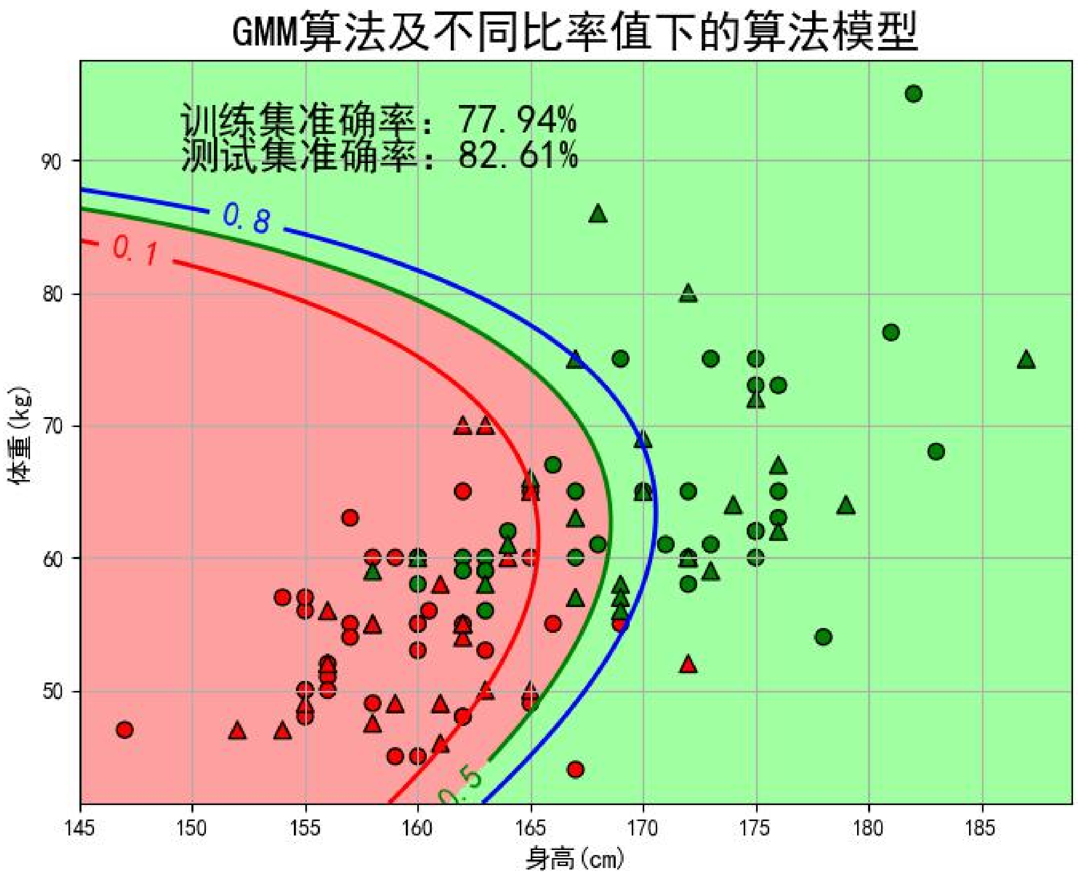
- 使用EM算法做分类问题



案例二：GMM算法分类及参数选择案例

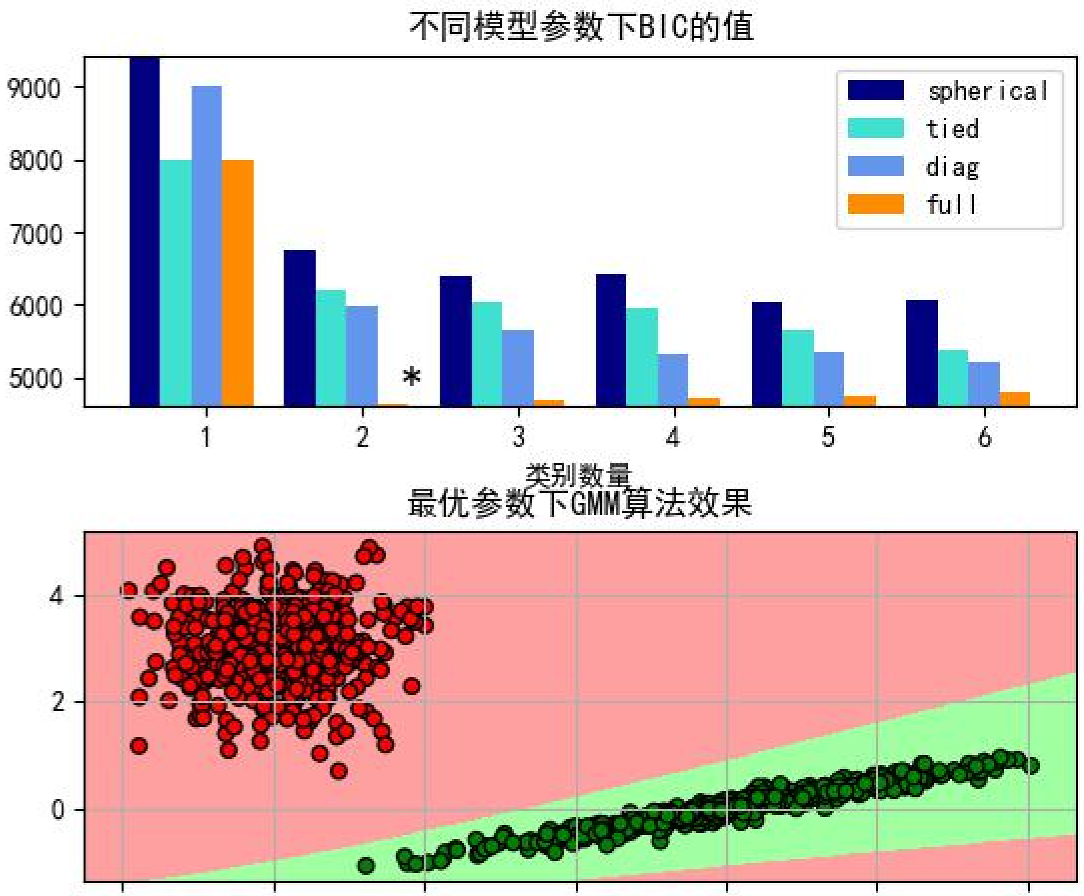
- GMM算法分类及参数选择

```
gmm = GaussianMixture(n_components=2, covariance_type='full', random_state=28)
gmm.fit(x, y)
```



案例三：GMM的不同参数

- GMM的不同参数结果如下图



案例四：EM无监督算法分类鸢尾花数据

- 使用EM算法对鸢尾花数据进行分类操作，数据来源：[鸢尾花数据]