# 无监督学习-EM算法

- 1.极大似然概率
- 2. EM算法推导
- 3. EM算法实例

- 1.似然估计

  - 现在有一个正反面不是很匀称的硬币，如果正面朝上记为H，反面朝上记为T，抛10次的结果如下:

  - T,T,T,H,T,T,H,T,T,T

  - 问正面朝上的概率是多少？

- 1.似然估计
  - 设反面朝上的概率为u, 则正面朝上的概率为1-u；
  - T,T,T,H,T,T,H,T,T,T
  - 那么出现上述可能性的概率是多大?
  - 为：u*u*u*(1-u)*u*u*(1-u)*u*u*u

- 1.似然估计

  - 为：u*u*u*(1-u)*u*u*(1-u)*u*u*u
  - 更一般的形式，我们假设正面朝上的x=1,反面朝上x=0
  - 一次的概率为: $u^x(1-u)^{(1-x)}$

$$p(\mathbf{X};\mu)=\prod_{i=1}^{n}p(x_i;\mu)=\prod_{i=1}^{n}\mu^{x_i}(1-\mu)^{1-x_i}$$

- 4.似然估计

$$
\begin{aligned}
\log p(\mathbf{X};\mu) &= \log \prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i} \\
&= \sum_{i=1}^{n} \log \left\{\mu^{x_i}(1-\mu)^{1-x_i}\right\} \\
&= \sum_{i=1}^{n} \left[\log \mu^{x_i} + \log(1-\mu)^{1-x_i}\right] \\
&= \sum_{i=1}^{n} \left[x_i \log \mu + (1-x_i)\log(1-\mu)\right]
\end{aligned}
$$

- 4.似然估计

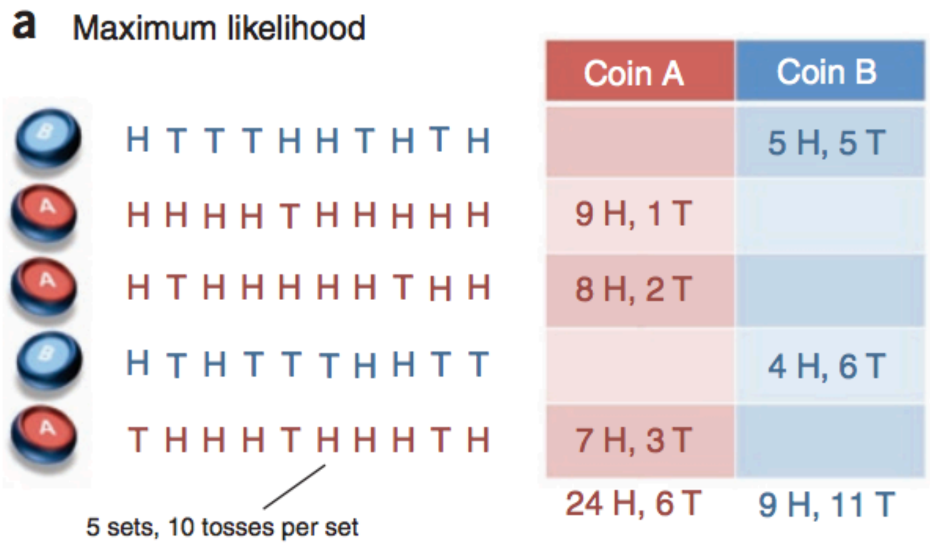
$$\begin{aligned}
\frac{\partial}{\partial\mu}\log p(\mathbf{X};\mu) &= \sum_{i=1}^{n}\frac{\partial}{\partial\mu}[x_i\log\mu+(1-x_i)\log(1-\mu)] \\
&= \sum_{i=1}^{n}x_i\frac{\partial}{\partial\mu}\log\mu+\sum_{i=1}^{n}(1-x_i)\frac{\partial}{\partial\mu}\log(1-\mu) \\
&= \frac{1}{\mu}\sum_{i=1}^{n}x_i-\frac{1}{1-\mu}\sum_{i=1}^{n}(1-x_i)
\end{aligned}$$

- 4.似然估计

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- 发现是正面朝上的概率是0.2， 我们实验了10次，有两次是正面。

• 如果现在有两个硬币A和B，要估计的参数是它们各自翻正面（head）的概率：




知道每次选的是A还是B，
利用上述的例子，可得估计值为：

$$\hat{\theta}_A = \frac{24}{24+6} = 0.80$$

$$\hat{\theta}_B = \frac{9}{9+11} = 0.45$$

• 如果不知道每次选的是A还是B，那如何估计？

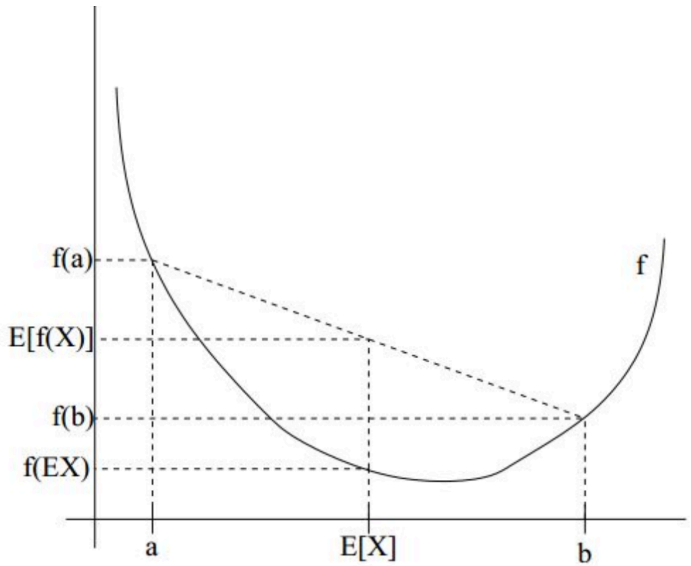

这就是我们EM要做的事情。

- EM算法推导

- 复习一下Jensen不等式：

  - 凸函数：

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



- 基本的Jensen不等式

若 $\theta_1, \ldots, \theta_k \geq 0, \theta_1 + \cdots + \theta_k = 1$

则 $f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$

如果是凹函数，符号相反。

- 我们的目标：在观察变量X 和给定观察样本x1; x2; :::; xn 的情况下，极大化对数似然函数：

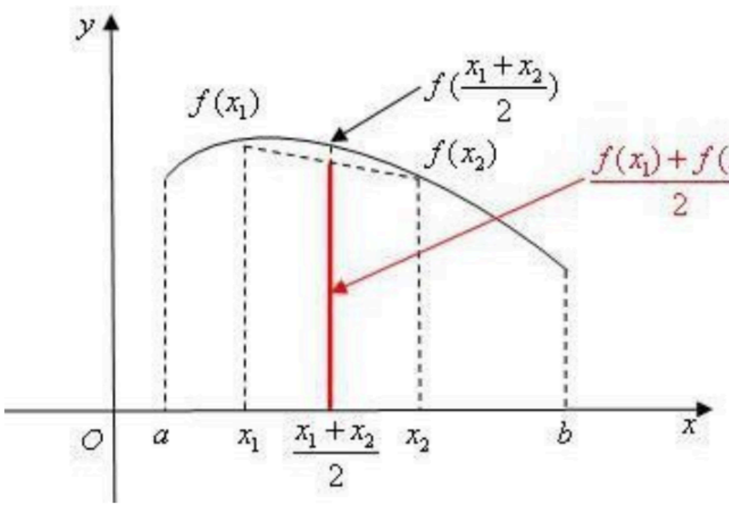
$$l(\theta)=\sum_{i=1}^{m}\log p(x;\theta)$$

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

Z为隐藏变量，不知道的参数，我不知道投的硬币是A还是B

令Qi是z的某一个分布，Qi≥0，有：

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




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

$$\geq\sum_{i=1}^{m}\sum_{z^{(i)}}Q_i(z^{(i)})\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}$$

- 等号成立的话：

$$\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}=c$$

一个概率分布的和为1

$$\sum_z Q_i\left(z^{(i)}\right)=1$$

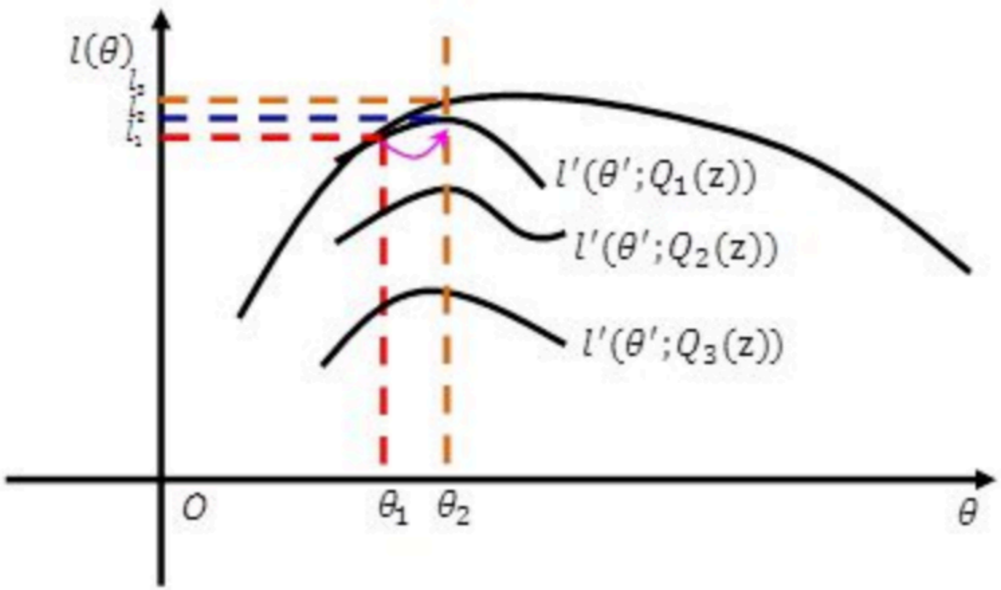
因此：

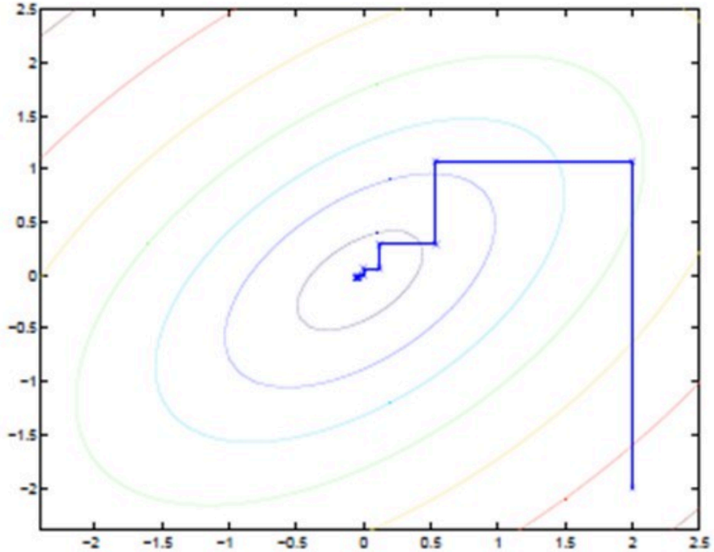
$$Q_i(z^{(i)})=\frac{p(x^{(i)},z^{(i)};\theta)}{\sum_z p(x^{(i)},z^{(i)};\theta)}$$

$$=\frac{p(x^{(i)},z^{(i)};\theta)}{p(x^{(i)};\theta)}$$

$$=p(z^{(i)}\mid x^{(i)};\theta)$$

固定一个 $\theta$ ，求出期望，E-step；
固定一个 z，最大化似然函数，M=step；
迭代知道满足收敛条件。



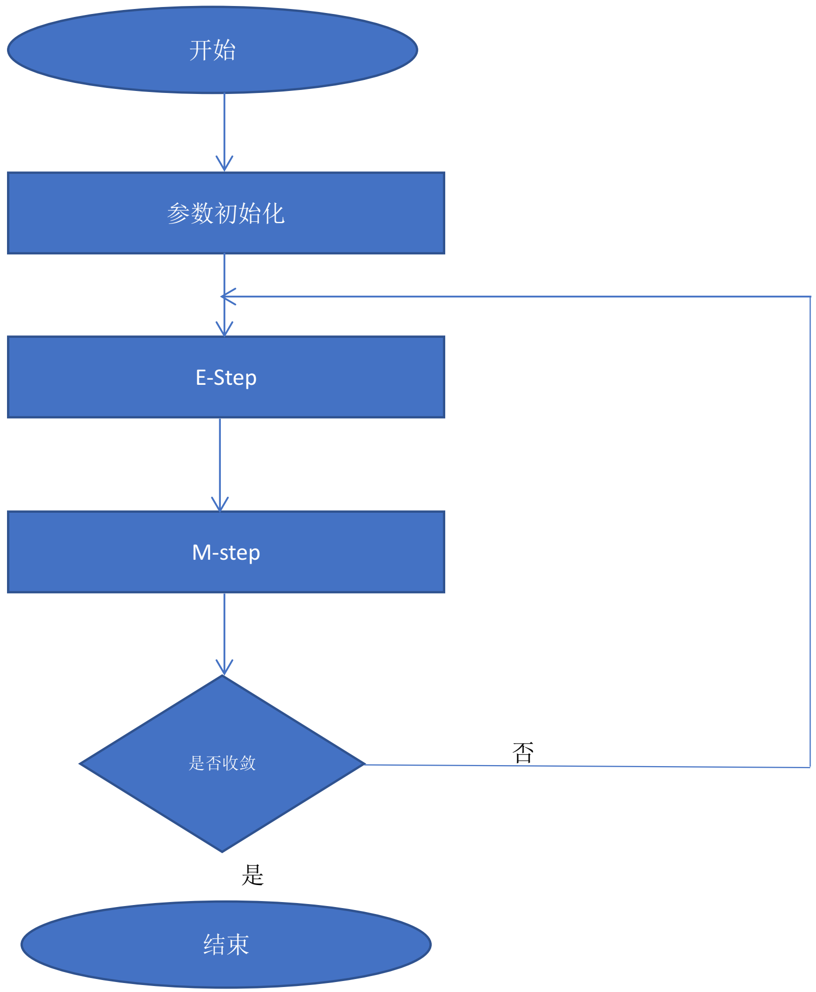



步骤：
1. 初始化参数

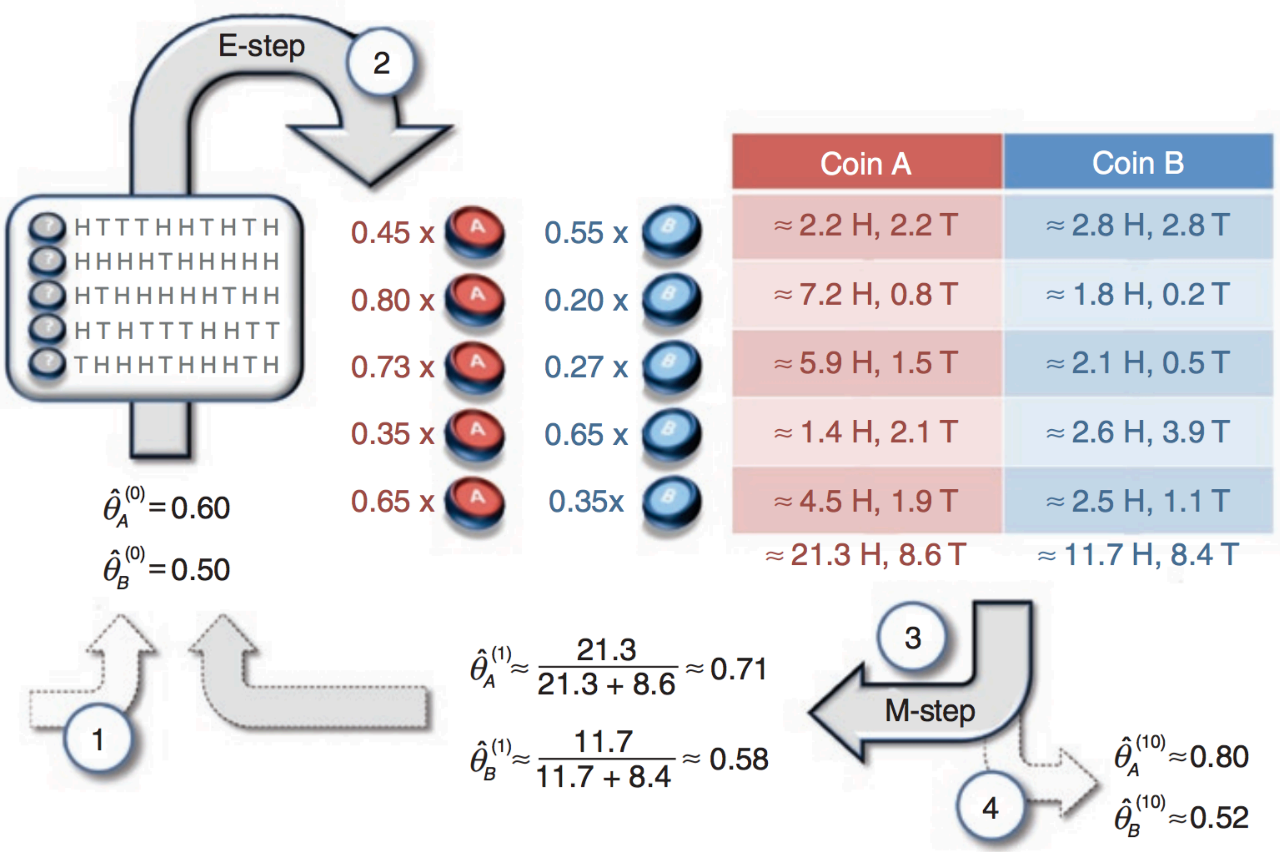
2. E-step

3. M-step

4. 结束迭代



上述例子的步骤：


E-step
2
H T T T H H T H T H
H H H H T H H H H H
H T H H H H H T H H
H T H T T T H H T T
T H H H T H H H T H
0.45 x
0.55 x
0.80 x
0.20 x
0.73 x
0.27 x
0.35 x
0.65 x
0.65 x
0.35x
Coin A
Coin B
≈ 2.2 H, 2.2 T
≈ 2.8 H, 2.8 T
≈ 7.2 H, 0.8 T
≈ 1.8 H, 0.2 T
≈ 5.9 H, 1.5 T
≈ 2.1 H, 0.5 T
≈ 1.4 H, 2.1 T
≈ 2.6 H, 3.9 T
≈ 4.5 H, 1.9 T
≈ 2.5 H, 1.1 T
≈ 21.3 H, 8.6 T
≈ 11.7 H, 8.4 T
$\hat{\theta}_A^{(0)} = 0.60$
$\hat{\theta}_B^{(0)} = 0.50$
1
$\hat{\theta}_A^{(1)} \approx \frac{21.3}{21.3 + 8.6} \approx 0.71$
$\hat{\theta}_B^{(1)} \approx \frac{11.7}{11.7 + 8.4} \approx 0.58$
3
M-step
4
$\hat{\theta}_A^{(10)} \approx 0.80$
$\hat{\theta}_B^{(10)} \approx 0.52$

- $Z1=C_{10}^{5}x(0.6)^5*(0.4)^5=0.20$
- $Z2=C_{10}^{5}x(0.5)^5*(0.5)^5=0.24$

- $Z1/Z2=0.45$

- 迭代计算，收敛得结果。

- Em算法收敛性：

  - 在EM框架下，求得的参数θ一定是收敛的，能够找到似然函数的最大值。

  - 只能保证收敛到稳定点，不能保证收敛到极大值点，因此EM算法受初值的影响较大。

- 如果样本的分布假设为高斯分布时，算法又叫做高斯混合模型（GMM）。
    - 高斯混合模型就是用高斯概率分布密度精确地量化事物，它是一个将事物分解为若干的基于高斯概率密度函数（正态分布曲线）形成的模型。

E步：计算概率

$$\gamma(i,k)=\frac{\pi_k N(x_i \mid \mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_i \mid \mu_j,\Sigma_j)}$$

M步：更新概率

$$N_k=\sum_{i=1}^{N}\gamma(i,k)$$

$$\mu_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)x_i$$

$$\Sigma_k=\frac{1}{N_k}\sum_{i=1}^{N}\gamma(i,k)(x_i-\mu_k)(x_i-\mu_k)^T$$

$$\pi_k=\frac{N_k}{N}=\frac{1}{N}\sum_{i=1}^{N}\gamma(i,k)$$

- K-means 回顾

1. 给定K的值，代表有K个不同的类别。
2. 对每一个类别，猜测其中心点。
3. 在已知K个中心点的情况下，计算每个点到这K的中心点的距离，距离最小的那个中心点所代表的类就是该点所属的类别，这样对所有样本完成分类。
4. 针对每一个类重新计算中心点，即将该类中所有点加和取平均，该均值则为新的中心点
5. 重复3~4的过程直到中心点收敛。

- 谢谢大家