EM算法

@author: natasha_yang

@e-mail: ityangna0402@163.com

极大似然估计:参数估计的方法之一
已知样本X满足某种概率分布,未知θ
怎么求:可微,求导数(多个,求偏导)

$$L(\theta)=L(x_0,x_1,x_2,\dots,x_n|\theta)\xrightarrow{样本独立}P(x_0|\theta)*P(x_1|\theta)*\dots P(x_n|\theta)=\prod_{i=0}^{n}P(x_i|\theta)\xrightarrow{对数求和}\sum_{i=0}^{n}\log P(x_i|\theta)$$

似然函数:样本X[x_i]的联合概率

y=g(x),g是连续函数
x离散,分布P(x=x_i)=p_i => E(y)=E(g(x))=求和[0-n]g(x_i)p_i
x连续,概率密度函数f(x)=>积分[+/-inf]g(x)f(x)dx

$$\hat{\theta}=\operatorname{argmax}(L(\theta))$$

样本X[x_i]已知,求θ使得样本X出现的P最大(找θ使得L(θ)最大)

问题:如果存在多个分布
不知道X[x_i]属于哪个分布的
不知道每个分布的θ

求解:1>x_i属于哪个分布
2>每个分布的θ

关系:1与2互相依赖

解法:先随便哪一个记F初始化值记FV,另外一个记B根据FV计算得出BV,然后F根据BV在计算出FV,循环往复,直到收敛

表示隐藏变量的分布

L(θ)

$$\sum_{i=0}^{n}\log P(x_i|\theta)\xrightarrow{展开z_j}\sum_{i=0}^{n}\log\big(P(x_i,z_0|\theta)+P(x_i,z_1|\theta)+\dots+P(x_i,z_m|\theta)\big)=\sum_{i=0}^{n}\log\sum_{j=0}^{m}P(x_i,z_j|\theta)$$

$$\sum_{i=0}^{n}\log\sum_{j=0}^{m}P(x_i,z_j|\theta)=\sum_{i=0}^{n}\log\sum_{j=0}^{m}\boxed{Q_i(z_j)}\frac{P(x_i,z_j|\theta)}{Q_i(z_j)}\xrightarrow{X=z_j;P_i(X)=Q_i(z_j);g(X)=\frac{P(x_i,z_j|\theta)}{Q_i(z_j)}}\sum_{i=0}^{n}\log\sum_{j=0}^{m}P_i(X)*g(X)$$

$$\sum_{i=0}^{n}\log\sum_{j=0}^{m}P_i(X)*g(X)\xrightarrow{E(g(X))=\sum P(X)*g(X)}\sum_{i=0}^{n}\log(E_i(g(X)))\xrightarrow{f(x)=\log(x)}\sum_{i=0}^{n}f(E_i(g(X)))$$

$$\sum_{i=0}^{n}f(E_i(g(X)))\xrightarrow{f:\log 凹函数}\geq\sum_{i=0}^{n}E_i(f(g(X)))\xrightarrow{E(f(x))=\sum P(x)*f(x)}\sum_{i=0}^{n}\sum_{j=0}^{m}P_i(X)*f(g(X))=\boxed{\sum_{i=0}^{n}\sum_{j=0}^{m}Q_i(z_j)*\log\frac{P(x_i,z_j|\theta)}{Q_i(z_j)}}$$

J(z,Q)

Jensen不等式(判断凹凸[二阶导数>=0])
凸函数:E(f(x))>=f(E(x))+凹函数:E(f(x))<=f(E(x))
等式何时成立:x是常数

问题:L(θ)>=J(z,Q)下界,通过不断提高下界求最大的L(θ)
解法:固定θ,调整Q=>J(z,Q)上升至L(θ);固定Q,调整θ=>J(z,Q)下界达到最大;建立下界,最大化下界,依次往复
求解:如何求θ使J(z,Q)下界等于L(θ)
为何一定会收敛

使J(z,Q)下界等于L(θ)
Jensen不等式等号成立,就是随机变量g(X)变成constant

下界=似然函数的条件

$$g(X)=\frac{P(x_i,z_j|\theta)}{Q_i(z_j)}\xrightarrow{\text{Jensen}不等式等号成立}g(X)=C\xrightarrow{分别对z_j求和}\frac{\sum_{j=1}^{m}P(x_i,z_j|\theta)}{\sum_{j=1}^{m}Q_i(z_j)}=C\xrightarrow{\sum_{j=1}^{m}Q_i(z_j)=1}\sum_{j=1}^{m}P(x_i,z_j|\theta)=C$$

$$P(x_i,z_j|\theta)\xrightarrow{A=z_j,B=x_i,P(AB)=P(A|B)*P(B)}P(z_j|x_i,\theta)*P(x_i|\theta)\xrightarrow{Q_i(z_j)=P(z_j|x_i,\theta)}Q_i(z_j)*P(x_i|\theta)$$

$$Q_i(z_j)=\frac{P(x_i,z_j|\theta)}{C}=\frac{P(x_i,z_j|\theta)}{\sum_{j=1}^{m}P(x_i,z_j|\theta)}=\frac{P(x_i,z_j|\theta)}{P(x_i|\theta)}\xrightarrow{P(AB)=P(A|B)*P(B)}P(z_j|x_i,\theta)$$ 计算Q

初始化θ

E step

计算联合分布的条件概率的期望

$$Q_i(z_j)=P(z_j|x_i,\bar{\theta})\xrightarrow{联合分布的条件概率期望}L(\theta,\bar{\theta})=\sum_{i}^{n}\sum_{j=1}^{m}Q_i(z_j)\log(x_i,z_j|\bar{\theta})$$

M step

找使得L(θ)最大的θ

$$\theta_{new}=\operatorname{argmax}(L(\theta,\bar{\theta}))$$

θ

验证上面步骤是否有效(结果是否收敛)

证明极大似然估计是单调递增的,那么最终会得到最大的极大似然估计
https://blog.csdn.net/pipisorry/article/details/42550815