EM

问题与答案

1. **EM的核心思想是什么？如何理解？**

是一种迭代算法，含有隐变量的概率模型参数的极大似然估计或极大后验概率估计。（样本产生和隐变量有关，极大似然求导无法求解，此时用EM）

E步：利用当前估计的参数值，求出在该参数下隐含变量的条件概率值（已知thita,推断隐变量Z）；

M步：结合E步求出的隐含变量条件概率，求出似然函数下界函数的最大值（已知Z，极大似然估计thita）

1. **用 EM 算法推导解释 Kmeans。**

k-means运行之前需归一化处理，不然可能会因为样本在某些维度上过大导致距离计算失效。

k-means中每个样本所属的类可以看做隐变量。在E步中，固定每个类的中心，对于每个点x_i找当前最近的聚类中心\mu_{y_i}；在M步，重新更新每个类的中心点，该步骤可以通过对目标函数求导实现，最终可得新的类中心就是类中样本的均值。

数学证明：

[**https://www.zhihu.com/question/49972233?sort=created**](https://www.zhihu.com/question/49972233?sort=created)

1. **采用 EM 算法求解的模型有哪些，为什么不用牛顿法或梯度下降法？**

一般有混合高斯、协同过滤、k-means。**牛顿法或梯度下降法**等优化方法的求和项数会随着隐变量的数目指数上升，会给梯度计算带来麻烦。

EM优缺点：

优点：简单性和普适性，是一种非梯度优化方法（解决梯度下降等优化方法的缺陷：见上）

缺点：对初值敏感,不同初值可能得到不同参数估计值；算法一定收敛，但可能局部最优。

[**http://blog.csdn.net/qq\_34896915/article/details/75040578**](http://blog.csdn.net/qq_34896915/article/details/75040578)

《机器学习-西瓜书》

<http://www.cnblogs.com/zuochongyan/p/5407053.html>

参数估计方法：

极大似然（参数确定值）、贝叶斯估计（参数是变量，有先验）、极大后验概率估计（参数固定值，具有先验）、EM(含有隐变量)、梯度下降