

第七章，聚类

1. 步骤，初始化质心，计算距离划分，更新质心。。。。
2. K均值聚类的特点
 1. 只适用于数值属性
 2. 对噪声数据是敏感的
 3. 算法的收敛速度和结果容易受初始质心的影响
3. EM算法的思想：用于含有隐变量（通常是多维的，一维的可以求）（Hidden variable）的概率模型参数的最大似然估计。（K均值的思想），E-step和M-step更新的参数分别是什么？
4. EM 算法具备收敛性，但并不保证找到全局最大值，有可能找到局部最大值。解决方法是初始化几次不同的参数进行迭代，取结果最好的那次。
5. 代价函数？
6. 先验概率：是指根据以往经验和分析得到的概率
7. 后验概率：事情已经发生，要求这件事情发生的原因是由某个因素引起的可能性的大小[1]，和 Bayes公式的作用差不多
8. 高斯混合模型（GMM）？理解和实现？