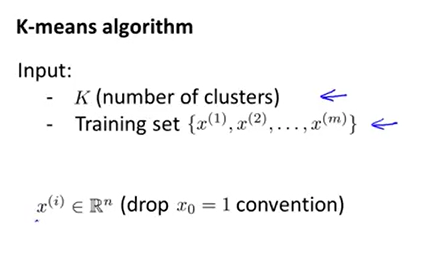
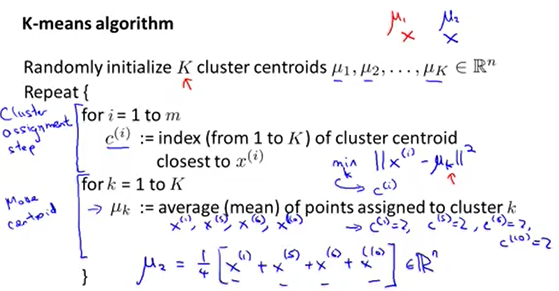
# 非监督学习

监督学习的训练数据有标记，非监督学习的数据无标记，非监督学习算法将未标记的训练数据送入特定的算法，分析出数据的结构。

## 1、K均值算法



K均值算法步骤：

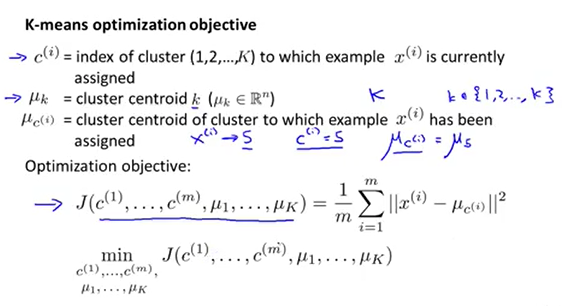


如果没有点分配给它的聚类中心，那么就直接移除该中心点，最终将会得到K-1个簇。

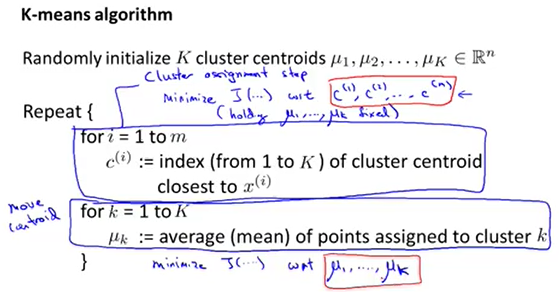
### 优化目标

所有算法都有一个优化目标函数，或者某个代价函数需要进行最小化。

**优化目标函数如下**：



K均值算法：

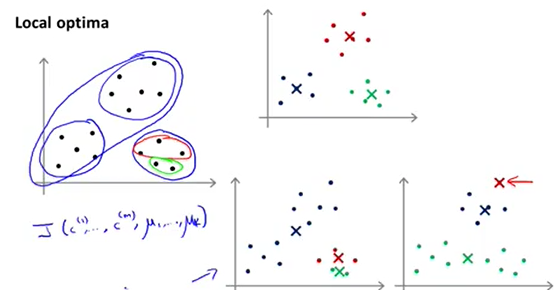


由上图可知，K均值算法的第一个步骤相当于固定均值变量，求损失函数相对于样本类的最小值，第二个步骤是固定类变量，求损失函数相对于均值的最小值。

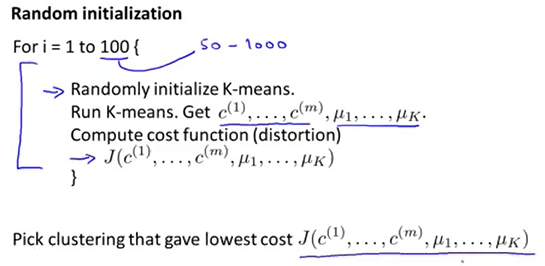
### 1.2 随机初始化

随机初始化是如何避开局部最优来构建K均值算法，K均值算法的第一步是随机初始化聚类中心，随机初始化K均值聚类中心一般是随机挑选K个训练样本，且该样本为聚类中心。

**局部最优化：**



解决局部最优的方法是尝试多次随机的初始化，来保证我们最终能得到一个足够好的结果。



运行100次的K均值初始随机化，对每一次随机化求得对应的损失函数，去损失函数最小值对应的K簇。若K值较小（2<=k<=10），运用多次的随机初始化可以得到很好的局部最优解，若K值过大，多次的随机初始化不可能会有太大的影响。