，K均值：

原理：

无监督，先随机选K个点作为中心点，然后对数据集中的每一个数据选个与中心点最近的相关联，聚成一类，计算每一个组的平均值，将中心点移到平均值的位置，然后反复迭代，直至中心点无变化。

损失函数：

$$J(c^{(1)}, c^{(2)}, \ldots, c^{(m)}, u_1, \ldots, u_k) = \frac{1}{m} \sum_{i=1}^{m} ||X^{(1)} - u_{c^{(i)}}||^2$$

训练的好坏取决于初始k的选择，训练也可能出现局部最小处，因此需要多次运行算法，然后选取代价最小的结果。

然后根据经验判断该选择几个K比较好。

高斯混合模型(GMM)：

核心思想是，数据都是通过多个高斯模型生成的，最后呈现的结果就是多个高斯模型堆叠而成的，我们要做的就是把它分离开。

公式：

$$p(x) = \sum_{i=1}^{k} \pi_i N(x|u_i, \sum_i)$$

通常我们并不能直接得到高斯混合模型的参数，而是观察到了一系列 数据点，给出一个类别的数量K后，希望求得最佳的K个高斯分模型。因此，高斯 混合模型的计算，便成了最佳的均值μ，方差Σ、权重π的寻找，这类问题通常通过 最大似然估计来求解。遗憾的是，此问题中直接使用最大似然估计，得到的是一 个复杂的非凸函数，目标函数是和的对数，难以展开和对其求偏导。

在这种情况下，可以用EM算法。 EM算法是在最大化目标函数时，先固定一个变量使整体函数变为凸优化函数，求导得到最值，然后利用最优参数更新被固定的变量，进入下一个循环。具体到高 斯混合模型的求解，EM算法的迭代过程如下。

首先，初始随机选择各参数的值。然后，重复下述两步，直到收敛。

- E步骤。根据当前的参数，计算每个点由某个分模型生成的概率。
- M步骤。使用E步骤估计出的概率，来改进每个分模型的均值，方差和权重。