**Thinking1 KNN与KMeans中的K分别代表什么？**

K近邻法（knn）是一种基本的分类与回归方法。k-means是一种简单而有效的聚类方法。

**knn**

算法思路： 如果一个样本在特征空间中的k个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别，则该样本也属于这个类别。

k近邻模型的三个基本要素：

1. k值的选择：k值的选择会对结果产生重大影响。较小的k值可以减少近似误差，但是会增加估计误差；较大的k值可以减小估计误差，但是会增加近似误差。一般而言，通常采用交叉验证法来选取最优的k值。
2. 距离度量：距离反映了特征空间中两个实例的相似程度。可以采用欧氏距离、曼哈顿距离等。
3. 分类决策规则：往往采用多数表决。

**k-means**

算法步骤： 1. 从n个数据中随机选择 k 个对象作为初始聚类中心； 2. 根据每个聚类对象的均值（中心对象），计算每个数据点与这些中心对象的距离；并根据最小距离准则，重新对数据进行划分； 3. 重新计算每个有变化的聚类簇的均值，选择与均值距离最小的数据作为中心对象； 4. 循环步骤2和3，直到每个聚类簇不再发生变化为止。

k-means方法的基本要素：

1. k值的选择：也就是类别的确定，与K近邻中k值的确定方法类似。
2. 距离度量：可以采用欧氏距离、曼哈顿距离等。

**Thinking2 都有哪些常用的启发式算法？**

模拟退火算法（SA）、遗传算法（GA）、蚁群算法（ACO）、人工神经网络（ANN）

**Thinking3 遗传算法的原理是怎样的？**

遗传算法：

* 通过模拟自然进化过程（达尔文生物进化论）搜索最优解的方法，遗传操作包括：选择、交叉和变异
* 算法核心：参数编码、初始群体的设定、适应度函数、遗传操作设计、控制参数设定
* 以一种群体中的所有个体为对象，利用随机化技术指导对一个被编码的参数空间进行高效搜索

遗传算法特点：

* 直接对结构对象进行操作，不存在求导和函数连续性的限定
* 具有内在的隐并行性和更好的全局寻优能力
* 采用概率化的寻优方法，不需要确定的规则就能自动获取和指导优化的搜索空间，自适应地调整搜索方向