

## 挖掘初始模型

### 1. 概括模型的输入输出特征

- 1) 本案例中输入组为 6 个政治关系数和 1 个亲戚为官数, 数据取值为数值型数据;
- 2) 输出组为官员类别, 数据取值为类别型数据

### 2. 根据变量和目标选择挖掘工具

- 1) 本次分析中输入组是 7 个数值型变量, 输出组是 1 个 “官员类型”, 属于类别型变量
- 2) 本案例的数据挖掘目标是官员的特征描述

综合而言, 我们选择可以按照内在相似性将数据划分为多个类别的**聚类算法**。

### 3. K-means 聚类分析

- 1) K-means 聚类分析: 是给定 K, 将数据构建成 K 个互不相交的群, 通过反复迭代改变分组的复位技术, 使得每一次改进之后的分群方案都比前一次好
- 2) 使用 K-means 算法对官员进行聚类分析, 具体操作请查看课程视频