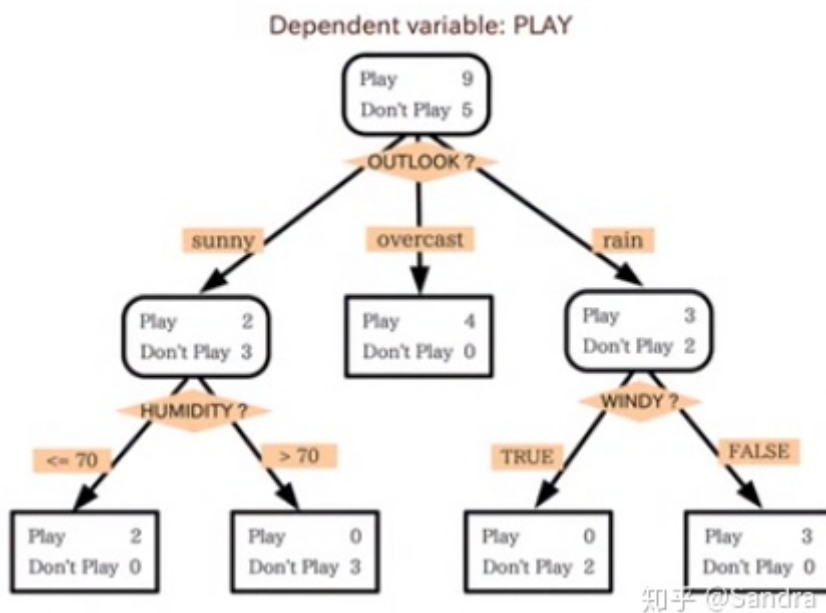


决策树



一层层做决策的模型

优点：有一个清晰易懂的结构，模仿了人类思维的方式

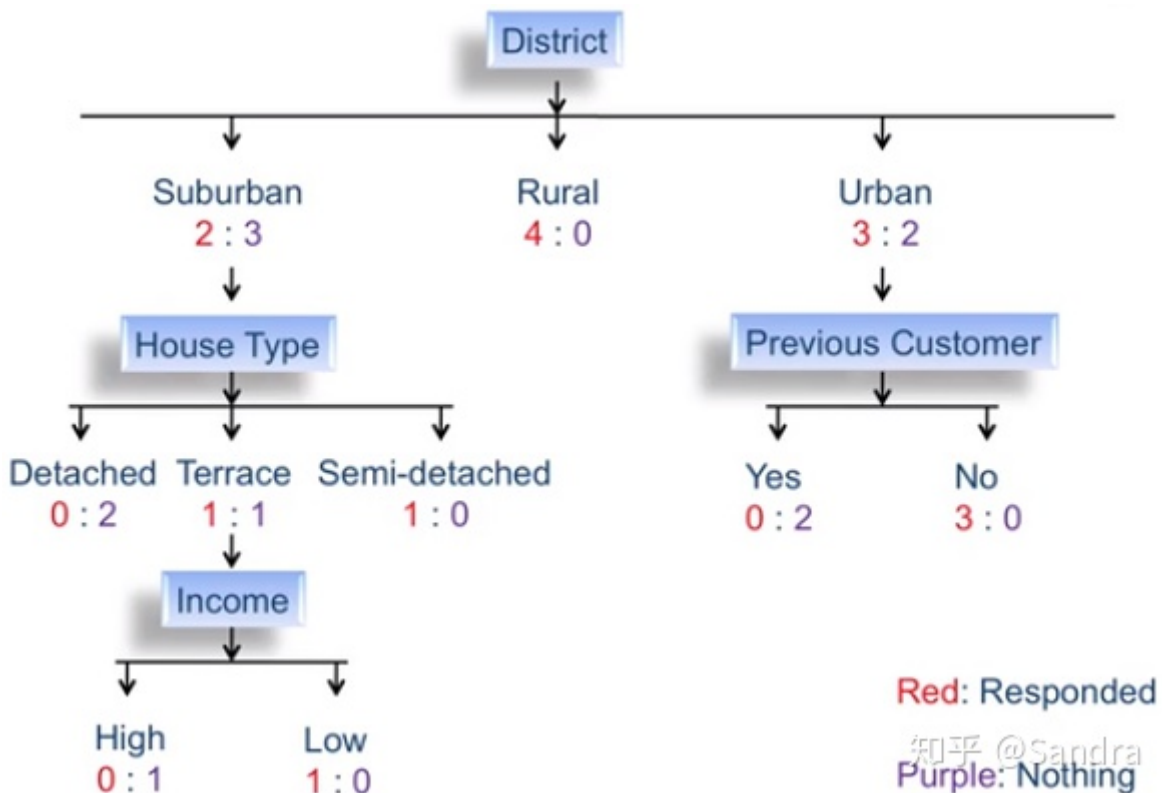
- ❖ Given the data collected from a promotion activity.
 - Could be tens of thousands of such records.
- ❖ Can we find any interesting patterns?
 - All rural households responded ...
- ❖ To find out which factors most strongly affect a household's response to a promotion.
 - Better understanding of potential customers
- ❖ Need a classifier to examine the underlying relationships and make future predictions.
- ❖ Send promotion brochures to selected households next time.
 - Targeted Marketing

根据接收、拒绝推荐的数据-->建模，查找潜在用户

| District | House Type | Income | Previous Customer | Outcome |
|----------|---------------|--------|-------------------|-----------|
| Suburban | Detached | High | No | Nothing |
| Suburban | Detached | High | Yes | Nothing |
| Rural | Detached | High | No | Responded |
| Urban | Semi-detached | High | No | Responded |
| Urban | Semi-detached | Low | No | Responded |
| Urban | Semi-detached | Low | Yes | Nothing |
| Rural | Semi-detached | Low | Yes | Responded |
| Suburban | Terrace | High | No | Nothing |
| Suburban | Semi-detached | Low | No | Responded |
| Urban | Terrace | Low | No | Responded |
| Suburban | Terrace | Low | Yes | Responded |
| Rural | Terrace | High | Yes | Responded |
| Rural | Detached | Low | No | Responded |
| Urban | Terrace | High | Yes | Nothing |

调查数据（最后一列是标签）

当无法做出决策时，继续采用其中的属性作为判断标准，继续划分，但极有可能有些分支分到最后仍是无法完全划分开的，树并不唯一



最理想的状态

- ❖ Rules can be easily extracted from the built tree.
 - (District = Rural) → (Outcome = Responded)
 - (District = Urban) AND (Previous Customer = Yes) → (Outcome = Nothing)
- ❖ One dataset, many possible trees
- ❖ Occam's Razor
 - The term *razor* refers to the act of shaving away unnecessary assumptions to get to the simplest explanation.
 - "When you have two competing theories that make exactly the same predictions, the simpler one is the better."
 - "The explanation of any phenomenon should make as few assumptions as possible, eliminating those making no difference in the observable predictions of the explanatory hypothesis or theory."
- ❖ Simpler trees are generally preferred.

知乎 @Sandra

剃刀模型：用尽可能简单的模型和方法进行分类