Advanced AI

学习课程

罗平

1-2

从信息熵到决策树

信息熵（Entropy）

- 一个简单的游戏：微信掷骰子

随机变量：Y，离散型随机变量

6个取值：6个可能的点数

问题是：评估“预测骰子点数的问题”的难度

Case 1：

$$P(y_1) = \cdots = P(y_6) = \frac{1}{6}$$

Case 2：

$$P(y_1) = \cdots = P(y_5) = 0, P(y_6) = 1$$

信息熵

- 评估预测“某支中超球队比赛胜负”问题的难度

随机变量：Y，离散型随机变量

3个取值：胜、平、负

CHINA FOOTBALL ASSOCIATION SUPER LEAGUE

中国足球协会超级联赛

$$P(y_1) = 0.3; P(y_2) = 0.267; P(y_3) = 0.433$$

$$P(y_1) = 0.633; P(y_2) = 0.333; P(y_3) = 0.033$$

形式化定义：信息熵

- 度量“预测随机变量Y的取值”的难度
- 已知：随机变量Y的分布P(Y)
  - 离散型随机变量
  - 连续型随机变量

$$H(Y):Y\in\mathcal{Y}\rightarrow\mathcal{R}$$

- 函数H需要满足的性质（考虑离散型随机变量的情况）

难度最高
H(Y)最大

$$P(y_1)=\cdots=P(y_n)=\frac{1}{n}$$

难度最低
H(Y)最小

$$P(y_i)=1,P(y_1)=\cdots=P(y_n)=0$$

物理意义与具体形式：信息熵

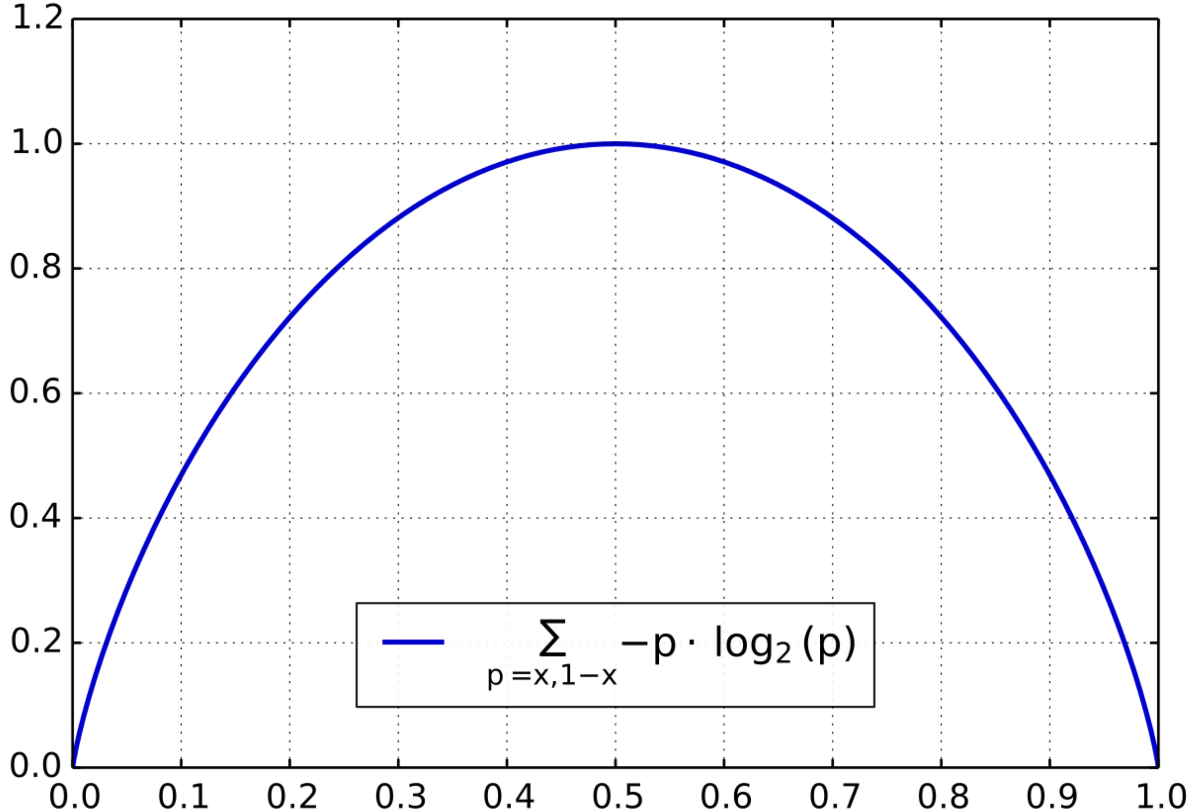
- 度量“预测随机变量Y的取值”的难度
- 度量“随机变量Y”的不确定性
- H(Y)越大，预测的难度越大，系统的不确定性越高

Entropy Name	Entropy
Shannon Entropy	$\mathcal{H}_{sha}(\pi) = \sum_{i=1}^{m} p_i log_2 \frac{1}{p_i}$
Pal Entropy	$\mathcal{H}_{pal}(\pi) = \sum_{i=1}^{m} p_i e^{1-p_i}$
Gini Index	$\mathcal{H}_{gin}(\pi) = \sum_{i=1}^{m} p_i (1 - p_i)$
Goodman-Kruskal coefficient	$\mathcal{H}_{goo}(\pi) = 1 - max_{i=1}^{m} p_i$

Ping Luo, et al.: On Defining Partition Entropy by Inequalities. IEEE Trans. Information Theory, 2007

香农信息熵

$$H(Y)=\sum_{i=1}^{n}P(y_i)\log\frac{1}{P(y_i)}$$



条件信息熵（Conditional Entropy）

- Y：需要预测的随机变量
- X：作为预测条件的随机变量
  - 考虑离散型随机变量的情况
  - 例如：X代表“主客场”，有“主场”和“客场”两种选者

- 度量：在X的条件下，预测Y的难度
  - 例如：在“主客场”已知的前提下，度量预测“胜负平”的难度

条件信息熵

- 在“主客场”已知的前提下，度量预测“胜负平”的难度
  - **x1**：主场 **x2**：客场
- H(Y|x1)

$$P(y_1|x1) = 0.4; P(y_2|x1) = 0.6; P(y_3|x1) = 0.0$$

- H(Y|x2)

$$P(y_1|x2) = 0.867; P(y_2|x2) = 0.067; P(y_3|x2) = 0.067$$

- H(Y|X)

$$\begin{aligned} H(Y|X) &= P(x_1)H(Y|x_1) + P(x_2)H(Y|x_2) \\ &= \frac{H(Y|x_1) + H(Y|x_2)}{2} \end{aligned}$$

形式化定义：条件信息熵

- 条件信息熵：度量“在随机变量X的前提下，预测随机变量Y”的难度
- 已知：P(Y|X)，P(X)

$$H(Y|X) : Y, X \in \mathcal{Y} \to \mathcal{R}$$

$$H(Y|X) = \sum_i P(x_i) H(Y|x_i)$$

- 信息增益（Information Gain）
  - 度量X对预测Y的能力

$$Gain(X, Y) = H(Y) - H(Y|X)$$

决策树：条件信息熵的应用

- 分类问题：预测恒大比赛的胜负
- 历史数据

主客场	时间	天气	郜飞机 是否首发	… …	胜负 情况
主场	星期六晚上	晴朗	是		胜
客场	星期六晚上	晴朗	是		胜
主场	星期天晚上	雨	否		平
客场	星期天晚上	晴朗	是		胜
客场	星期六晚上	阴天	否		胜
…	…	…	…	…	…
客场	星期六晚上	雪	是		负

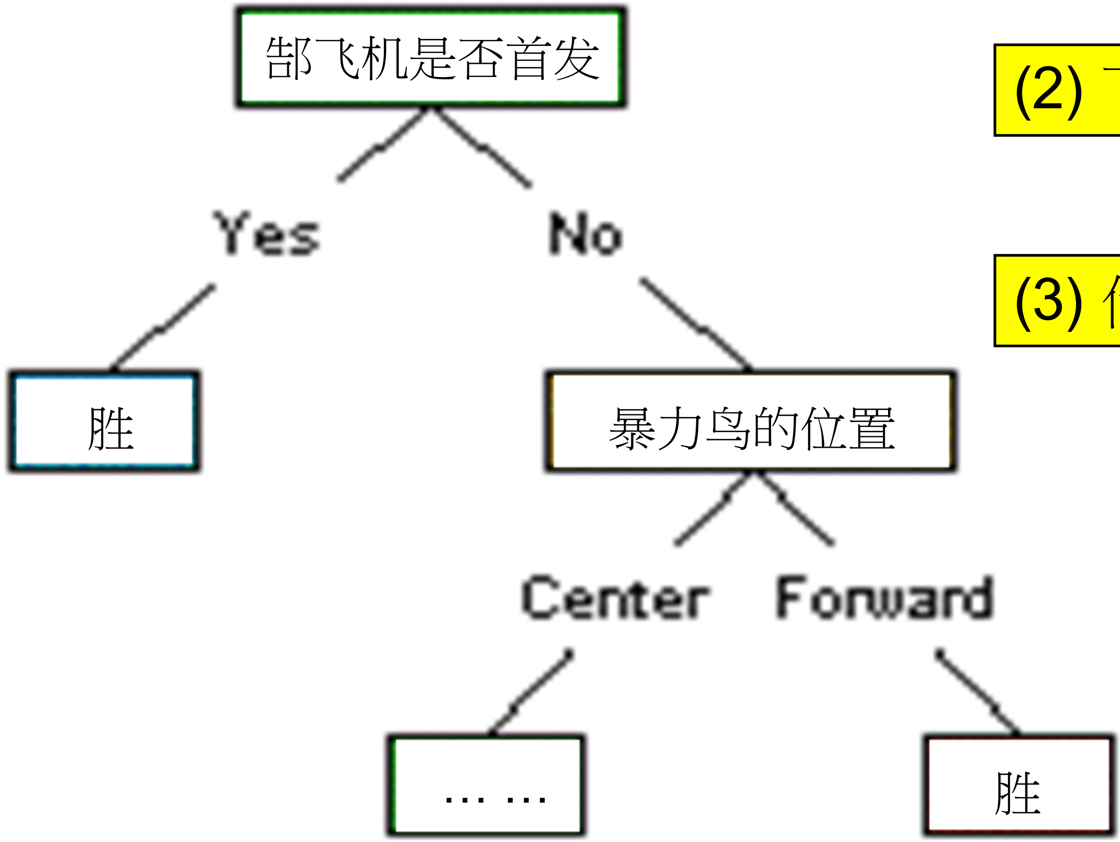
决策树：条件信息熵的应用

- 分类问题：基于历史数据，建立预测模型，对未知数据进行分类
- 预测下场球赛的结果

主客场	时间	天气	郜飞机 是否首发	…… ……	胜平负 情况
主场	星期六晚上	晴朗	否		？？？

Ross Quinlan, Inventor of Decision Tree

决策树：示例



(1) 根节点放置哪个条件属性

(2) 下面的节点放置什么属性

(3) 什么时候停止树的生长

决策树：构造

- 决策树构造准则
  - 分类准确
  - 树结构简单
- 选择根节点属性：贪心算法
  - 最大化信息增益，等价于最小化条件熵
  - 基于当前的数据计算
    - H(胜平负|主客场), H(胜平负|天气), H(胜平负|部飞机),…
    - 选择条件熵最小的条件属性，作为根节点

决策树：构造

- 计算：H(胜平负|部飞机是否首发)
- H(胜平负|部飞机首发)

主客场	时间	天气	部飞机 是否首发	……	胜负 情况
主场	星期六晚上	晴朗	是		胜
客场	星期六晚上	晴朗			胜
客场	星期天晚上	晴朗			胜
…	…	…		…	…
客场	星期六晚上	雪			负

决策树：构造

- 计算：H(胜平负|部飞机是否首发)
- H(胜平负|部飞机未首发)

主客场	时间	天气	部飞机 是否首发	……	胜负 情况
主场	星期天晚上	雨	否		平
客场	星期六晚上	阴天			胜

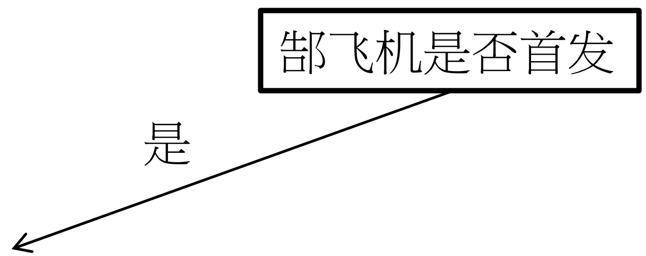
H(胜平负|部飞机是否首发)=

P(部飞机首发)H(胜平负|部飞机首发)+

P(部飞机未首发)H(胜平负|部飞机未首发)

决策树：构造

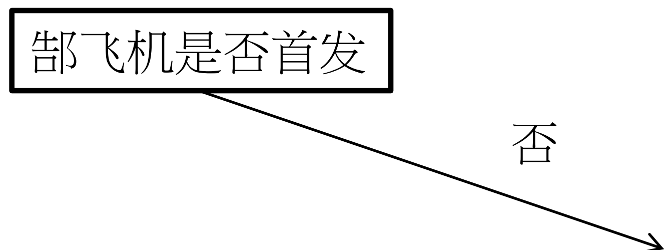
- 决策树的生长
  - 假设：根节点为“部飞机是否首发”



<table>
<tr><th>主客场</th><th>时间</th><th>天气</th><th>部飞机
是否首发</th><th>… …</th><th>胜负
情况</th></tr>
<tr><td>主场</td><td>星期六晚上</td><td>晴朗</td><td rowspan="5">是</td><td></td><td>胜</td></tr>
<tr><td>客场</td><td>星期六晚上</td><td>晴朗</td><td></td><td>胜</td></tr>
<tr><td>客场</td><td>星期天晚上</td><td>晴朗</td><td></td><td>胜</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>客场</td><td>星期六晚上</td><td>雪</td><td></td><td>负</td></tr>
</table>

决策树：构造

- 决策树的生长
  - 假设：根节点为“部飞机是否首发”



主客场	时间	天气	部飞机是否首发	……	胜负情况
主场	星期天晚上	雨	否		平
客场	星期六晚上	阴天			胜

- 结束生长
  - 当所对应的数据的熵足够的小

决策树：其它问题

- 连续型属性
- 决策树剪枝
- 决策树常用算法：C4.5

小结

- 概念：信息熵、条件熵（信息增益）
- 分类问题
- 决策树：最基本的分类模型