18 decision trees

问题：Chose which feature to split (according to importance)
↓ reduces uncertainty the most

Note：选择feature的顺序不影响结果
但会影响decision tree的大小

entropy：$H(x) = -\sum_{k} p(x=x_k)\log_2 p(x=x_k)$

(k代表x所有可能的取值)

into gain：$B(q) = -(q\log_2 q + (1-q)\log_2(1-q))$

实际计算：
- n — #negative samples
- p — #positive samples
- d — #outcomes of feature A

① $Rem(A) = \sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B\left(\frac{p_k}{p_k+n_k}\right)$

② $Gain(A) = B\left(\frac{p}{p+n}\right) - Rem(A)$

pruning：start with a full tree
from top to down：
对 a test is uninformative (↓ statistical significance testing), replace the node with a leaf

Compact representation of decision process

explanatory (可解释的)