**machine learning classification techniques:**

* decision tree based methods
* Naïve bayes & Bayesian belief networks
* ensemble methods
* artificial neural networks
* rule-based methods
* memory based reasoning
* support vector machines

**decision tree:**

* use existing data attributes & values
* can be used to classify new instances
* can be used to profile existing data
* robust to noise & missing values
* each tree can be viewed as “if-then-else” statements (high readability)
* can construct by hand

components:

* nodes:
  + leaf nodes: do not have children
  + non-leaf nodes: decision attributes
* branches: values of decision attributes

Iterative Dichotomiser 3 (ID3)

* each step determines the best decision attribute A for the next node
* assign A as decision attribute for node
* for each value of A create new descendant
* sort training examples to that node according to the attribute value of the branch
* if all training examples are homogenous, stop
* assume class attribute is categorical

**entropy:**

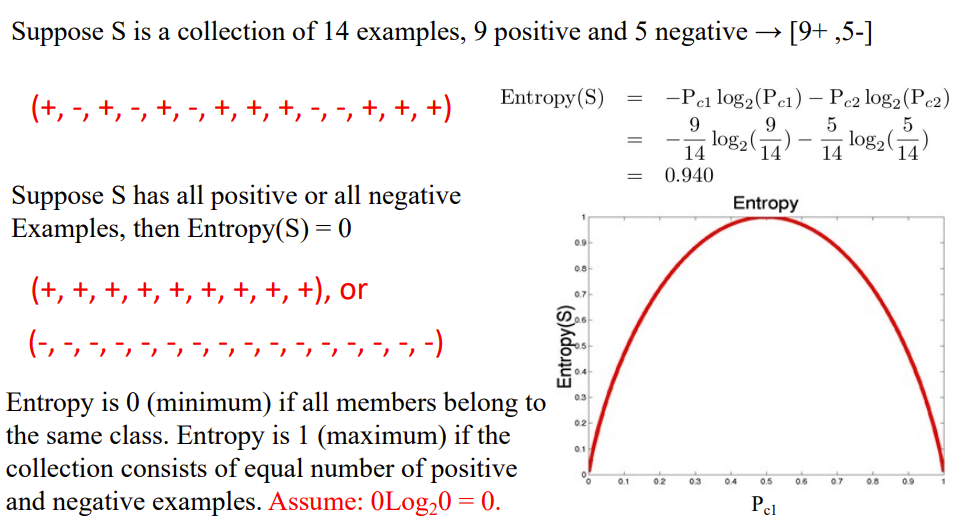
* measure the uncertainty in a random variable/message
* indicate how much information (impurity) there is in an event
* the more uncertain/random the event is, the more information it will contain



S: a collection of training examples

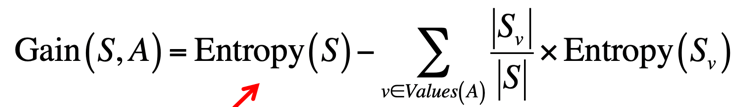
Pc1: the proportion of positive examples in S

Pc2: the proportion of negative examples in S



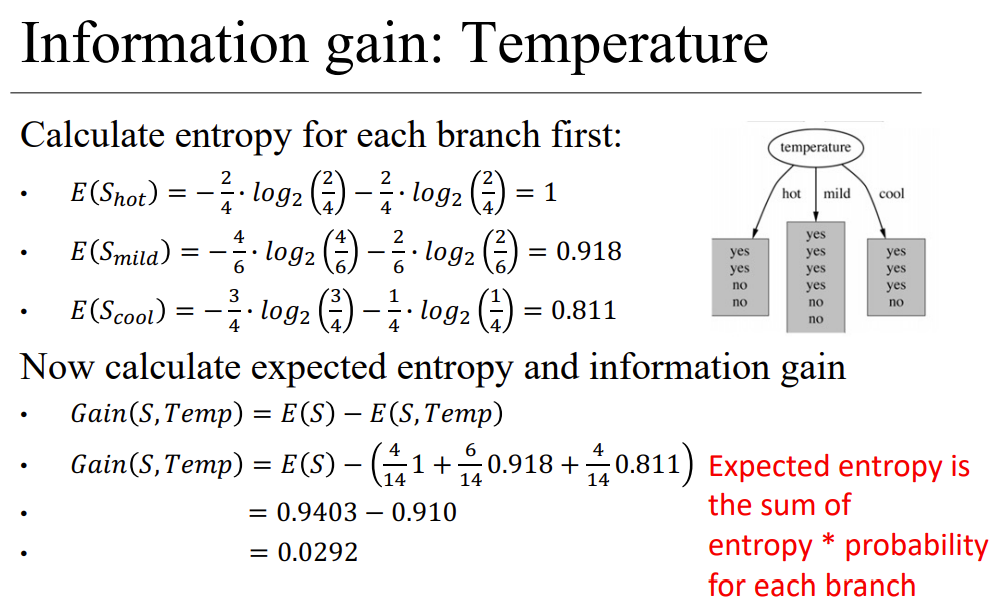
**Information gain:**

Expected reduction in entropy caused by partitioning the examples according to an attribute.



A: attribute A

Sv: subset of S for which attribute A has value v



**metrics for performance evaluation**

precision=TP/(TP+FP)

sensitivity=TP/(TP+FN)