树：是一种有层次性的结构，有深度优先、广度优先不同的算法，可以把很多的概念进行有规律的组织起来，便于分类。

❖ Classification is one of the fundamental skills for survival.
 ▪ Food vs. Predator

❖ A kind of **supervised** learning
 ▪ Techniques for deducing a function from data
 ▪ <Input, Output>
 ▪ Input: a vector of features
 ▪ Output: a Boolean value (binary classification) or integer (multiclass)

❖ "Supervised" means:
 ▪ A teacher or oracle is needed to label each data sample.

❖ We will talk about **unsupervised** learning later.

分类是一种有监督的学习（有标签，有输入有输出的一种组合），输出可以是布尔值，也可以是多类。
聚类是一种无监督的学习（无标签）

{boy, girl}

$Z = f(x, y)$

Height   Weight
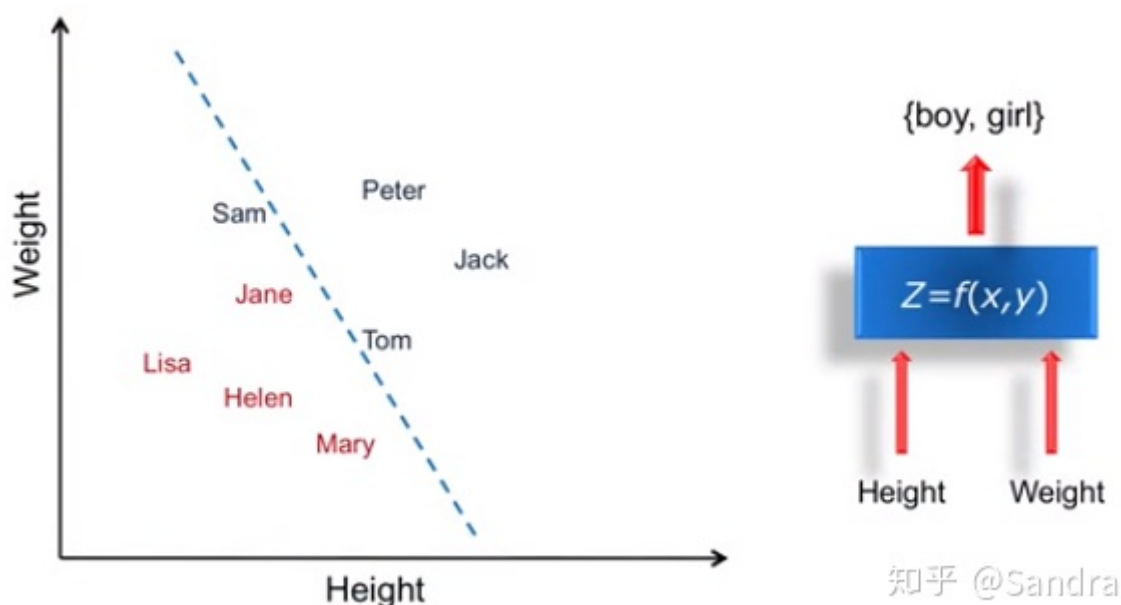
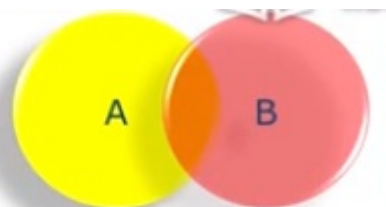存在分错的现象；中间可以视作一个黑盒，各种分类算法都是function，用不同的方式形成function

分类问题 = 函数，输入是它的属性（身高、体重、发色等），输出是分类的结果

学习的过程是有指导的，指导的就是数据的标签，形成一种更加抽象的表达模型，当再次遇到的时候可以进行判断。

贝叶斯公式：（描述的是后验概率）

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Likelihood of evidence B if A is true

Prior probability

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Posterior probability of A given the evidence B

Prior probability that evidence B is true

* Salmon vs. Tuna
* Grab a fish at random.
* $P(\omega_1) = P(\omega_2)$
* $P(\omega_1) > P(\omega_2)$
* Additional information

$$P(\omega_i \mid x) = \frac{P(x \mid \omega_i)P(\omega_i)}{P(x)}$$

我们知道不同于出现的概率可以有偏向的进行猜测；当我们知道不同种类的鱼的颜色以及已经捕捞上的鱼的颜色，就会进行更合理的猜测（贝叶斯）

- ❖ Probability of Kill
  - P(A): 0.6
  - P(B): 0.5

- ❖ The target is killed with:
  - One shoot from A
  - One shoot from B

- ❖ What is the probability that it is shot down by A?
  - C: The target is killed.

$$P(A \mid C) = \frac{P(C \mid A)P(A)}{P(C)} = \frac{1 \times 0.6}{0.6 \times 0.5 + 0.4 \times 0.5 + 0.} = \frac{3}{}$$

A命中，B命中不是互斥的事件，还是应该套用公式

- ❖ ω_1: Cancer; ω_2: Normal

- ❖ $P(\omega_1)=0.008$; $P(\omega_2)=0.992$

- ❖ Lab Test Outcomes: + vs. −

- ❖ $P(+|\omega_1)=0.98$; $P(-|\omega_1)=0.02$

- ❖ $P(+|\omega_2)=0.03$; $P(-|\omega_2)=0.97$

- ❖ Now someone has a positive test result...

- ❖ Is he/she doomed?

$$P(\omega_1 \mid +) \propto P(+ \mid \omega_1)P(\omega_1) = 0.98 \times 0.008 = 0.0078$$

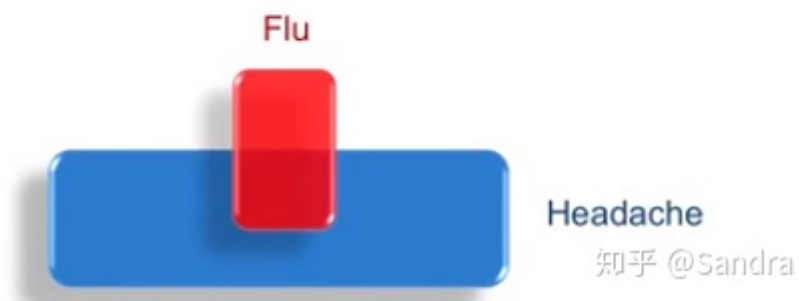$$P(\omega_2 \mid +) \propto P(+ \mid \omega_2)P(\omega_2) = 0.03 \times 0.992 = 0.0298$$

$$P(\omega_1 \mid +) < P(\omega_2 \mid +)$$

$$P(\omega_1 \mid +) = \frac{0.0078}{0.0078 + 0.0298} = 0.21 \gg P(\omega_1)$$

阳性结果，并且检测准确（不是假阳性）的概率没有过大，但也不可被忽略（因为还是远大于得癌症的先验概率的）

$$P(F \mid H) = \frac{P(H \mid F)P(F)}{P(H)} = \frac{1/2 \times 1/40}{1/10} = \frac{1}{8}$$

Flu

Headache

得流感会有50%头疼 不等于 头疼50%是得了流感