Artificial Intelligence ——Foundation of Mathematics

Yanghui Rao
Assistant Prof., Ph.D
School of Data and Computer Science,
Sun Yat-sen University
raoyangh@mail.sysu.edu.cn

概率

- 在计算机科学领域，概率模型首先出现在人工智能研究中（比如医疗诊断）
- 1972年de Bombal等人的系统对严重腹痛的正确诊断率平均超过90%， 远远高于当时专家级别的医生的正确诊断率平均值

Computer-aided Diagnosis of Acute Abdominal Pain

F. T. de DOMBAL, D. J. LEAPER, J. R. STANILAND, A. P. McCANN, JANE C. HORROCKS

British Medical Journal, 1972, 2, 9-13

Summary

This paper reports a controlled prospective unselected real-time comparison of human and computer-aided diagnosis in a series of 304 patients suffering from abdominal pain of acute onset.

The computing system's overall diagnostic accuracy (91·8%) was significantly higher than that of the most senior member of the clinical team to see each case (79·6%). It is suggested as a result of these studies that the provision of such a system to aid the clinician is both feasible in a real-time clinical setting, and likely to be of practical value, albeit in a small percentage of cases.

Introduction

We have already described our general operational experience

概率

- Frequentist (频率派)
  - 事件的概率是当我们无限次重复试验时，事件发生次数的比值。
  - 掷骰子、投掷硬币、纸牌游戏等。

概率

- Frequentist (频率派)
  - 事件的概率是当我们无限次重复试验时，事件发生次数的比值。
  - 掷骰子、投掷硬币、纸牌游戏等。
- 概率视为一种主观置信度
  - 明天下雨的概率是50%
  - 你愿意押1赔3（赢+1元，输-3元），在你的观念中，明天下雨的概率是多少?

概率

- $P(A,B)=P(A)P(B)$?
  - A: 第一枚硬币正面朝上; B: 第二枚硬币正面朝上
  - A: 第一天下雨; B: 第二天下雨

概率

- **乘法法则:**

$$P(A,B)=P(A)P(B|A)=P(B,A)=P(B)P(A|B)$$

$$P(A,B_1,B_2,B_3)=P(A)P(B_1|A)P(B_2|A,B_1)P(B_3|A,B_1,B_2)$$

$$P(\mathrm{Grade}=\mathrm{A} \mid \mathrm{Student}=\mathrm{Smart})=0.6$$

$$P(\mathrm{Grade}=\mathrm{A})=0.2$$

$$P(\mathrm{Student}=\mathrm{Smart})=0.3$$

$$P(\mathrm{Student}=\mathrm{Smart} \mid \mathrm{Grade}=\mathrm{A})=?$$

概率

- **乘法法则:**

$P(A,B)=P(A)P(B|A)=P(B,A)=P(B)P(A|B)$

$P(A,B_1,B_2,B_3)=P(A)P(B_1|A)P(B_2|A,B_1)P(B_3|A,B_1,B_2)$

$P(\text{Grade} = \text{A} \mid \text{Student} = \text{Smart}) = 0.6$

$P(\text{Grade} = \text{A}) = 0.2$

$P(\text{Student} = \text{Smart}) = 0.3$

$P(\text{Student} = \text{Smart} \mid \text{Grade} = \text{A}) = 0.9$

If $P(\text{Grade} = \text{A}) = 0.4$, then

$P(\text{Student} = \text{Smart} \mid \text{Grade} = \text{A}) = ?$

概率

- **乘法法则:**
  $$P(A,B)=P(A)P(B|A)=P(B)P(A|B)$$
  $$P(A,B_1,B_2,B_3)=P(A)P(B_1|A)P(B_2|A,B_1)P(B_3|A,B_1,B_2)$$

  P(两只大眼睛,四条腿,白肚皮,绿衣服)

  鸭妈妈说： 两只大眼睛 -> 大金鱼

  大金鱼说： 四条腿 -> 大乌龟

  大乌龟说： 白肚皮 -> 大白鹅

  大白鹅说： 绿衣服 -> 青蛙

概率

- **乘法法则:**

$$P(A,B)=P(A)P(B|A)=P(B)P(A|B)$$

$$P(A,B_1,B_2,B_3)=P(A)P(B_1|A)P(B_2|A,B_1)P(B_3|A,B_1,B_2)$$

P(两只大眼睛,四条腿,白肚皮,绿衣服)

鸭妈妈说：两只大眼睛 -> 大金鱼

大金鱼说：白肚皮 -> 大白鹅

大白鹅说：四条腿 -> 大乌龟

大乌龟说：绿衣服 -> 青蛙

概率

- **乘法法则:**
$P(A,B)=P(A)P(B|A)=P(B)P(A|B)$
$P(A,B_1,B_2,B_3)=P(A)P(B_1|A)P(B_2|A,B_1)P(B_3|A,B_1,B_2)$

- **加法法则:** $P(A)=P(A,B)+P(A,B^c)$

$$P(A)=\sum_B P(A,B)=\sum_{i=1}^{n} P(A,B_i)$$

$$=\sum_{i=1}^{n} P(A|B_i)P(B_i)$$

概率

- What's the value of $\sum_G P(G \mid X = \text{boy})$
  - 1
  - P(X=boy)
  - None of the above

概率

- What's the value of $\sum_G P(G \mid X = \text{boy})$
  - 1
  - P(X=boy)
  - None of the above

概率

- Exercise: 假设有一盒骰子，里面有4面的（点数为1、2、3、4），6面的、8面的、12面的、20面的均匀骰子各1个。如果我随机从盒子中选一个骰子，投掷它得到了点数5。那么我选中的骰子为4面、6面、8面、12面、20面的概率各是多少?

概率

- Exercise：假设有一盒骰子，里面有4面的（点数为1、2、3、4），6面的、8面的、12面的、20面的均匀骰子各1个。如果我随机从盒子中选一个骰子，投掷它得到了点数5。那么我选中的骰子为4面、6面、8面、12面、20面的概率各是多少?

答案: 0, 0.392, 0.294, 0.196, 0.118

概率

- Exercise: Suppose there are k types of fruits, and that each new one collected is, independent of previous ones, a type j fruit with probability p_j, $\sum_{j=1}^{k} p_j = 1$ Find the probability that the n-th fruit collected is a different type than any of the preceding n-1.

概率

- Exercise: Suppose there are k types of fruits, and that each new one collected is, independent of previous ones, a type j fruit with probability p_j, $\sum_{j=1}^{k} p_j = 1$ Find the probability that the n-th fruit collected is a different type than any of the preceding n-1.

Solution:

$$P(N) = \sum_{j=1}^{k} P(N \mid T_j) P(T_j)$$

$$= \sum_{j=1}^{k} (1 - p_j)^{n-1} p_j$$

变量类型

- 离散型变量
  - A discrete (离散) variable has a finite or countably infinite set of values.
  - Such variables can be categorical, such as gender, or numeric, such as counts.
  - Discrete variables are often represented using integer values.
  - Binary (二元) variables are a special case of discrete variables and assume only two values, e.g. true/false, yes/no, or 0/1.

变量类型

- 连续型变量
  - A continuous (连续) variable is one whose values are real numbers.
  - Examples include temperature, height or weight.
  - Continuous attributes are represented as floating point variables typically.

Expectation (期望)

- If X is a discrete random variable

$$E[X] = \sum_i x_i P\{X = x_i\}$$

- If X is a continuous random variable having probability density function f

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

期望

- If rolling one die (6-sided) and X is the value on its face, then: $E[X]$?

期望

- If rolling one die (6-sided) and X is the value on its face, then: $E[X]$?

$$E[X] = \sum_{x=1}^{6} xp(x) = \frac{1}{6}\sum_{x=1}^{6} x = \frac{21}{6}$$

Median (中位数)

- Sort n variables
  - $X(1) <= X(2) <= \ldots <= X(n)$
- If n is odd number
  - $X((n+1)/2)$
- If n is even number
  - $(X(n/2)+X(1+n/2))/2$

Mode (众数)

- 10 5 9 12
- 6 5 9 8 5
- 25 28 28 36 25 42

Variance (方差)

- $\mathrm{Var}(X) = E[(X\text{-}E[X])^2] = E[X^2]\text{-}(E[X])^2$

X	$E(X)$	$(X\text{-}E(X))^2$	X^2
1	2	1	1
2	2	0	4
3	2	1	9

Covariance (协方差)

- Cov(X,Y)=$E[(X-E(X))(Y-E(Y))]$

$$= E[XY - E(X)Y - XE(Y) + E(X)E(Y)]$$

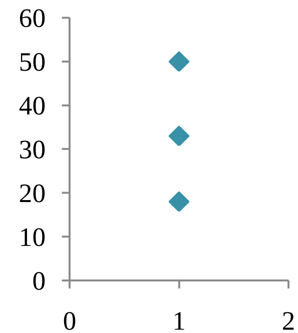
$$= E[XY] - E(X)E[Y] - E[X]E(Y) + E(X)E(Y)$$

$$= E[XY] - E[X]E[Y]$$

Correlation (相关系数)

- If X and Y are independent random variables, then Cov(X,Y)=0

性别	年龄
1	18
1	50
1	33

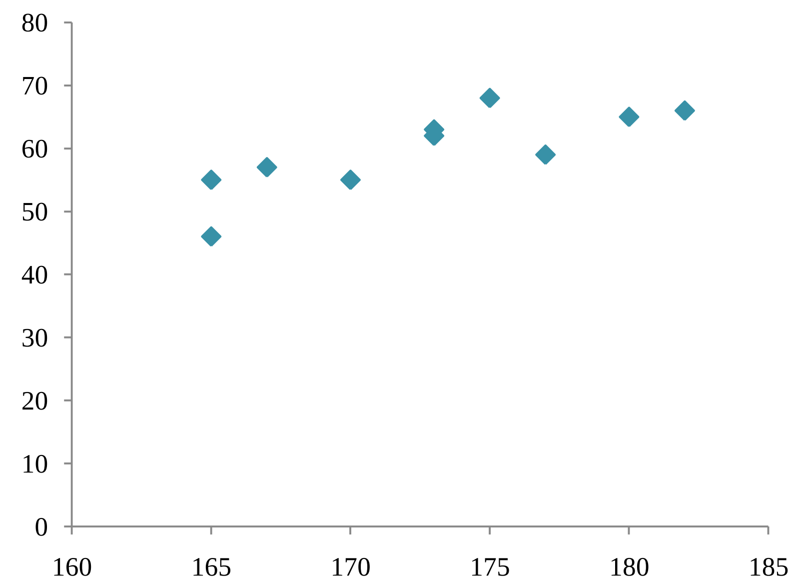


- The *correlation* between two random variables X and Y is:

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Correlation (相关系数)

身高(cm)	体重(kg)
165	46
177	59
170	55
180	65
173	63
165	55
167	57
182	66
173	62
175	68



10位同学身高与体重的相关系数：0.80

均匀分布

- *Uniformly distributed* (均匀分布) *random variables*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$$

$$E(x) = \frac{1}{b-a}\int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

$$E(x^2) = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + b^2 + ab}{3}$$

$$\mathrm{Var}(x) = \frac{1}{12}(b-a)^2$$

正态分布

- *Normal* (正态/高斯) *random variables*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$E[X] = \mu$$

$$\mathrm{Var}(X) = \sigma^2$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$$

The distribution function of a standard normal random variable

距离

- The Euclidean distance d between two vectors $\mathbf{x}$ and $\mathbf{y}$ is given by

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

where

  - n is the number of dimensions
  - x_k and y_k are the k-th item of $\mathbf{x}$ and $\mathbf{y}$

距离

- The Euclidean distance measure is generalized by the *Minkowski* distance metric as follows:

$$d(\mathbf{x},\mathbf{y}) = \left(\sum_{k=1}^{n} | x_k - y_k |^r \right)^{\frac{1}{r}}$$

- Three common examples of *Minkowski* distances:
  - r=1: City block distance (L_1 norm)
  - r=2: Euclidean distance (L_2 norm)
  - r=∞: Supremum distance (L_{max} or L_∞ norm), which is the maximum difference between any item of the vectors.

距离

- Suppose x and y coordinates of four vectors:
  - p1 = <0, 2>
  - p2 = <2, 0>
  - p3 = <3, 1>
  - p4 = <5, 1>

距离

L_1	p1	p2	p3	p4
p1	0.0	4.0	4.0	6.0
p2	4.0	0.0	2.0	4.0
p3	4.0	2.0	0.0	2.0
p4	6.0	4.0	2.0	0.0

L_2	p1	p2	p3	p4
p1	0.0	2.8	3.2	5.1
p2	2.8	0.0	1.4	3.2
p3	3.2	1.4	0.0	2.0
p4	5.1	3.2	2.0	0.0

L_{max}	p1	p2	p3	p4
p1	0.0	2.0	3.0	5.0
p2	2.0	0.0	1.0	3.0
p3	3.0	1.0	0.0	2.0
p4	5.0	3.0	2.0	0.0

距离

新闻标题	公众“感动”的概率
少年 救出 溺水 男童	0. 9
老人 参加 高考	0. 5
男童 救出 溺水 老人	?

少年	救出	溺水	男童	老人	参加	高考	公众“感动”的概率
0.25	0.25	0.25	0.25	0	0	0	0.9
0	0	0	0	0.33	0.33	0.33	0.5
0	0.25	0.25	0.25	0.25	0	0	?