信息、控制与计算

Fundamental of Information, Control and Computation

刘雷 研究员
浙江大学信息与电子工程学院
Email: lei_liu@zju.edu.cn

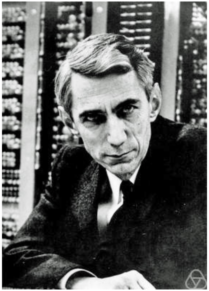
信息的度量—熵

1. 熵
2. 熵的性质
3. 联合熵、条件熵

信息的度量—熵

1. 熵
2. 熵的性质
3. 联合熵、条件熵

信息论之父



克劳德 · 香农（Claude Shannon）

“信息论之父”

在我看来，二三百年后，当人们回顾我们这个时代时，不会记得谁是美国总统，不会记得谁是电影明星或摇滚明星，但克劳德 · 香农的名字在仍然会被人们熟知，学校仍然会教授信息论。

—理查德 · 布拉胡特，2000

信息哲学

- 信息就是信息，而不是物质或能量。

—诺伯特 · 维纳，1948

- 自然必须被解释为物质、能量和信息。

—杰里米 ·C· 坎贝尔，1984

- 宇宙的基础可能不是能量或物质，而是信息。

—菲利普 · 佩里，2017

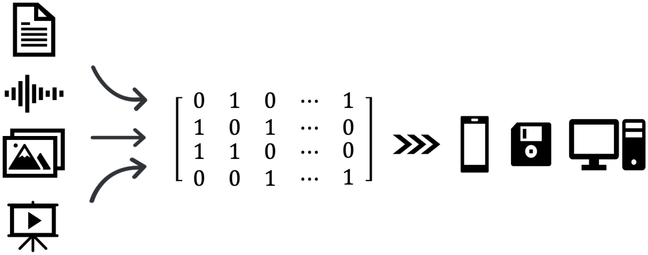
- 物质和能量构成了宇宙的“身体”，信息是“灵魂”。

$$\textbf{宇宙} = \textbf{物质} + \textbf{能量}$$
$$\textbf{自然} = \textbf{物质} + \textbf{能量} + \textbf{信息}$$

- 信息是通过物质和能量来产生、传递、存储、加工和感知的物质和能量是通过信息来表达和控制的。

Q1：什么是信息？

认识信息



对信息的认识：

- 信息**载体**：文本、语音、图片、视频等
- 信息储存**形式**：比特（0、1）序列
- 信息**度量**：比特数

还是有些抽象

认识信息

投硬币

2 种可能

需 1 比特

投骰子

6 种可能

需 3 比特

信息：对不确定的可能性数数！

Q2：信息的严格数学定义？

信息量和随机事件

信息量：描述随机事件的最少平均比特数

信息量和随机事件

信息量：描述随机事件的最少平均比特数

- 事件发生的可能性越大，信息就越少
  $\Rightarrow$ 明天天气晴朗/多雨/多云

信息量和随机事件

信息量：描述随机事件的最少平均比特数

- 事件发生的可能性越大，信息就越少
  $\Rightarrow$ 明天天气晴朗/多雨/多云
- 事件发生的可能性越小，信息就越多 (惊喜)
  $\Rightarrow$ 明天下雪/地震

信息量和随机事件

信息量：描述随机事件的最少平均比特数

- 事件发生的可能性越大，信息就越少
  $\Rightarrow$ 明天天气晴朗/多雨/多云
- 事件发生的可能性越小，信息就越多 (惊喜)
  $\Rightarrow$ 明天下雪/地震
- 确定事件没有任何信息
  $\Rightarrow$ 昨天天气

信息量和随机事件

信息量：描述随机事件的最少平均比特数

- 事件发生的可能性越大，信息就越少
  $\Rightarrow$ 明天天气晴朗/多雨/多云
- 事件发生的可能性越小，信息就越多 (惊喜)
  $\Rightarrow$ 明天下雪/地震
- 确定事件没有任何信息
  $\Rightarrow$ 昨天天气

信息量 $\Leftrightarrow$ 不确定度

随机变量

- X：随机变量
- x 或 x_i：变量取值（观测值）
- $\mathcal{X}$：随机变量 X 的样本空间（所有取值的集合）
- $|\mathcal{X}|$：集合 $\mathcal{X}$ 的元素个数，也称集合 $\mathcal{X}$ 的势
- p_x：概率分布函数

$$p_x = p_X(x) = \Pr[X = x]$$

p_x 满足

$$0 \le p_x \le 1, \quad \sum_{x \in \mathcal{X}} p_x = 1.$$

信息量是概率 p_x 的函数

事件“$X=x$”不确定度 $\Rightarrow$ 事件“$X=x$”发生概率 p_x

- ι_x：表示事件“$X=x$”发生所含有的信息量

信息量是概率 p_x 的函数

事件 "$X = x$" 不确定度 $\Rightarrow$ 事件 "$X = x$" 发生概率 p_x

- ι_x：表示事件 "$X = x$" 发生所含有的信息量
- 性质：
  - 单调性：ι_x 关于 p_x 单调递减

信息量是概率 p_x 的函数

事件“$X=x$”不确定度 $\Rightarrow$ **事件“$X=x$”发生概率** p_x

- ι_x：表示事件“$X=x$”发生所含有的信息量
- 性质：
  - 单调性：ι_x 关于 p_x 单调递减
  - 确定性：当 $p_x=1$ 时，$\iota_x=0$

信息量是概率 p_x 的函数

事件 “$X = x$” 不确定度 $\Rightarrow$ 事件 “$X = x$” 发生概率 p_x

- ι_x：表示事件 “$X = x$” 发生所含有的信息量
- 性质：
  - 单调性：ι_x 关于 p_x 单调递减
  - 确定性：当 $p_x = 1$ 时，$\iota_x = 0$
  - 无穷性：当 $p_x \to 0$ 时，$\iota_x \to \infty$

信息量是 p_x 的对数函数

考虑独立随机变量 X_1 和 X_2

- 事件“$X_1 = x_1$”发生概率$p_{x_1} \Rightarrow$ 信息量ι_{x_1}
- 事件“$X_2 = x_2$”发生概率为$p_{x_2} \Rightarrow$ 信息量ι_{x_2}
- “$X_1 = x_1$ & $X_2 = x_2$”发生概率$p_{x_1,x_2} = p_{x_1} p_{x_2} \Rightarrow$ 信息量ι_{x_1,x_2}

问：ι_{x_1,x_2} 是什么形式？

- $\iota_{x_1,x_2} = \iota_{x_1} + \iota_{x_2}$，即

$$\iota(p_{x_1} p_{x_2}) = \iota(p_{x_1}) + \iota(p_{x_2})$$

$$\Rightarrow \iota(p) = \log(f(p))$$

信息量的定义

ι_x 性质:

- 单调性: ι_x 关于 p_x 单调递减

信息量的定义

ι_x 性质:

- 单调性：ι_x 关于 p_x 单调递减
- 确定性：当 $p_x = 1$ 时，$\iota_x = 0$

信息量的定义

ι_x 性质:

- 单调性：ι_x 关于 p_x 单调递减
- 确定性：当 $p_x = 1$ 时，$\iota_x = 0$
- 无穷性：当 $p_x \to 0$ 时，$\iota_x \to \infty$

信息量的定义

ι_x 性质:

- 单调性: ι_x 关于 p_x 单调递减
- 确定性: 当 $p_x = 1$ 时, $\iota_x = 0$
- 无穷性: 当 $p_x \to 0$ 时, $\iota_x \to \infty$
- 对数性: $\iota_x = \log(f(p_x))$

信息量的定义

ι_x 性质:

- 单调性: ι_x 关于 p_x 单调递减
- 确定性: 当 $p_x = 1$ 时, $\iota_x = 0$
- 无穷性: 当 $p_x \to 0$ 时, $\iota_x \to \infty$
- 对数性: $\iota_x = \log(f(p_x))$

什么函数满足上述性质?

信息量的定义

ι_x 性质:

- 单调性：ι_x 关于 p_x 单调递减
- 确定性：当 $p_x = 1$ 时，$\iota_x = 0$
- 无穷性：当 $p_x \to 0$ 时，$\iota_x \to \infty$
- 对数性：$\iota_x = \log(f(p_x))$

什么函数满足上述性质？

- $\iota_x = \frac{1}{p_x}$?
  - ✓ 单调性
  - ✓ 无穷性
  - × 确定性
  - × 对数性

信息量的定义

ι_x 性质:

- 单调性：ι_x 关于 p_x 单调递减
- 确定性：当 $p_x = 1$ 时，$\iota_x = 0$
- 无穷性：当 $p_x \to 0$ 时，$\iota_x \to \infty$
- 对数性：$\iota_x = \log(f(p_x))$

什么函数满足上述性质？

- $\iota_x = \frac{1}{p_x}$?
  - ✓ 单调性
  - ✓ 无穷性
  - × 确定性
  - × 对数性
- $\iota_x = \log p_x$?
  - × 单调性
  - × 无穷性
  - ✓ 确定性
  - ✓ 对数性

信息量的定义

ι_x 性质:

- 单调性: ι_x 关于 p_x 单调递减
- 确定性: 当 $p_x = 1$ 时, $\iota_x = 0$
- 无穷性: 当 $p_x \to 0$ 时, $\iota_x \to \infty$
- 对数性: $\iota_x = \log(f(p_x))$

什么函数满足上述性质?

- $\iota_x = \frac{1}{p_x}$?
  - ✓ 单调性
  - ✓ 无穷性
  - × 确定性
  - × 对数性
- $\iota_x = \log p_x$?
  - × 单调性
  - × 无穷性
  - ✓ 确定性
  - ✓ 对数性
- $\iota_x = -\log p_x$?
  - ✓ 单调性
  - ✓ 无穷性
  - ✓ 确定性
  - ✓ 对数性

信息量的定义

ι_x 性质:

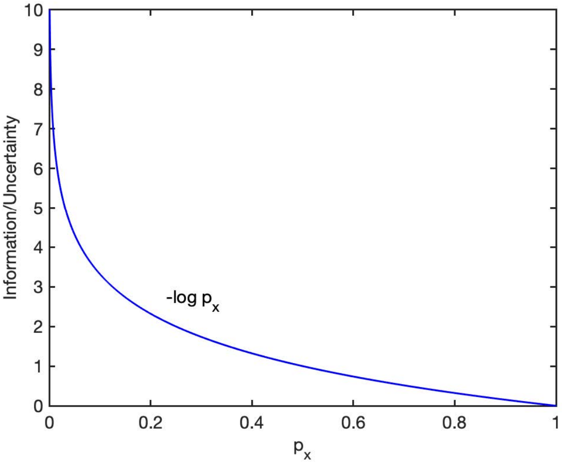
- 单调性: ι_x 关于 p_x 单调递减
- 确定性: 当 $p_x = 1$ 时, $\iota_x = 0$
- 无穷性: 当 $p_x \to 0$ 时, $\iota_x \to \infty$
- 对数性: $\iota_x = \log(f(p_x))$

什么函数满足上述性质?

- $\iota_x = \frac{1}{p_x}$?
  - ✓ 单调性
  - ✓ 无穷性
  - × 确定性
  - × 对数性
- $\iota_x = \log p_x$?
  - × 单调性
  - × 无穷性
  - ✓ 确定性
  - ✓ 对数性
- $\iota_x = -\log p_x$?
  - ✓ 单调性
  - ✓ 无穷性
  - ✓ 确定性
  - ✓ 对数性

$\Rightarrow$ 事件 $X = x$ 的信息量: "$\iota_x = -\log p_x$"

事件 $X = x$ 的信息量



10
9
8
7
6
5
4
3
2
1
0
Information/Uncertainty
-log p_x
0
0.2
0.4
0.6
0.8
1
p_x

举例

令 Y 为描述某天天气的随机变量。

- 样本空间：$\mathcal{Y}=${晴，多云，雨，雪}
- 样本总数：$|\mathcal{Y}|=4$.
- 概率分布

$$p_y = \begin{cases} 5/16 & \text{if } y = \text{晴}, \\ 5/16 & \text{if } y = \text{多云}, \\ 5/16 & \text{if } y = \text{雨}, \\ 1/16 & \text{if } y = \text{雪}, \end{cases}$$

满足

$$0 \le p_y \le 1 \text{ 且 } \sum_{y\in\mathcal{Y}} p_y = 1.$$

举例

令 Y 为描述某天天气的随机变量。

- 样本空间：$\mathcal{Y}=${晴，多云，雨，雪}
- 样本总数：$|\mathcal{Y}|=4.$
- 概率分布

$$p_y = \begin{cases} 5/16 & \text{if } y = \text{晴}, \\ 5/16 & \text{if } y = \text{多云}, \\ 5/16 & \text{if } y = \text{雨}, \\ 1/16 & \text{if } y = \text{雪}, \end{cases}$$

满足

$$0 \leq p_y \leq 1 \text{ 且 } \sum_{y \in \mathcal{Y}} p_y = 1.$$

- 晴/云/雨：大概率 ($\frac{15}{16}$) $\Rightarrow$ 信息量小
- 雪 (Snowy)：小概率 ($\frac{1}{16}$)，惊喜！$\Rightarrow$ 信息量大

“违反直觉”

经验： **“不要杞人忧天！”**

- “$p_x = 0$” $\Rightarrow$ “$X = x$” 不会发生 $\Rightarrow$ 信息量应该为 0 而非 ∞？
- “概率越小信息量越大”是否正确？

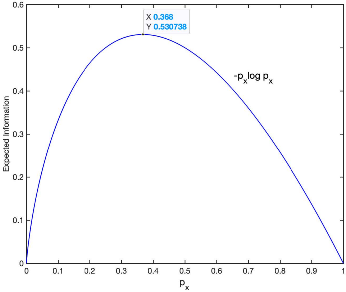
似乎与信息的定义相悖，为什么？

- 上述信息定义的提前： “事件 $X = x$ 发生”
- 事件 $X = x$ 的期望信息量应乘上事件发生概率 p_x，即，

$$\mathcal{I}_x = p_x \iota_x = -p_x \log p_x$$

应关注期望信息量更大的事件

事件 $X = x$ 的期望信息量



- 不可能事件：$\lim_{p_x \to 0} \mathcal{I}_x \to 0$.
- $\mathcal{I}_x$ 是 p_x 的凸函数
- $\mathcal{I}_x$ 的极大值点为 $p_x = 1/e$（解 $\mathrm{d}\mathcal{I}_x/\mathrm{d}p_x = 0$）

“关注概率 $p_x \approx 1/e$” 的事件”

Q3：信息度量？$\Rightarrow$ 信息熵

熵

定义：熵为随机变量 X 的不确定度，即期望信息量：

$$H(X) = -\sum_{x\in\mathcal{X}} p_x \log p_x = \mathrm{E}_X\{-\log p_X\}$$

- $-\log p_x$：事件 $X = x$ 的不确定性（信息量）
- $\mathrm{E}_X\{\cdot\}$：对所有可能的 $X \in \mathcal{X}$ 求期望

注：

- 对数以 2 为底，单位为比特（bit）
- 对数以 e 为底，单位为奈特（nat）
- 根据假设，$0\log 0 = 0$

熵是信息论的核心概念！

举例：投掷硬币

- 50% 正面，50% 反面

$$H(X) = -\sum_{x\in\mathcal{X}} p_x \log p_x$$
$$= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$\Rightarrow$1 bit

- 10% 正面，90% 反面

$-0.1\log_2 0.1 - 0.9\log_2 0.9 \approx 0.47$

$\Rightarrow$ 0.47 bit

- 独立投掷两次

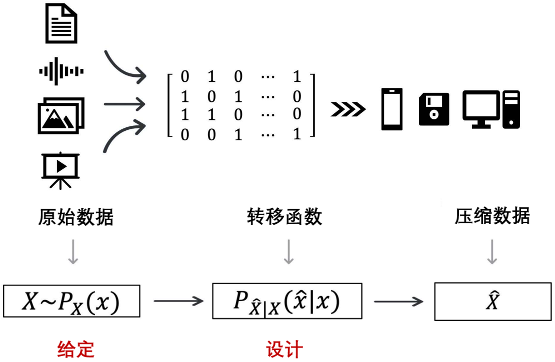
$\Rightarrow$ 投掷一次信息量的两倍

Q4: 有啥用？

物理意义

- 熵： 描述一次实验结果的最小平均比特数

信息压缩



目标：减少存储数据需要的空间

八卦图


乾 qián 天
巽 xùn 風
坎 kǎn 水
艮 gèn 山
坤 kūn 地
震 zhèn 雷
離 lí 火
兑 duì 澤
111
011
010
001
000
100
101
110

- “– –” 表示 0
- “——” 表示 1

赛马

四匹马比赛，假设它们赢的概率为：

马	概率
A	1/2
B	1/4
C	1/8
D	1/8

传输信息描述比赛结果，至少需要多少比特？

举例：赛马

马	概率	消息
A	1/2	00
B	1/4	01
C	1/8	10
D	1/8	11

(a) 2 bits

马	概率	消息
A	1/2	0
B	1/4	10
C	1/8	110
D	1/8	111

(b) 期望长度

$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 2(3 \cdot \frac{1}{8}) = 1.75$ bits

- (b) 方案节约了 0.25 比特
- 比赛结果的信息熵为

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - 2\left(\frac{1}{8}\log\frac{1}{8}\right) = 1.75.$$

熵 $H(X)$：描述随机变量 X 的最少比特数

信息的度量—熵

1. 熵
2. 熵的性质
3. 联合熵、条件熵

熵的性质

- $H(X) \leq \log |\mathcal{X}|$

熵的性质

- $H(X) \leq \log |\mathcal{X}|$

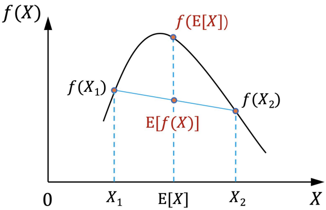
  等号成立条件为 X 在 $\mathcal{X}$ 上均匀分布，即均匀分布最大化信息熵

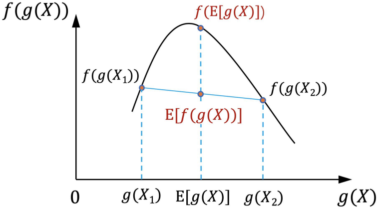
熵的性质

- $H(X) \leq \log|\mathcal{X}|$

  等号成立条件为 X 在 $\mathcal{X}$ 上均匀分布，即均匀分布最大化信息熵

证明： $\mathrm{E}_X\left\{\log\frac{1}{p_X}\right\} \leq \log\underbrace{\mathrm{E}_X\left\{\frac{1}{p_X}\right\}}_{\sum_{x\in\mathcal{X}} p_x/p_x} = \log|\mathcal{X}|$，因为 $\log(\cdot)$ 是凸函数。





$$f(x) = \log_2(x), \quad g(x) = 1/x$$

熵的性质

- $H(X) \leq \log |\mathcal{X}|$

  等号成立条件为 X 在 $\mathcal{X}$ 上均匀分布，即均匀分布最大化信息熵

熵的性质

- $H(X) \le \log |\mathcal{X}|$

  等号成立条件为 X 在 $\mathcal{X}$ 上均匀分布，即均匀分布最大化信息熵

- $H(C) = 0$，其中 C 是常数

  因为常数没有随机性

熵的性质

- $H(X) \leq \log|\mathcal{X}|$

  等号成立条件为 X 在 $\mathcal{X}$ 上均匀分布，即均匀分布最大化信息熵

- $H(C) = 0$，其中 C 是常数

  因为常数没有随机性

- $H(X) \geq 0$

  因为 $0 < p_x \leq 1$，所以 $\log \frac{1}{p_x} \geq 0$，因此 $H(X) = \mathrm{E}\left\{\log \frac{1}{p_x}\right\} \geq 0$

二元随机变量

- 二元随机变量 $X \in \mathcal{X} = \{0, 1\}$ 服从伯努利分布:

$$\Pr(x = 0) = 1 - p,$$
$$\Pr(x = 1) = p,$$

其中 $0 \leq p \leq 1$.

- p=0:

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

- p=0.1:

0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0

- p=0.5:

1 0 1 0 1 0 0 1 1 1 0 1 0 0 0 1 0 0 1 1 0 0 0 1 0 0 1 1 0 1 0 0 1 1 0 0 1 0

- p=1:

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

二元随机变量的熵

$$H(X) = -\sum_{x\in\mathcal{X}} p_x \log p_x$$

- X 是伯努利分布的二元随机变量：

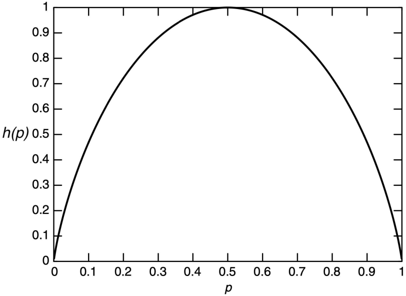
$$H(x) = h(p) = -p\log p - (1-p)\log(1-p)$$

二元随机变量的熵

$$H(X) = -\sum_{x \in \mathcal{X}} p_x \log p_x$$

- X 是伯努利分布的二元随机变量：

$$H(x) = h(p) = -p \log p - (1-p) \log(1-p)$$



信息的度量—熵

1. 熵
2. 熵的性质
3. 联合熵、条件熵

联合熵

联合熵：一组随机变量 $(X, Y) \sim p_{X,Y}$ 的不确定性：

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{x,y} \log p_{x,y}$$

联合熵

联合熵：一组随机变量 $(X, Y) \sim p_{X,Y}$ 的不确定性：

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{x,y} \log p_{x,y}$$

$p(Y, X)$ \ X / Y	1	2
1	1/3	1/3
2	1/3	0

$$H(X, Y) = 3 \cdot \left(-\frac{1}{3} \log \frac{1}{3}\right) = \log 3 \text{ bits}$$

条件熵

1. **条件熵** $H(Y|x)$：给定 $X=x$ 后 Y 的不确定性：

$$H(Y|x) = -\sum_{y\in\mathcal{Y}} p_{y|x} \log p_{y|x}.$$

条件熵

1. **条件熵** $H(Y|x)$：给定 $X=x$ 后 Y 的不确定性：

$$H(Y|x) = -\sum_{y\in\mathcal{Y}} p_{y|x} \log p_{y|x}.$$

2. **条件熵** $H(Y|X)$：

$$H(Y|X) = \sum_{x\in\mathcal{X}} p_x H(Y|x) = \mathrm{E}_X\{H(Y|x)\}$$

条件熵

1. **条件熵** $H(Y|x)$：给定 $X=x$ 后 Y 的不确定性：

$$H(Y|x) = -\sum_{y\in\mathcal{Y}} p_{y|x} \log p_{y|x}.$$

2. **条件熵** $H(Y|X)$：

$$H(Y|X) = \sum_{x\in\mathcal{X}} p_x H(Y|x) = \mathrm{E}_X\{H(Y|x)\} \text{ 或}$$

$$H(Y|X) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{x,y} \log p_{y|x} = \mathrm{E}_{X,Y}\{-\log p_{y|x}\}$$

条件熵

1. **条件熵** $H(Y|x)$：给定 $X=x$ 后 Y 的不确定性：

$$H(Y|x) = -\sum_{y\in\mathcal{Y}} p_{y|x} \log p_{y|x}.$$

2. **条件熵** $H(Y|X)$：

$$H(Y|X) = \sum_{x\in\mathcal{X}} p_x H(Y|x) = \mathrm{E}_X\{H(Y|x)\} \text{ 或}$$

$$H(Y|X) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{x,y} \log p_{y|x} = \mathrm{E}_{X,Y}\{-\log p_{y|x}\}$$

注：$H(Y|X)$ **是一个数，而** $H(Y|x)$ **是事件** $X=x$ **的函数**

条件熵

$p(Y,X)$ Y \ X	1	2
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{3}$	0

条件熵

$p(Y,X)$ \ X / Y	1	2
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{3}$	0

$p(Y\|X)$ \ X / Y	1	2
1	$\frac{1}{2}$	1
2	$\frac{1}{2}$	0

条件熵

$p(Y,X)$ — Y \ X	1	2
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{3}$	0

$p(Y\|X)$ — Y \ X	1	2
1	$\frac{1}{2}$	1
2	$\frac{1}{2}$	0

X	p_X
1	$\frac{2}{3}$
2	$\frac{1}{3}$

条件熵

$p(Y,X)$ (Y \ X)	1	2
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{3}$	0

$p(Y\|X)$ (Y \ X)	1	2
1	$\frac{1}{2}$	1
2	$\frac{1}{2}$	0

X	p_X
1	$\frac{2}{3}$
2	$\frac{1}{3}$

- $H(Y|X=2)=0$ bit

条件熵

$p(Y,X)$ \ X / Y	1	2
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{3}$	0

$p(Y\|X)$ \ X / Y	1	2
1	$\frac{1}{2}$	1
2	$\frac{1}{2}$	0

X	p_X
1	$\frac{2}{3}$
2	$\frac{1}{3}$

- $H(Y|X=2) = 0$ bit
- $H(Y|X=1) = 2 \cdot \left(-\frac{1}{2}\log\frac{1}{2}\right) = 1$ bit

条件熵

$p(Y,X)$ \ X / Y	1	2
1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{1}{3}$	0

$p(Y\|X)$ \ X / Y	1	2
1	$\frac{1}{2}$	1
2	$\frac{1}{2}$	0

X	p_X
1	$\frac{2}{3}$
2	$\frac{1}{3}$

- $H(Y|X=2) = 0$ bit
- $H(Y|X=1) = 2 \cdot \left(-\frac{1}{2}\log\frac{1}{2}\right) = 1$ bit
- $$\begin{aligned} H(Y|X) &= p(X{=}1)H(Y|X{=}1) + p(X{=}2)H(Y|X{=}2) \\ &= \tfrac{2}{3}\cdot 1 + \tfrac{1}{3}\cdot 0 \\ &= \tfrac{2}{3}\ \text{bit} \end{aligned}$$

条件熵的性质

- **条件减少信息熵**: $H(Y|X) \leq H(Y)$

  但是 $H(Y|x) \nleq H(Y)$

条件熵的性质

- **条件减少信息熵**：$H(Y|X) \leq H(Y)$

  但是 $H(Y|x) \nleq H(Y)$

- 如果 X 和 Y 相互独立，则 $H(Y|X) = H(Y)$

条件熵的性质

- **条件减少信息熵**：$H(Y|X) \le H(Y)$

  但是 $H(Y|x) \not\le H(Y)$

- 如果 X 和 Y 相互独立，则 $H(Y|X) = H(Y)$

- **确知函数** g **的条件熵**：$H(g(X)|X) = 0$

条件熵的性质

- **条件减少信息熵**: $H(Y|X) \leq H(Y)$

  但是 $H(Y|x) \not\leq H(Y)$
- 如果 X 和 Y 相互独立，则 $H(Y|X) = H(Y)$
- **确知函数 g 的条件熵**: $H(g(X)|X) = 0$
- **以双射函数 g 为条件的熵**: $H(X|g(X)) = 0$

  —如果 g 不是双射函数则不成立

条件熵的性质

- **条件减少信息熵**：$H(Y|X) \leq H(Y)$

  但是 $H(Y|x) \not\leq H(Y)$

- 如果 X 和 Y 相互独立，则 $H(Y|X) = H(Y)$

- **确知函数 g 的条件熵**：$H(g(X)|X) = 0$

- **以双射函数 g 为条件的熵**：$H(X|g(X)) = 0$

  —如果 g 不是双射函数则不成立

- 一般情况下，$H(Y|X) \neq H(X|Y)$

熵的链式法则

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

熵的链式法则

- $H(X, Y) = H(X) + H(Y|X)$
- $H(X_1, X_2, \ldots, X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1}, \ldots, X_1)$

熵的链式法则

- $H(X,Y) = H(X) + H(Y|X)$
- $H(X_1, X_2, \ldots, X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1}, \ldots, X_1)$
- **联合熵的上界（相互独立）**:

$$H(X,Y) \leq H(X) + H(Y)$$

熵的链式法则

- $H(X, Y) = H(X) + H(Y|X)$
- $H(X_1, X_2, \ldots, X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1}, \ldots, X_1)$
- 联合熵的上界（相互独立）:

$$H(X, Y) \leq H(X) + H(Y)$$

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$$

小结

- 信息 $\leftrightarrow$ 不确定性 $\leftrightarrow$ 熵
- 熵是描述一个随机试验结果的最小比特数
- 单变量熵及其性质

  定义，非负性，上界（均匀分布）
- 条件熵及其性质

  条件减小熵，链式法则，函数的熵，联合熵上界（独立）
- 不仅要记住信息的数学定义，还要理解其物理意义

作业

- 复习授课内容
- 预习互信息和散度
- 独立完成习题
  - 2.1
  - 2.2
  - 2.4
  - 2.5
  - 2.6