引入

信息的定义

- 我们对信息的度量将依赖于概率分布 $p(x)$，因此我们想要寻找一个函数 $h(x)$，它是概率 $p(x)$ 的单调递增 function，表达了信息的内容。

  ① $h(x,y) = h(x) + h(y)$

  ② $p(x,y) = p(x)p(y)$

  $\Rightarrow h(x) = -\log p(x) \geq 0$ 低概率事件 x 对应于高的信息量

熵

- 期望：传输的平均信息

  $H[x] = -\sum p(x)\log p(x)$

△ $\lim_{p\to 0} p\log p = 0$

- 熵是传输一个随机变量状态值所需的比特位的下界

信息熵

离散变量	• X是离散变量，$p(x)=P_r(X=x)$ 信息熵为：$H(x)=-\sum p(x)\log p(x)$.
	△ $0\log 0=\lim_{a\to 0^+} a\log a=0$
性质	• $H(x)\geq 0$
e.g.	• $X=\begin{cases}1 & p\\ 0 & 1-p\end{cases}$ $H(x)=-p\log p-(1-p)\log(1-p)$
联合熵	• (X,Y) 离散 $H(X,Y)=-\sum_x\sum_y p(x,y)\log p(x,y)$
条件熵	• $(X,Y)\sim p(x,y)$ 则：$H(Y\|X)=\sum_{x\in X}p(x)H(Y\|X=x)$ $=\sum_{x\in X}p(x)\sum_{y\in Y}p(y\|x)\log p(y\|x)$ $=-\sum_{x\in X}\sum_{y\in Y}p(x,y)\log(y\|x)$
定理	• $H(X,Y)=H(X)+H(Y\|X)$.

连续变量的熵

- $H(x) = -\int_S f_X(x) \log f_X(x) dx$.

e.g.
- 均匀分布 $[0, a]$

  $H[x] = -\int_0^a \frac{1}{a} \log \frac{1}{a} dx = \log a$

  $\log a < 0 \quad a < 1$.

联合熵
- $H(x_1, \cdots x_n) = -\int f(x_1, \dots, x_n) \log f(x_1, \cdots, x_n) dx_1 \cdots dx_n$.