信息论基础

李　莹
liying2009@ecust.edu.cn

第二章：信息的度量

一、自信息和互信息

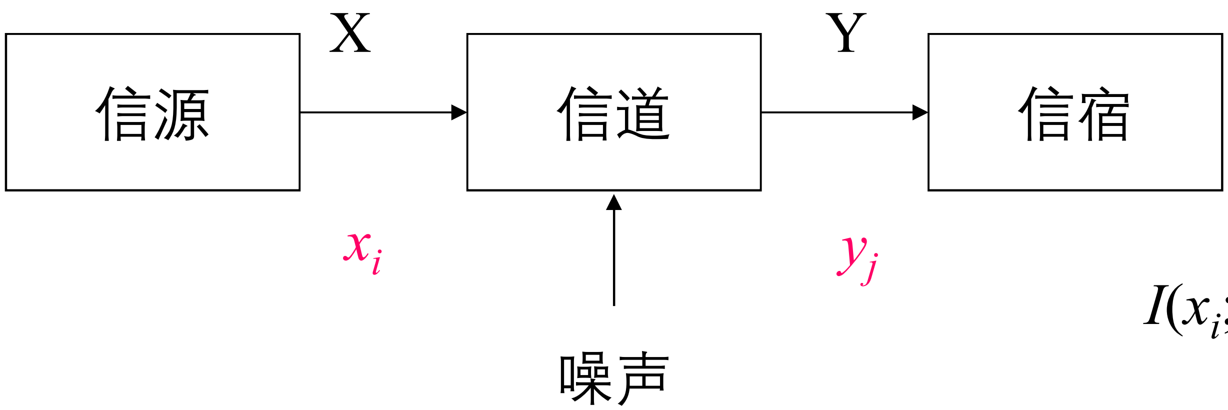
二、平均自信息

三、平均互信息

1. 平均互信息的概念
2. 平均互信息的性质
3. 数据处理定理

1. 平均互信息的概念

$I(X;Y)$



$I(x_i; y_j)$

平均交互信息量、交互熵

$$I(X;Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)I(x_i;y_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log\frac{p(x_i\mid y_j)}{p(x_i)}$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log\frac{1}{p(x_i)}-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_iy_j)\log p(x_i\mid y_j)$$

$$=H(X)-H(X\mid Y)$$

例　　对某城市进行交通忙闲的调查，并把天气分成晴雨两种状态，气温分成冷暖两种状态。调查结果得到的各数据联合出现相对频率如表所示，

忙				闲			
晴		雨		晴		雨	
冷	暖	冷	暖	冷	暖	冷	暖
12	8	27	16	8	13	4	12

若把这些频度看作概率测度，求：

（1）　忙闲的无条件熵；

（2）　天气状态和气温状态同时已知时忙闲的条件熵；

（3）从天气状态和气温状态同时获得的关于忙闲的信息量。

解：

$$X=\{忙，闲\}=\{0,1\}$$
$$Y=\{晴，雨\}=\{0,1\}$$
$$Z=\{冷，暖\}=\{0,1\}$$

则联合分布：

XY					
P(xyz)		00	01	10	11
X	0	0.12	0.08	0.27	0.16
	1	0.08	0.13	0.04	0.12
		$p_{yz}(00)=0.2$	$p_{yz}(01)=0.21$	$p_{yz}(10)=0.31$	$p_{yz}(11)=0.28$

$$H(X)=H(0.63,0.37)=0.951 \quad 比特/符号$$

$$H(X\mid YZ)=\mathrm{H}(XYZ)-H(YZ)=H(.12,.08,.27,.16,.08,.13,.04,.12)$$
$$-H(.2,.21,.31,.28)=2.819-1.976=0.843 \quad 比特/符号$$

$$I(X;YZ)=H(X)-H(X/YZ)=0.951-0.843=0.108 \quad 比特/符号$$

平均互信息与各类熵的关系

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y) \\
&= H(Y) - H(Y \mid X) \\
&= H(X) + H(Y) - H(XY)
\end{aligned}
$$

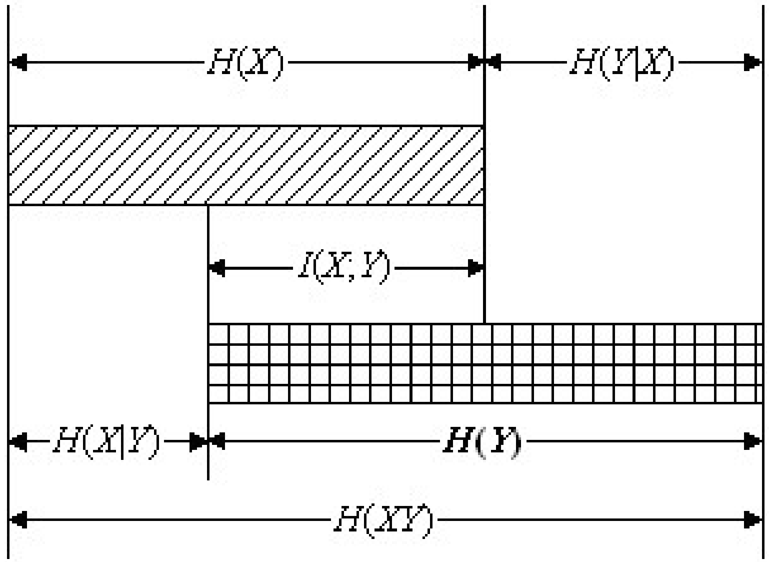


图1.3 平均互信息和各类熵之间的关系

2. 平均互信息的性质

平均互信息量不是从两个具体消息出发，而是从随机变量X和Y的整体角度出发，在平均意义上观察问题，所以平均互信息不会出现负值。

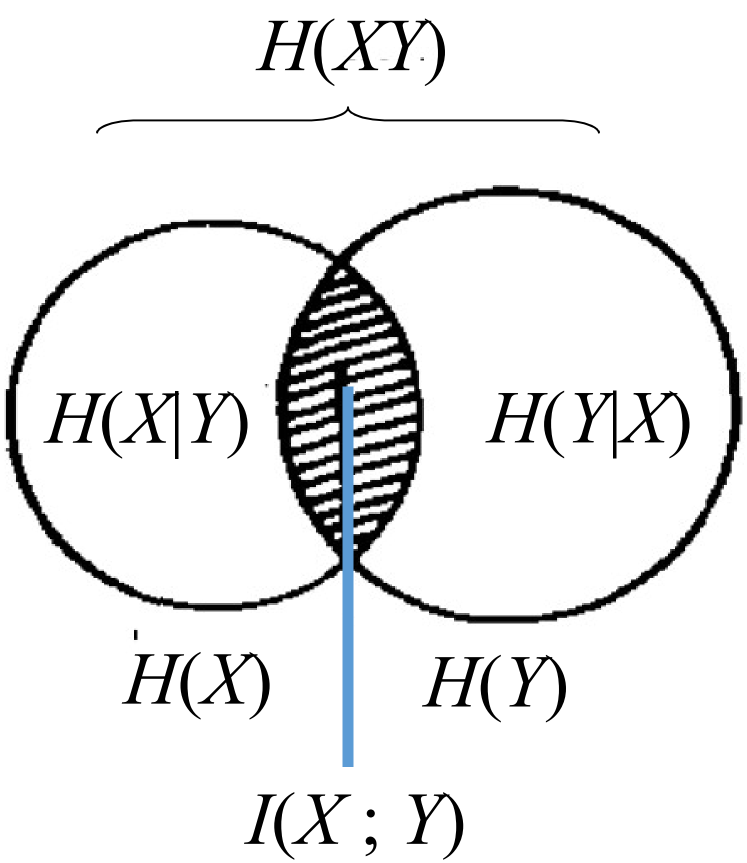
1. 对称性：$I(X;Y)=I(Y;X)$
2. 非负性： $I(X;Y)\geq 0$
3. 极值性： $I(X;Y)\leq \min\{H(X),H(Y)\}$

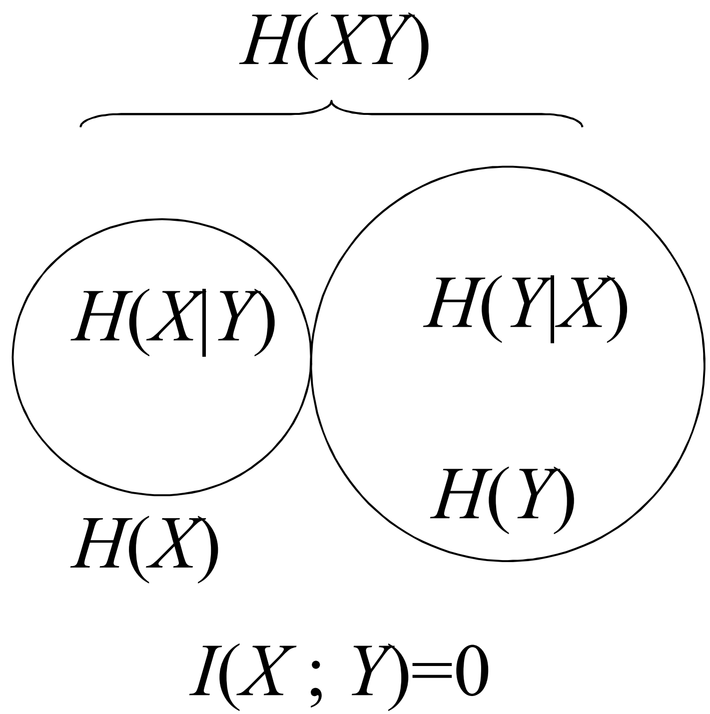
4. 平均互信息和各类熵的关系：

$$I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)=H(X)+H(Y)-H(XY)$$

5. $I(X;Y)$是信源概率分布$p(x_i)$的上凸函数和信道传递概率$p(y_j|x_i)$的下凸函数。

$$I(X;Y)=f[p(x_i),p(y_j|x_i)]$$


H(XY)
H(X|Y)
H(Y|X)
H(X)
H(Y)
I(X ; Y)


H(XY)
H(X|Y)
H(Y|X)
H(X)
H(Y)
I(X ; Y)=0

例　　已知二元随机变量X、Y，输出符号均为$\{0,1\}$，$p_X(0)=\omega$，$(0\le\omega\le 1)$，条件概率$p_{Y|X}(0|0)=p_{Y|X}(1|1)=1-p$，$(0\le p\le 1)$，求$I(X;Y)$，并讨论其凸函数性。

解:　　已知　$p_{Y|X}(0|0)=p_{Y|X}(1|1)=1-p$,

故有　$p_{Y|X}(1|0)=p_{Y|X}(0|1)=p$

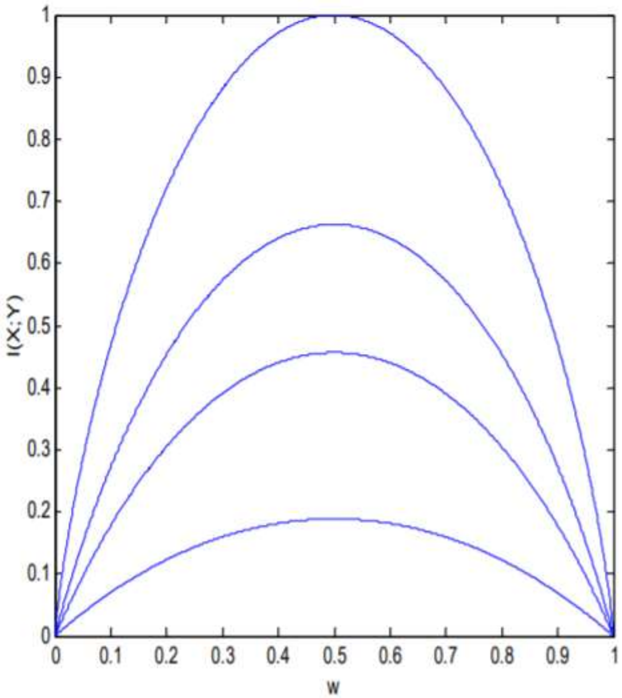
$$\left(p_Y(0)\quad p_Y(1)\right)=\left(\omega\quad 1-\omega\right)\begin{pmatrix}1-p & p\\ p & 1-p\end{pmatrix}$$

$$=\left(p+\omega-2\omega p\quad 1-p-\omega+2\omega p\right)$$

故得 $H(Y)=H(p+\omega-2\omega p)$, $H(Y|X)=\omega H(p)+(1-\omega)H(p)=H(p)$

因此 $I(X;Y)=H(p+\omega-2\omega p)-H(p)$

$I(X;Y)$ 是 ω 的上凸函数



当 $p=1/2$时， $I(X;Y)=0$；

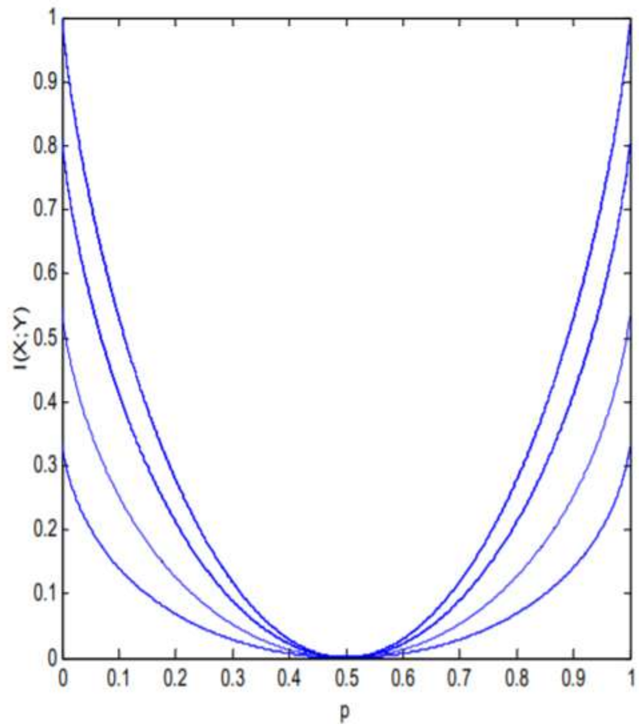
当$p\neq 1/2$有如下结论：

（1）因$I(\omega,p)=I(\omega,1-p)$，所以p和1−p对应的是同一条曲线；

（2）因$I(\omega,p)=I(1-\omega,p)$，所以曲线关于$\omega=1/2$对称；

（3）当 $p+\omega-2\omega p=1/2$ 时，有 $(1-2\omega)(p-1/2)=0$，所以当$\omega=1/2$ 时，$I(X;Y)=\log 2-H(p)$，达到极大值，当 $\omega=0$ 或1时，$I(X;Y)$取最小值0。

$I(X;Y)$ 是p的下凸函数



当 $\omega=0$或1时$I(X;Y)=0$

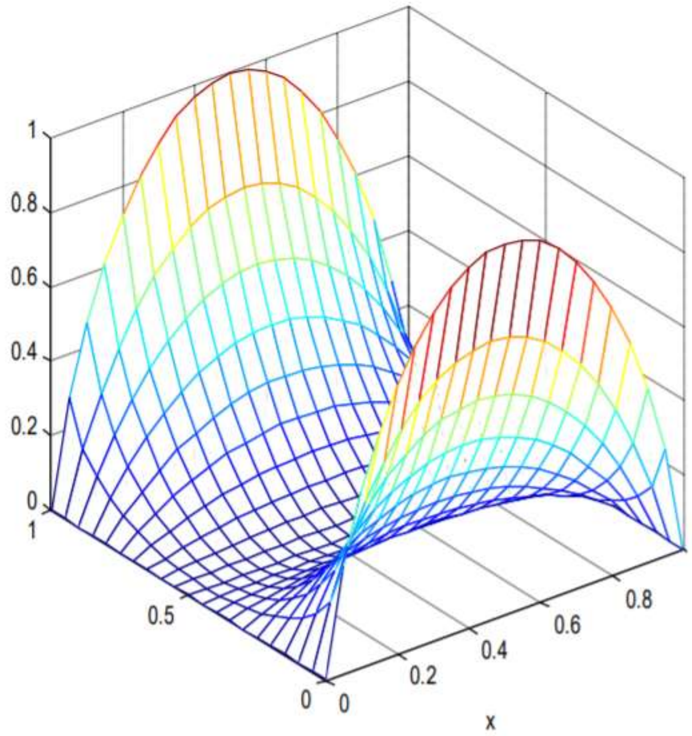
$\omega\neq 0$或1时有如下结论：

（1）因 $I(\omega,p)=I(1-\omega,p)$，所以$\omega$ 和 $1-\omega$ 对应的是同一条曲线；

（2）因 $I(\omega,p)=I(\omega,1-p)$，所以曲线关于 $p=1/2$ 对称；

（3）当 $p=0$ 或1时，$I(X;Y)=H(\omega)$或$H(1-\omega)$ 达到最大值，当 $p=1/2$ 时，$I(X;Y)$取最小值0。

$I(\omega, p)$的图形



3. 数据处理定理

Z条件下，X与Y之间的平均互信息

定义2.7 平均条件互信息

$$I(X;Y|Z)=E\left[I(x;y|z)\right]=\sum_x\sum_y\sum_z p(xyz)\log\frac{p(x|yz)}{p(x|z)}$$

定义2.8 平均联合互信息

$$I(X;YZ)=E\left[I(x;yz)\right]=\sum_x\sum_y\sum_z p(xyz)\log\frac{p(x|yz)}{p(x)}$$

可以证明

$$I(X;YZ)=\sum_x\sum_y\sum_z p(xyz)\log\frac{p(x|z)\cdot p(x|yz)}{p(x)\cdot p(x|z)}=I(X;Z)+I(X;Y|Z)$$

同理 $I(X;YZ)=I(X;Y)+I(X;Z|Y)$

定义　马尔科夫链

如果随机变量 Z的分布仅依赖于Y的分布，而与X是独立的，则称随机变量X,Y,Z构成一个马尔可夫链，记为 $X \to Y \to Z.$

即要求

$$p(xyz) = p(x)p(y \mid x)p(z \mid y)$$

定理2.3 （数据处理定理）

如果随机变量 X，Y，Z构成一个马尔可夫链，则有以下关系成立：

$$I(X;Z) \le I(X;Y)$$

$$I(X;Z) \le I(Y;Z)$$

等号成立的条件分别是对于任意的 x，y，z

$$p(x \mid yz) = p(x \mid z)$$

$$p(z \mid xy) = p(z \mid x)$$

证明：

$$I(X;YZ)=I(X;Z)+\underbrace{I(X;Y|Z)}_{\geq 0}=I(X;Y)+\underbrace{I(X;Z|Y)}_{=0}$$

$$\Downarrow$$

$$I(X;Z)\leq I(X;Y)$$

等号成立：

$$I(X;Y|Z)=0$$

$$I(X;Y|Z)=E\left[I(x;y|z)\right]=\sum_x\sum_y\sum_z p(xyz)\log\frac{p(x|yz)}{p(x|z)}$$

$$p(x|yz)=p(x|z)$$

同理可证其他结论。

例　已知：X为随机变量，$Y = f(X)$

判断： $H(Y) \le H(X)$ **？**

解：　$I(X;Y) = I(Y;X)$

$$H(X) - \underbrace{H(X|Y)}_{\geq 0} = H(Y) - \underbrace{H(Y|X)}_{=0}$$

$$H(X) \geq H(Y)$$