Joint Entropy, Conditional entropy and Mutual Information

An informal introduction

Consider the following two statements

1. Sindhu, an Indian won medal in Olympics, in badminton.
2. India has got one Silver medal in Olympics

- Both the statements have something in common and something not.
  - Common: India won medal in Olympics
  - Unique to 1st statement: Who got? & Which game?
  - Unique to 2nd statement: What medal?

Joint entropy

$$H(X,Y) = \sum_{xy \in \mathcal{A}_X \mathcal{A}_Y} P(x,y) \log \frac{1}{P(x,y)}$$

Independent random variables

- Two independent probability distributions $P_X = \{p_1, . . . , p_N\}$ & $Q_Y = \{q_1, . . . , q_M\}$

$$H = -\sum_{i=1}^{N} \sum_{j=1}^{M} p(i, j) \log (p(i, j))$$

$$p(i, j) = p(i) q(j)$$

$$H = -\sum_{i=1}^{N} \sum_{j=1}^{M} p(i) q(j) \log (p(i) q(j))$$

$$= -\sum_{i=1}^{N} \sum_{j=1}^{M} p(i) q(j) [\log p(i) + \log q(j)]$$

$$= -\sum_{i=1}^{N} \sum_{j=1}^{M} p(i) q(j) \log p(i) + -\sum_{i=1}^{N} \sum_{j=1}^{M} p(i) q(j) \log q(j)$$

$$= -\sum_{i=1}^{N} p(i) \log p(i) \{q_1 + \ldots + q_M\} + -\sum_{j=1}^{M} q(j) \log q(j) \{p_1 + \ldots + p_N\}$$

$$\{q_1 + \ldots + q_M\} = \{p_1 + \ldots + p_N\} = 1$$

$$H = -\sum_{i=1}^{N} p(i) \log p(i) + -\sum_{j=1}^{M} q(j) \log q(j)$$

$$= H_{P_X} + H_{Q_Y}$$

Conditional entropy of X given $y=b_k$

$$H(X \mid y = b_k) \equiv \sum_{x \in \mathcal{A}_X} P(x \mid y = b_k) \log \frac{1}{P(x \mid y = b_k)}$$

Conditional entropy of X given $y=b_1$

$$H(X \mid y=b_1) \equiv \sum_{x \in \mathcal{A}_X} P(x \mid y=b_1) \log \frac{1}{P(x \mid y=b_1)}$$

Conditional entropy of X given $y=b_2$

$$H(X \mid y=b_2) \equiv \sum_{x \in \mathcal{A}_X} P(x \mid y=b_2) \log \frac{1}{P(x \mid y=b_2)}$$

Conditional entropy of X given $y=b_3$

$$H(X \mid y=b_3) \equiv \sum_{x \in \mathcal{A}_X} P(x \mid y=b_3) \log \frac{1}{P(x \mid y=b_3)}$$

Conditional entropy of X given Y (for various y s)

$$P(y=b_1).\sum_{x \in \mathcal{A}_X} P(x \mid y=b_1) \log \frac{1}{P(x \mid y=b_1)}$$

$$+ P(y=b_2).\sum_{x \in \mathcal{A}_X} P(x \mid y=b_2) \log \frac{1}{P(x \mid y=b_2)}$$

$$+ P(y=b_3).\sum_{x \in \mathcal{A}_X} P(x \mid y=b_3) \log \frac{1}{P(x \mid y=b_3)}$$

$+ \ldots$

Recall, P(y).P(x|y) = P(x,y)

Conditional entropy of X given Y (i.e. all possible y s)

$$\sum_{xy \in \mathcal{A}_X \mathcal{A}_Y} P(x, y) \log \frac{1}{P(x \mid y)}$$

Chain rule for information

$$\log \frac{1}{P(x,y)} = \log \frac{1}{P(x)} + \log \frac{1}{P(y \mid x)}$$

Information content of x and y

Information content of x

Information content of y given x

Chain rule for entropy

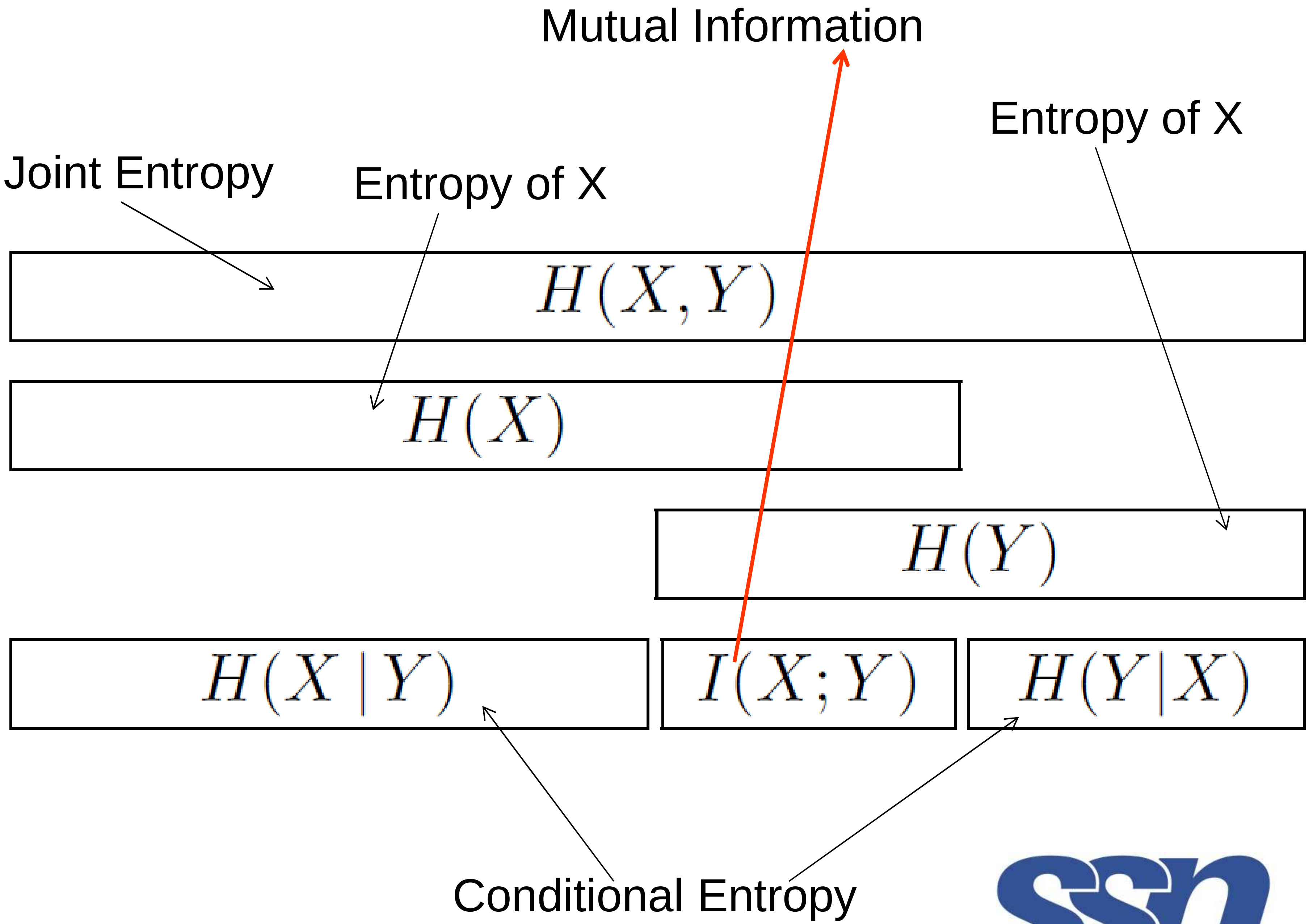
$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$$

Mutual Information

$$\underbrace{I(X;Y)} \equiv \underbrace{H(X)} - H(X|Y)$$

Entropy (information) about event X, after getting the knowledge of Y

Entropy (information) about event X, without the knowledge of Y


Mutual Information
Entropy of X
Joint Entropy
Entropy of X
$H(X,Y)$
$H(X)$
$H(Y)$
$H(X \mid Y)$
$I(X;Y)$
$H(Y|X)$
Conditional Entropy

Example – Find H(X,Y), H(X), H(Y), & H(X|Y)

$P(x,y)$		x = 1	2	3	4
y	1	$1/8$	$1/16$	$1/32$	$1/32$
	2	$1/16$	$1/8$	$1/32$	$1/32$
	3	$1/16$	$1/16$	$1/16$	$1/16$
	4	$1/4$	0	0	0

Solution for H(X,Y)

Find out information content

	1	2	3	4
1	1/8	1/16	1/32	1/32
2	1/16	1/8	1/32	1/32
3	1/16	1/16	1/16	1/16
4	1/4	0	0	0

	1	2	3	4
1	3	4	5	5
2	4	3	5	5
3	4	4	4	4
4	2	0	0	0

Find out average information content

	1	2	3	4
1	1/8	1/16	1/32	1/32
2	1/16	1/8	1/32	1/32
3	1/16	1/16	1/16	1/16
4	1/4	0	0	0

X

	1	2	3	4
1	3	4	5	5
2	4	3	5	5
3	4	4	4	4
4	2	0	0	0

	1	2	3	4
1	3/8	4/16	5/32	5/32
2	4/16	3/8	5/32	5/32
3	4/16	4/16	4/16	4/16
4	2/4	0	0	0

Find out average information content

	1	2	3	4
1	3/8	4/16	5/32	5/32
2	4/16	3/8	5/32	5/32
3	4/16	4/16	4/16	4/16
4	2/4	0	0	0

Sum up

108/32 = 27/8 = 3.4 bits

Solution for H(X) and H(Y)

$P(x,y)$		x = 1	2	3	4	$P(y)$	I(y)
y	1	$1/8$	$1/16$	$1/32$	$1/32$	$1/4$	2
	2	$1/16$	$1/8$	$1/32$	$1/32$	$1/4$	2
	3	$1/16$	$1/16$	$1/16$	$1/16$	$1/4$	2
	4	$1/4$	0	0	0	$1/4$	2
$P(x)$		$1/2$	$1/4$	$1/8$	$1/8$		
I(x)		1	2	3	3		

Find average of H(X)

$P(x)$	$1/2$	$1/4$	$1/8$	$1/8$

X

I(x)	**1**	**2**	**3**	**3**
=	1/2	2/4	3/8	3/8

Sum up

14/8 = 7/4 = 1.75 bits

Find average of H(Y)

$P(y)$
$1/4$
$1/4$
$1/4$
$1/4$

X

$\mathbf{I}(y)$	=
2	2/4
2	2/4
2	2/4
2	2/4

Sum up
2 bits

Solution for H(X|Y)

Recall, $P(x,y) = P(x|y).P(y)$

Solution for H(X|Y)

$P(x,y)$		x				$P(y)$
		1	2	3	4	
	1	$1/8$	$1/16$	$1/32$	$1/32$	$1/4$
y	2	$1/16$	$1/8$	$1/32$	$1/32$	$1/4$
	3	$1/16$	$1/16$	$1/16$	$1/16$	$1/4$
	4	$1/4$	0	0	0	$1/4$
$P(x)$		$1/2$	$1/4$	$1/8$	$1/8$	

$P(x \mid y)$		x			
		1	2	3	4
	1	$1/2$	$1/4$	$1/8$	$1/8$
y	2	$1/4$	$1/2$	$1/8$	$1/8$
	3	$1/4$	$1/4$	$1/4$	$1/4$
	4	1	0	0	0

Find out information content of (X|Y)

	1	2	3	4
1	1/2	1/4	1/8	1/8
2	1/4	1/2	1/8	1/8
3	1/4	1/4	1/4	1/4
4	1	0	0	0

	1	2	3	4
1	1	2	3	3
2	2	1	3	3
3	2	2	2	2
4	0	0	0	0

Find out average information content

	1	2	3	4
1	1/2	1/4	1/8	1/8
2	1/4	1/2	1/8	1/8
3	1/4	1/4	1/4	1/4
4	1	0	0	0

X

	1	2	3	4
1	1	2	3	3
2	2	1	3	3
3	2	2	2	2
4	0	0	0	0

	1	2	3	4
1	1/2	2/4	3/8	3/8
2	2/4	1/2	3/8	3/8
3	2/4	2/4	2/4	2/4
4	0	0	0	0

	1	2	3	4	(sum-up row wise) H(X|Y)	
1	1/2	2/4	3/8	3/8	14/8	← H(X|Y=1)
2	2/4	1/2	3/8	3/8	14/8	← H(X|Y=2)
3	2/4	2/4	2/4	2/4	2	← H(X|Y=3)
4	0	0	0	0	0	← H(X|Y=4)

- H(X) = 1.75 bits
- H(X|Y=4) = 0 & is less than H(X) - **learning decreases entropy**
- H(X|Y=3) = 2 & is greater than H(X) - **learning increases entropy**
- H(X|Y=2) = 1.75 & is equal to H(X) - **learning doesn't change entropy**
- H(X|Y=1) = 1.75 & is equal to H(X) - **learning doesn't change entropy**

	1	2	3	4	(sum-up row wise) H(X|Y)
1	1/2	2/4	3/8	3/8	14/8
2	2/4	1/2	3/8	3/8	14/8
3	2/4	2/4	2/4	2/4	2
4	0	0	0	0	0

Find average to get H(X|Y):

$=P(Y=1).\ H(X|Y=1)+P(Y=2).\ H(X|Y=2)$

$+P(Y=3).\ H(X|Y=3)+P(Y=4).\ H(X|Y=4)$

$=(14/8).(1/4) + (14/8).(1/4) + 2.(1/4)+0$

$=14/32+14/32+2/4 = 44/32 = 11/8 = 1.37$ bits

On an average, $H(Y|X) < H(X)$ i.e. $1.37 < 1.75$

Mutual Information

$$I(X;Y) \equiv H(X) - H(X \mid Y)$$

$= 1.75 – 1.37 = 0.38$ bits

- Compute $H(Y|X) = 1.62$ bits
- $H(Y) = 2$ bits
- $I(X;Y) = H(Y) – H(Y|X) = 2- 1.62 = 0.38$ bits