Information and Computation

Lecture 1

Joint entropy, conditional entropy and mutual information

A probability refresher

- **Definition.** A joint ensemble XY is a tuple $(\mathrm{A}_X \times A_Y, p_{XY})$ with

$$p_X(x\,) = \sum_{y \in A_Y} p_{XY}(x, y)$$

$$p_Y(y\,) = \sum_{x \in A_Y} p_{XY}(x, y)$$

$$p_{XY}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

Joint entropy

- **Definition.** The entropy (rate) of an ensemble XY is

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

- **Example.**

$p(\text{tails}, \text{tails}) = \qquad = \frac{1}{4}$

$H(XY) = -\sum \frac{1}{4} \log \frac{1}{4}$

$= \frac{1}{4}(2 + 2 + 2 + 2) = 2$

Conditional entropy

- **Definition**. Conditional information,

$$h(x|y) = -\log(p(x|y))$$

- **Definition**. Conditional entropy given event y

$$H(X|y) = -\sum_{x \in A_X} p(x|y) \log(p(x|y))$$

- **Definition**. Conditional entropy

$$H(X|Y) = \sum_{y} p_Y(y) H(X|y)$$

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. How much information do the messages 0 and 1 carry?

	Days with no rain	Days with rain
Rotterdam	212	153
Atacama desert	360	5

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. How much information do the messages 0 and 1 carry?

	Days with no rain	Days with rain
Summer	182	1
Winter	30	152

Exercise 1.1 - Let us assume that you are living in the Atacama desert where it rarely rains. You receive a 0. How much information does this message carry?

Exercise 1.2 - Now let us assume that you live in the Netherlands where it does rain quite often, but certainly not every day. You receive a 0. Does the message contain information?

Exercise 1.3 - Finally, let us assume that you live in the Netherlands but (boldly) also that you are aware of the current season. You receive a 0. Does the message 0 carry the same information in summer and in winter?

	Days with no rain	Days with rain
Summer	182	1
Winter	30	152

- Compute $H(\text{Season}), H(\text{Weather})$
- Compute $H(\text{Season,Weather})$
- Compute $H(\text{Season}|\text{Weather=rain})$

P	No R	RAIN
Sum	$\frac{182}{365}$	$\frac{1}{365}$
Wint	$\frac{30}{365}$	$\frac{152}{365}$

Exercise

- Show that $H(X|Y) \geq 0$

Exercise. Chain rule

- Show that $H(XY) = H(X|Y) + H(Y)$

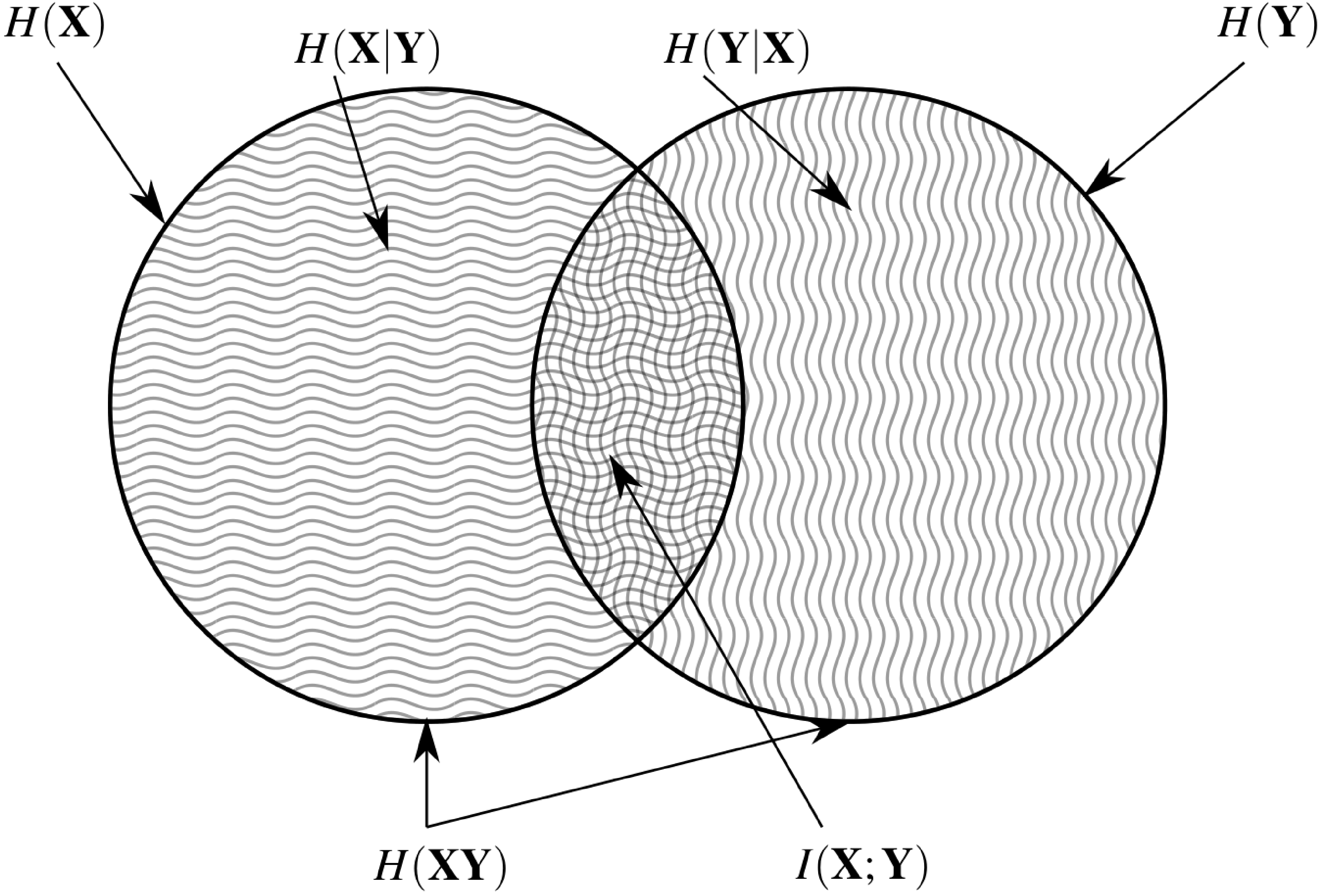
Exercise

- Show that
  - $H(X, Y) \geq H(X)$
  - $H(X) \geq H(X|Y)$ (conditioning always reduces entropy)

Mutual information

- **Definition**. Mutual information

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



$H(\mathbf{X})$
$H(\mathbf{X}|\mathbf{Y})$
$H(\mathbf{Y}|\mathbf{X})$
$H(\mathbf{Y})$
$H(\mathbf{XY})$
$I(\mathbf{X};\mathbf{Y})$

Exercise

- Show that
  - $\min\big(H(X), H(Y)\big) \geq I(X;Y) \geq 0$

Data processing inequality

- Show that $I(X;Y) \geq I(X;f(Y))$

Let $Z = f(Y)$ and let us expand $I(X:YZ)$ in two ways:

1) $I(X;YZ) = H(X) + H(YZ) - H(XYZ) + H(Y) - H(Y) + H(XY) - H(XY)$

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

$$= I(X:Y) + \underbrace{\cancel{H(Z|Y)}}_{0} - \underbrace{\cancel{H(Z|XY)}}_{0}$$

$$= I(X:Y)$$

2) $$I(X:YZ) = I(X:Z) + H(YZ) - H(XYZ) - H(Z) + H(XZ)$$

$$= I(X:Z) + H(X|Z) - H(XY|Z)$$

$$= I(X:Z) + \sum_z p(z)\Big(H(X|z) - H(XY|z)\Big)$$

$$= I(X:Z) + \sum_z p(z)\, I(X:Y|z)$$

Hence:

$$I(X:Y) = I(X:Z) + \sum_z p(z)\, I(X:Y|z)$$

$$\geq I(X:Z)$$

You will do great in the exam if you can

- Evaluate entropy, conditional entropy and mutual information
- Derive basic properties of the above

Resources

- Lecture notes
- Slides
- “Elements of information theory”. Cover and Thomas. Chapter 2.1-2.8
- “Information Theory, Inference, and Learning Algorithms”. MacKay. Chapter 2