EE 231A: Information Theory

Rick Wesel
Boelter 6412A
(310) 256-2150
wesel@ee.ucla.edu

1

EE 231E: Information Theory
Lecture 1

A. Introduction
B. Entropy
C. Relative Entropy
D. Mutual Information

2

Mutual information

- $I(X;Y)$ describes how much information X contains about Y and vice versa.
- Formal Definition is (2.28) on page 20, but that is not the first expression to learn.

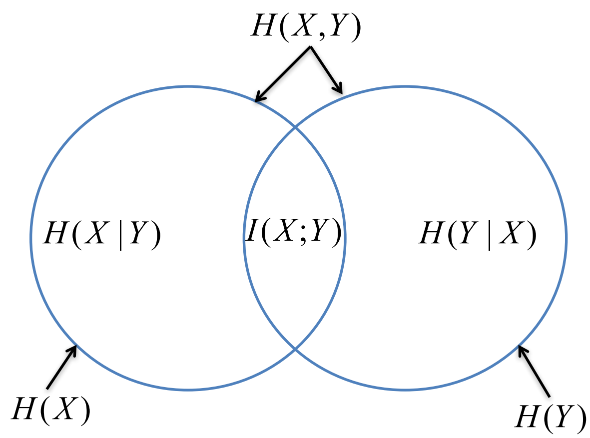
3

Operational Definition of Mutual information

$$
\begin{aligned}
I(X;Y) = I(Y;X) & \\
&= H(X) - H(X \mid Y) \\
&= H(Y) - H(Y \mid X) \\
&= H(X) + H(Y) - H(X,Y) \quad \text{Why?} \\
&= H(X) + H(Y) \underbrace{- H(Y) - H(X \mid Y)}_{-H(X,Y)} \\
&= H(X) + H(Y) - H(Y) - H(X \mid Y)
\end{aligned}
$$

4

Venn diagram



5

A special case of $D(p||q)$

- What if $q(x,y) = p(x)p(y)$

$$D(p(x,y)\,\|\,p(x)p(y)) = E_{p(x,y)}[-\log p(x)p(y)] - E[-\log p(x,y)]$$

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

$$= I(X;Y)$$

6

Mutual information and relative entropy

$$I(X;Y) \triangleq \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$
$$= D(p(x,y) \| p(x)p(y))$$

- So $I(X;Y)$ measures the penalty associated with compressing two dependent R.V.'s as if they were independent.
- i.e. $I(X;Y)$ measures how much information X has about Y and vice versa.

7

Information vs. Probability of Error

- The concept of mutual information provides a way for us to evaluate how informative some data is.
- i.e. Does this data tell me something I don't know?
- On your homework you will explore whether you should choose the weather forecaster having the lower probability of a wrong forecast or the higher mutual information.

8

Mutual information and entropy

- Let's use the formal definition to show

$$I(X;Y) = H(X) - H(X\,|\,Y)$$

- Proof:

$$\begin{aligned} I(X;Y) &\triangleq \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= \sum_x \sum_y p(x,y) \log \frac{p(x\,|\,y)}{p(x)} \\ &= \underbrace{\sum_x \sum_y p(x,y) \log p(x\,|\,y)}_{-H(X|Y)} - \sum_x \underbrace{\sum_y p(x,y)}_{p(x)} \log p(x) \\ &= H(X) - H(X\,|\,Y) \end{aligned}$$

9

Conditional mutual information

- $I(X;Y\,|\,Z) = H(X\,|\,Z) - H(X\,|\,Y,Z)$
- Chain rule for mutual information

$$I(X_1, X_2, ..., X_n; Y) = \sum_{i=1}^{n} I(X_i; Y\,|\,X_{i-1}, X_{i-2}, ..., X_1)$$

- Proof: $I(\underbrace{X_1, X_2, ..., X_n}_{X^n}; Y)$

$$\begin{aligned} &= H(Y) - H(Y\,|\,X^n) \\ &= H(X^n) - H(X^n\,|\,Y) \\ &= \sum_{i=1}^{n} H(X_i\,|\,X_{i-1}, X_{i-2}, ..., X_1) - \sum_{i=1}^{n} H(X_i\,|\,X_{i-1}, X_{i-2}, ..., X_1, Y) \\ &= \sum_{i=1}^{n} I(X_i; Y\,|\,X_{i-1}, X_{i-2}, ..., X_1) \end{aligned}$$

Which decomposition Is helpful?

10