What is Information Theory? The Basics

Sensor Reading Group

10 October 2003

Motivation

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

- C.E. Shannon, 1948

Not only for communication, but for sensing, classification, economics, computer science, mathematics, networking, …

The Main Idea

- Info theory examines the uncertain world.
  - Can't assume perfect communication (or sensing, classification, etc.), but we can characterize the uncertainty.
  - Quantities possess information, but are corrupted by noise.
- C.E. Shannon wrote "A Mathematical Theory of Communication"
  - Rigorous analysis of the essence of information

Entropy: What is H(X)?

- Entropy is a measure of the uncertainty in a random variable.
  - This is a function of the probability distribution, not of the individual samples of that distribution.
- Examples
  - 20 Questions – Entropy is the minimum expected # of yes/no questions needed to determine X.
  - Classification – Entropy represents the inherent uncertainty of the class of the target.

Entropy, H(X)

- Given a r.v. X with probability (mass) function $p(x)$, the entropy is defined as:

$$H(X) = -\sum_{x} p(x) \log p(x)$$

  - Example calculation

    Given two classes (e.g. friend or foe) that are distributed uniformly, the entropy, or uncertainty, in the class of the target is:

$$H(X) = -p(X = friend) \log p(X = friend) - p(X = foe) \log p(X = foe)$$

$$H(X) = -\tfrac{1}{2} \log \tfrac{1}{2} - \tfrac{1}{2} \log \tfrac{1}{2} = \log 2 = 1\, bit$$

Mutual Information: What is I(X;Y)?

- Mutual information is a measure of the amount of information shared between random variables.
  - What does knowing information about one thing tell you about another thing?
- Examples
  - Wheel of Fortune – You can guess what the missing words/letters are, based on the ones showing
  - Jeopardy – You're given the answer, you have to determine what it tells you about the question.

Mutual Information, I(X;Y)

- Given two (discrete) r.v.'s X, Y with pmf's $p(x)$, $p(y)$, $p(x,y)$, the mutual information is defined as:

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

$$= -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

- Jeopardy example:

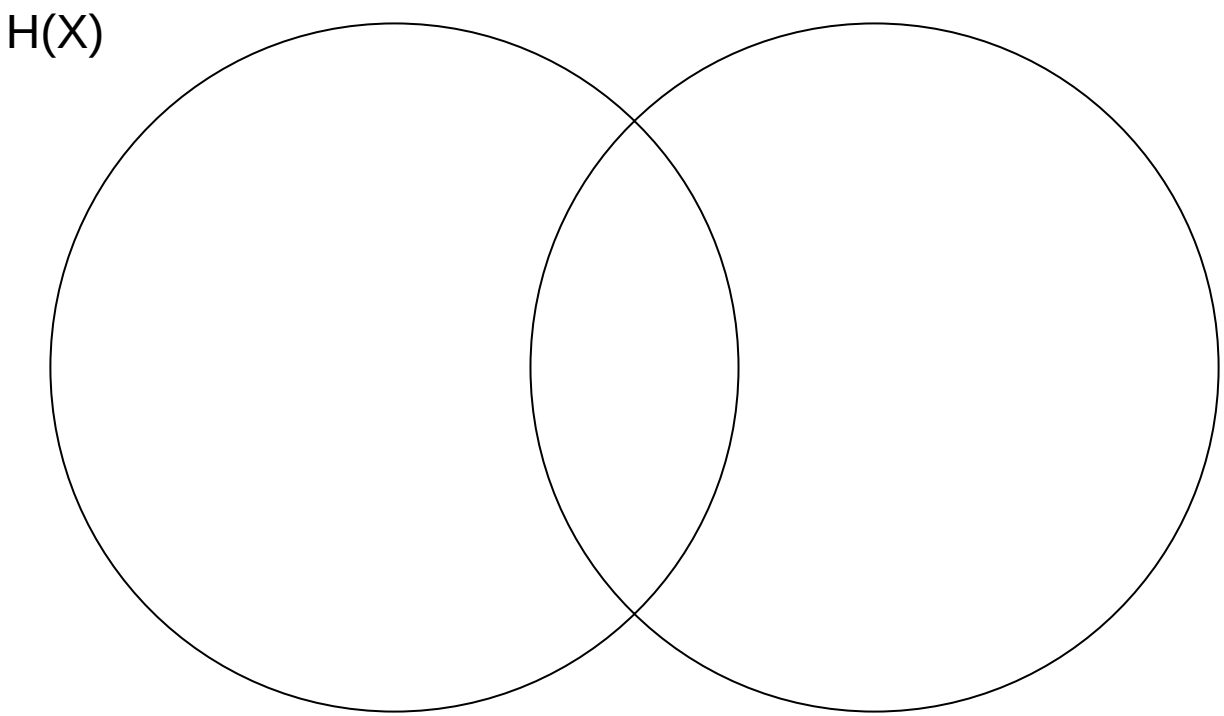
$$H(X) = I(X;Y) + H(X \mid Y)$$

"What is X?"

"The answer is"

This is the contestant's job.

Relationship between Entropy and Mutual Information



Fano's Inequality

- Probability of error of classification

$$P_e \geq \frac{H(X|Y)-1}{\log|\mathrm{X}|} = \frac{H(X)-I(X;Y)-1}{\log|\mathrm{X}|}$$

- This is the reason why some of us are interested in MAXIMIZING mutual information!