# More regarding noisy channels

Information Theory Lecture 5b

## Let's review what we know about entropies of two or more variables

- The joint entropy of $X,Y$ is

$$H(X,Y) = \sum_{xy \in A_x A_y} P(x,y)\log\frac{1}{P(x,y)}$$

$$H(X,Y) = H(X) + H(Y) \text{ iff } P(x,y) = P(x)P(y)$$

## Conditional Entropy

- Of $X$ given $y=b$ is the entropy of the probability distribution $P(x|y=b)$

$$H(X \mid y=b) = -\sum_{x \in A_x} P(x \mid y=b)\log\big(P(x \mid y=b)\big)$$

- This is the information that remains in $X$ after $y$ is known to be $b$

## Condition entropy

- Of X given Y is the average of the previous expression, over all possible values of $y$

$$H(X \mid Y) = -\left[\sum_{y \in A_y} P(y) \sum_{x \in A_x} P(x \mid y=b)\log\big(P(x \mid y=b)\big)\right]$$

$$= -\sum_{xy \in A_x A_y} P(x,y)\log\big(P(x \mid y)\big)$$

- This is the information that remains in $X$ after we know $Y$ in general

## Chain rule for entropy

- Relating the three previous expressions

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

- The information available in $X$ and $Y$ is the information in $X$ plus the information in $Y$ given $X$
- or vice versa

## Mutual information

- Between $X$ and $Y$
  - $I(X:Y) = H(X) - H(X|Y) = I(Y:X)$
  - $I(X:Y) \geq 0$
- This measures the average information obtained about $x$ given $y$, or vice versa

## Conditional Mutual Information

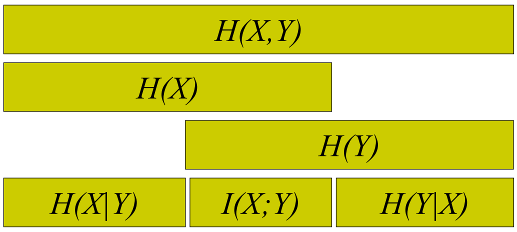
- Between $X$ and $Y$ given $z = C$

$$I(X;Y \mid z = C) = H(X \mid z = c) - H(X \mid Y, z = C)$$

- Averaging over all possible values of $Z$

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

## The relationship




## Let's return to the noisy channel

- The sender inputs symbol $x$, and we receive symbol $y$
- Our job is to infer $x$ given $y$

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)} = \frac{P(y \mid x)P(x)}{\sum_{x'} P(y \mid x')P(x')}$$

- But we also want to characterize average *rates* through this channel

## The Capacity of a Channel $Q$

- Is defined as the maximum information we can convey about $x$ by reading $y$
- We can accomplish this by picking the best probability distribution over $x$ (coding)

$$C(Q) = \max_{P_x} I(X;Y)$$

## The Noisy Typewriter Channel

- Consider a typewriter that sends one of 27 characters (A,B,…,Z,-)
- The letters are arranged in a circle, and the typist can "miss" and hit the higher or the lower character
- We can send information *perfectly* by only using every third character on the typewriter

## Shannon's Noisy Channel Coding Theorem

- Associated with each discrete, memoryless channel there is a non-negative capacity $C$ (called the channel capacity) with the following property:
- For any $\varepsilon > 0$ and $R < C$ there is a block code with block length $N$ and rate $\geq R$ and a decoding algorithm such that the maximal probability of block error is $< \varepsilon$

## For the noisy typewriter

| The Theorem | How it applies to the noisy typewriter |
|---|---|
| Channel capacity $C$ | $\log_2(27/3)$ |
| $\varepsilon$ and $R$ | We only need block length of $N=1$ |
| Block code of length $N$ | The block code only using every third character will require 3 characters to convey any one, so the rate is $\log_2(27/3)$ |
| Decoding algorithm | Map the received letter to the nearest code letter |
| Maximal probability of block error < $\varepsilon$ | Zero, in this case |

## Another version of the proof

- (not offered here)
- Like in the noisy typewriter, we could consider blocks at $x$ that map to non-overlapping $y$
- We then measure the density of these blocks in the possible input space
- This gives rate

## Pattern Recognition as Noisy Communication

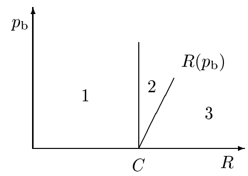
- Let's say we want to send symbols $A_x=\{0,1,2,3,4,\ldots 9\}$
- By writing characters in a 16 by 16 pixel box
- The input space is $A_x$
- The output space is $A_y = \{0,1\}^{256}$
- Our approach to pattern recognition is

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{\sum_{x'} P(y \mid x')P(x')}$$

## Beyond perfection

- If a bit-probability of error $p_b$ is acceptable, rates of up to $R(p_b)$ can be achieved

$$R(p_b) = \frac{C}{1 - H_2(p_b)}$$




- Rates higher than this cannot be achieved