2/17/2011

Lecture 14

Classical Information & Entropy

Entropy is the expected surprise upon learning the outcome of a random experiment

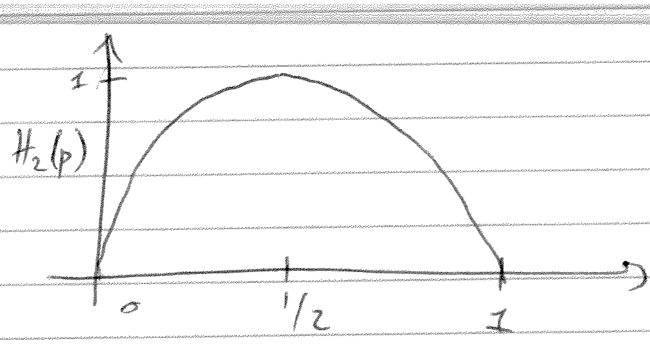
Review

$$i(x) \equiv -\log(p_X(x))$$

$$H(X) = \mathbb{E}_X\{i(X)\} = -\sum_x p_X(x)\log(p_X(x))$$

Special case: binary entropy function

$$H_2(p) = -p\log p - (1-p)\log(1-p)$$



concave as a function of p

max value at 1/2

2/17/2011

Mathematical properties of entropy

1) positivity (clear)

2) concavity (prove later)

3) Invariance under permutations

depends only on distribution - not on letters

4) minimum value = 0

for a deterministic variable

5) max. value is log d

$$H(X) \leq \log d$$

use Lagrangian optimization to figure out

$$\mathcal{L} \equiv H(X) + \lambda \left(\sum_x p_X(x) - 1 \right)$$

$$\frac{\partial \mathcal{L}}{\partial p_X(x)} = \frac{\partial \mathcal{L}}{\partial p(x)} \left[- \sum_x p(x) \log(p(x)) + \lambda \left(\sum_x p(x) - 1 \right) \right]$$

$$= - \log[p(x)] - 1 + \lambda$$

Null the partial derivative

$$0 = - \log p(x) - 1 + \lambda$$

$$\Rightarrow \quad p(x) = 2^{\lambda - 1}$$

↑ constant ⇒ uniform dist.

2/17/2011

Conditional Entropy

conditional information content $i(x|y) \equiv -\log[p(x|y)]$

entropy conditional on a particular value of y

is $H(X|Y=y) \equiv \mathbb{E}_X\{i(X|y)\}$

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

conditional entropy is expected info content

$$H(X|Y) = \mathbb{E}_{X,Y}\{i(X|Y)\}$$

$$= \sum_y p(y)\; H(X|Y=y)$$

$H(X|Y)$ quantifies uncertainty about X if Bob has access to Y

$\therefore$ intuitive that

$$H(X) \geq H(X|Y)$$

"conditioning does not increase entropy"

same statement as concavity of entropy

2/17/2011

Joint Entropy

don't know X or Y ?

uncertainty is $H(X,Y) \equiv \mathbb{E}_{X,Y} \{ i(X,Y) \}$

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

natural extension of entropy

can show that

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

similarly

$$= H(Y) + H(X|Y)$$

Thus, $H(X,Y) \leq H(X) + H(Y)$

Mutual Information

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

"how much knowing Y reduces the uncertainty about X"

can show it is symmetric

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

Also,

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

2/17/2011

$I(X;Y) \geq 0$

- can use that $I(X;Y) = H(X) - H(X|Y) \geq 0$

  & conditioning does not increase entropy

- max value is $\min\{\log d_X, \log d_Y\}$

Relative Entropy

How "far" is one distribution from another?

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

not symmetric, so not really a distance

interpretation in source coding as inefficiency

Suppose Alice gets sequences according to $p(x)$ but thinks it is $q(x)$ and codes this way. Then compression will require

$H(p) + D(p\|q)$ bits

Also, $I(X;Y) = D(p(x,y) \| p(x)p(y))$

mutual info. quantifies how far X,Y is from being independent.

2/17/2011

5a

Conditional Mutual Information

$$I(X;Y|Z) = \sum_z p(z)\; I(X;Y|Z=z)$$

or equivalently

$$I(X;Y|Z) = H(Y|Z) - H(Y|XZ)$$

What are the correlations between X & Y given Z?

Chain rule

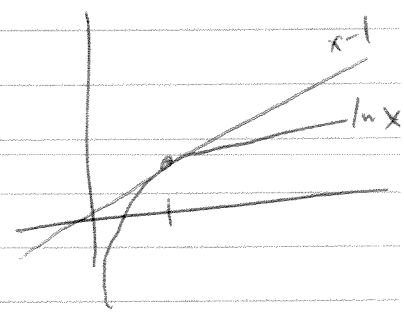
$$I(XZ;Y) = I(X;Y|Z) + I(Z;Y)$$

2/17/2011

Fundamental Classical Information Inequality

(All classical ~~results~~ information inequalities follow from this one)

$D(p||q) \geq 0$



prove w/ $\ln x \leq x-1$

$-\ln x \geq 1-x$

$$D(p||q) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

$$= \frac{-1}{\ln 2} \sum_x p(x) \ln\left(\frac{q(x)}{p(x)}\right)$$

$$\geq -\frac{1}{\ln 2} \sum_x p(x) \left(1 - \frac{q(x)}{p(x)}\right)$$

$$= -\frac{1}{\ln 2} \left(\sum_x p(x) - \sum_x q(x) \right) = 0$$

2/17/2011

Data Processing Inequality

$$X \longrightarrow \boxed{\mathcal{N}_1} \longrightarrow Y \longrightarrow \boxed{\mathcal{N}_2} \longrightarrow Z$$

$\mathcal{N}_1 \equiv p(y|x)$

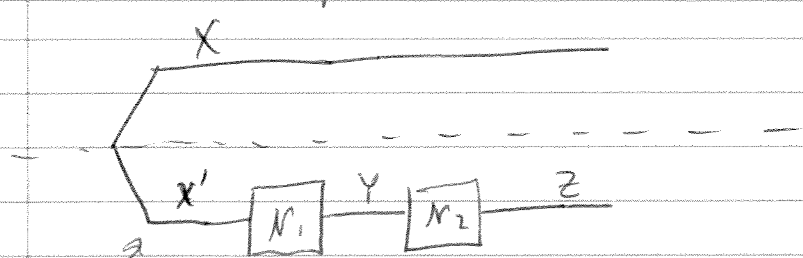
$\mathcal{N}_2 \equiv p(z|y)$

Markov property

Data Processing Ineq.: $I(X;Y) \geq I(X;Z)$

"~~data~~ noisy processing ~~decreases~~ cannot increase correlations"

Similar picture (useful when we get to quantum)



X' is a copy of X

2/17/2011

Proof $\quad p(z|y,x) = p(z|y)$ (Markov condition)

$$\therefore \; p(x,z|y) = p(z|y,x)\,p(x|y)$$
$$= p(z|y)\,p(x|y)$$

Consider

$$I(X;YZ) = I(X;Y) + \underbrace{I(X;Z|Y)}_{=0 \text{ b/c}}$$
$$= I(X;Y)$$

expand the other way

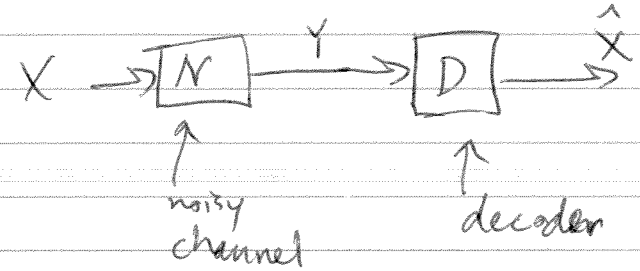
$$I(X;YZ) = I(X;Z) + I(X;Y|Z)$$

$$\therefore \; I(X;Y) = I(X;Z) + \underbrace{I(X;Y|Z)}_{\geq 0}$$

$$\Rightarrow \; I(X;Y) \geq I(X;Z)$$

2/17/2011

Fano's Inequality



probability of error is $\Pr\{\hat{X} \neq X\} \equiv p_e$

can think of $H(X|Y)$ as

given Y, how certain are you about X?

should be related to probability of error

Fano's is $H(X|Y) \leq H_2(p_e) + p_e \log(|\mathcal{X}|-1)$

number of letters for X

Proof: make an indicator R.V.

$$E = \begin{cases} 0 & : X = \hat{X} \\ 1 & : X \neq \hat{X} \end{cases}$$

Consider $H(EX|\hat{X}) = H(X|\hat{X}) + \underbrace{H(E|X\hat{X})}_{=0}$

$= H(X|\hat{X})$

2/17/2011

10

use data processing

$$I(X;Y) \geq I(X;\hat{X})$$

$$\therefore \quad H(X) - H(X|Y) \geq H(X) - H(X|\hat{X})$$

$$\therefore \quad H(X|\hat{X}) \geq H(X|Y)$$

Now consider

$$H(EX|\hat{X}) = H(E|\hat{X}) + H(X|E\hat{X})$$

$$\leq H(E) + H(X|E\hat{X})$$

$$= H_2(p_e) + p_e H(X|E=1,\hat{X}) +$$

$$(1-p_e) \underbrace{H(X|E=0,\hat{X})}_{\text{no error means no uncertainty}}$$

$$= H_2(p_e) + p_e H(X|E=1,\hat{X})$$

we know that $X \neq \hat{X}$
$\therefore$ $|\mathcal{X}|-1$ possibilities for X

$$\leq H_2(p_e) + p_e \log(|\mathcal{X}|-1)$$

given that $H(EX|\hat{X}) = H(X|\hat{X}) \geq H(X|Y)$, theorem follows.

2/17/2011

11

Quantum Entropy

suppose state is density operator ρ^A

Then its entropy is

$$H(A)_\rho = \underbrace{-\mathrm{Tr}\{\rho^A \log \rho^A\}}_{\text{How to compute?}}$$

diagonalize ρ^A as $\rho^A = \sum_x p(x)|x\rangle\langle x|$

Then

$$H(A)_\rho = -\mathrm{Tr}\{\sum_x p(x)|x\rangle\langle x| \log(\sum_{x'} p(x')|x'\rangle\langle x'|$$

$$= -\mathrm{Tr}\{\sum_x p(x)|x\rangle\langle x| \sum_{x'} \log p(x')|x'\rangle\langle x'|$$

$$= -\sum_{x',x} p(x)\log p(x')\,\mathrm{Tr}\{|x\rangle\langle x|x'\rangle\langle x'|\}$$

$$= \underbrace{-\sum_x p(x)\log p(x)}_{\text{Shannon entropy of eigenvalue}}$$

2/17/2011

Note that $H(A)_\rho = 0$ ~~where~~ when ρ is pure $|\psi\rangle\langle\psi|$

Joint Entropy

$$H(AB)_\sigma = -\mathrm{Tr}\{\sigma^{AB} \log \sigma^{AB}\}$$

where σ^{AB}

Mutual Information

$$I(A;B)_\sigma = H(A)_\sigma + H(B)_\sigma - H(AB)_\sigma$$

Consider pure bipartite state $|\psi\rangle^{AB}$. It has zero joint entropy $H(AB)_\psi = 0$

this cannot happen classically!!

do Schmidt decomposition

$$|\psi\rangle^{AB} = \sum_x \sqrt{p(x)}\, |x\rangle^A |x\rangle^B$$

i.e. $\mathrm{Tr}_A\{\psi\} = \sum_x p(x) |x\rangle\langle x|^B$

& $\mathrm{Tr}_B\{\psi\} = \sum_x p(x) |x\rangle\langle x|^A$

Thus $H(A)_\psi = H(B)_\psi$ & can be >0