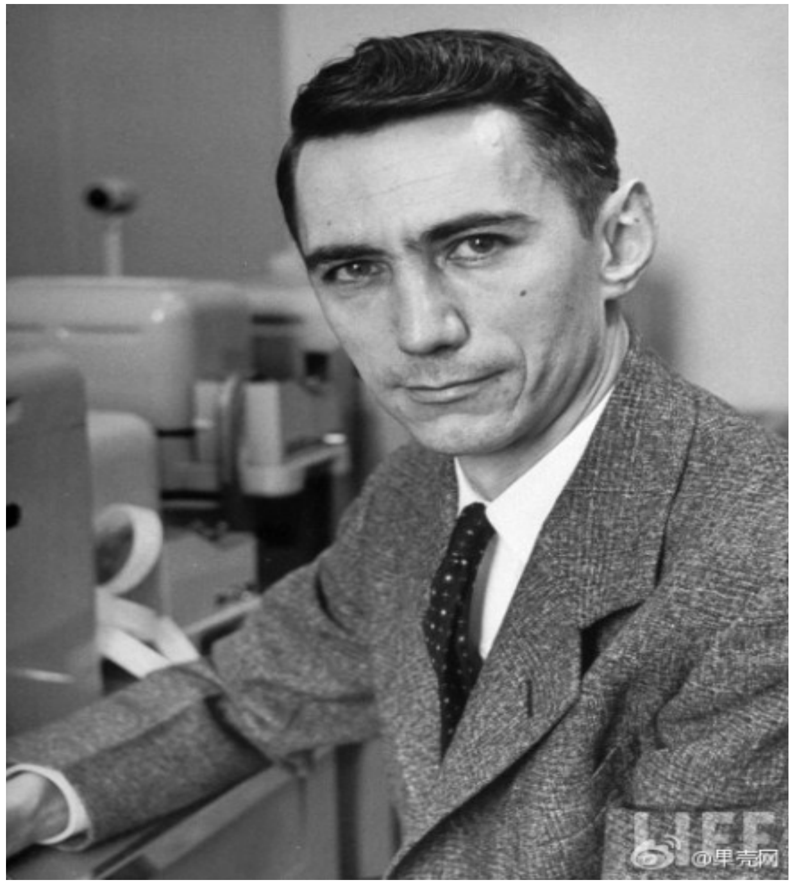

Homework 02 Information Measure

Ge Yao, alexgecontrol@qq.com

Contents

Ge Yao
01-25-2015

Homework 02
Information Measure

Problem 11

Prove that $H(p)$ is concave is p, i.e. , for $0 \le \lambda \le 1$ and $\bar{\lambda} = 1 - \lambda$

$$\lambda H(p_1) + \bar{\lambda} H(p_2) \le H\left(\lambda p_1 + \bar{\lambda} p_2\right)$$

Proof

First, we prove the following lemma that **relative entropy $\boldsymbol{D(p||q)}$ is a convex functional**, i.e., for $\forall (p_1, q_1), (p_2, q_2)$, $0 \le \lambda \le 1$

$$\lambda D(p_1||q_1) + (1-\lambda) D(p_2||q_2) \ge D(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2)$$

$$\lambda D(p_1||q_1) + (1-\lambda) D(p_2||q_2) = \sum \lambda\, p_1 log \frac{p_1}{q_1} + \sum (1-\lambda)\, p_2 log \frac{p_2}{q_2}$$

$$= \sum \left[\lambda p_1 log \frac{\lambda p_1}{\lambda q_1} + (1-\lambda) p_2 log \frac{(1-\lambda)p_2}{(1-\lambda)q_2}\right]$$

Apply the log-sum inequality to the summation kernel, we have:

$$\ge \sum [\lambda p_1 + (1-\lambda)p_2] log \frac{\lambda p_1 + (1-\lambda)p_2}{\lambda q_1 + (1-\lambda)q_2}$$

$$= D(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2)$$

So the convexity of relative entropy is proved.

No we apply the above lemma to the two pairs of distributions $(p_1, u), (p_2, u)$ in which u is the uniform distribution on the given alphabet, we will have:

$$\lambda \sum p_1 log p_1 + (1-\lambda) \sum p_2 log p_2 \geq \sum [\lambda p_1 + (1-\lambda) p_2] log [\lambda p_1 + (1-\lambda) p_2]$$

Which is equivalent to:

$$-\lambda H(p_1) - (1-\lambda) H(p_2) \geq -H(\lambda p_1 + (1-\lambda) p_2)$$

i.e.

$$\lambda H(p_1) + (1-\lambda) H(p_2) \leq H(\lambda p_1 + (1-\lambda) p_2)$$

So the concavity of entropy is proved.

Problem 12

Let $(X,Y) \sim p(x,y) = p(x)p(y|x)$

a) Prove that for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$;

b) Prove that for fixed $p(y|x)$, $I(X;Y)$ is a concave functional of $p(x)$.

Proof a

Define the joint distributions as

$$p_1(x,y) = p(x)p_1(y|x)$$

$$p_2(x,y) = p(x)p_2(y|x)$$

$$p_\lambda(x,y) = p(x)[\lambda p_1(y|x) + (1-\lambda)p_2(y|x)]$$

Define the marginal distributions of Y as

$$p_1(y) = \sum_x p_1(x,y)$$

$$p_2(y) = \sum_x p_2(x,y)$$

$$p_\lambda(y) = \sum_x p_\lambda(x,y) = \lambda p_1(y) + (1-\lambda)p_2(y)$$

Define the mutual information as

$$I_1(X;Y) = \sum p_1(x,y) log \frac{p_1(x,y)}{p(x)p_1(y)}$$

$$I_2(X;Y) = \sum p_1(x,y) log \frac{p_1(x,y)}{p(x)p_1(y)}$$

$$I_\lambda(X;Y) = \sum p_\lambda(x,y) log \frac{p_\lambda(x,y)}{p(x)p_\lambda(y)}$$

Since

$\lambda I_1(X;Y) + (1-\lambda)I_2(X;Y) =$

$$= \sum \lambda p_1(x,y) log \frac{p_1(x,y)}{p(x)p_1(y)} + \sum(1-\lambda)\, p_2(x,y) log \frac{p_2(x,y)}{p(x)p_2(y)}$$

$$= \sum \lambda p_1(x,y) log \frac{\lambda p_1(x,y)}{p(x)\lambda p_1(y)} + (1-\lambda)p_2(x,y) log \frac{(1-\lambda)p_2(x,y)}{p(x)(1-\lambda)p_2(y)}$$

Apply log-sum inequality to the summation kernel, we have:

$\lambda I_1(X;Y) + (1-\lambda)I_2(X;Y) \geq$

$$\geq \sum[\lambda p_1(x,y) + (1-\lambda)p_2(x,y)] log \frac{\lambda p_1(x,y)+(1-\lambda)p_2(x,y)}{p(x)[\lambda p_1(y)+(1-\lambda)p_2(y)]}$$

$$= \sum p_\lambda(x,y) log \frac{p_\lambda(x,y)}{p(x)p_\lambda(y)} = I_\lambda(X;Y)$$

So for fixed $p(x)$, $I(X;Y)$ is a convex functional of $p(y|x)$

Proof b

Define the joint distributions as

$$p_1(x,y) = p_1(x)p(y|x)$$

$$p_2(x,y) = p_2(x)p(y|x)$$

$$p_\lambda(x,y) = p(y|x)[\lambda p_1(x) + (1-\lambda)p_2(x)]$$

Define the marginal distributions of Y as

$$p_1(y) = \sum_x p_1(x,y)$$

$$p_2(y) = \sum_x p_2(x, y)$$

$$p_\lambda(y) = \sum_x p_\lambda(x, y) = \lambda p_1(y) + (1 - \lambda) p_2(y)$$

Define the mutual information as

$$I_1(X; Y) = H_1(Y) - H_1(Y|X)$$

$$I_2(X; Y) = H_2(Y) - H_1(Y|X)$$

$$I_\lambda(X; Y) = H_\lambda(Y) - H_\lambda(Y|X)$$

In which

$$H_t(Y) = \sum_y -log p_t(y)\, p_t(y), t = 1,2, \lambda$$

$$H_t(Y|X) = \sum_y -log \frac{p(y|x) p_t(x)}{p_t(x)} p_t(x, y), t = 1,2, \lambda$$

Define u as the uniform distribution on the given alphabet:

$$u(x_i) = \frac{1}{|\chi|}, x_i \in \chi$$

Since

$$\lambda H_1(Y) + (1 - \lambda) H_1(Y) - log|\chi|$$

$$= \sum_y \sum_x -\left[log \frac{\lambda p_1(x) p(y|x)}{\lambda u} \lambda p_1(x) p(y|x) + log \frac{(1 - \lambda) p_2(x) p(y|x)}{(1 - \lambda) u} (1 - \lambda) p_2(x) p(y|x)\right]$$

$$\leq \sum_y \sum_x -\left\{ log \frac{[\lambda p_1(x) + (1-\lambda)p_2(x)]p(y|x)}{u} [\lambda p_1(x) + (1-\lambda)p_2(x)]p(y|x) \right\}$$

$$= H_\lambda(Y) - log|\chi|$$

So $H(Y)$ is a concave functional of $p(x)$ given $p(y|x)$.

Since

$$\lambda H_1(Y|X) + (1-\lambda)H_1(Y|X)$$

$$= \sum_y \sum_x -\left[log \frac{\lambda p_1(x)p(y|x)}{\lambda p_1(x)} \lambda p_1(x)p(y|x) + log \frac{(1-\lambda)p_2(x)p(y|x)}{(1-\lambda)p_2(x)} (1-\lambda)p_2(x)p(y|x) \right]$$

$$\leq \sum_y \sum_x -\left\{ log \frac{[\lambda p_1(x) + (1-\lambda)p_2(x)]p(y|x)}{\lambda p_1(x) + (1-\lambda)p_2(x)} [\lambda p_1(x) + (1-\lambda)p_2(x)]p(y|x) \right\}$$

$$= H_\lambda(Y|X)$$

So $H(Y|X)$ is a concave functional of $p(x)$ given $p(y|x)$.

Since the linear combination of concave functional is still a concave functional, $I(X;Y) = H(Y) - H(Y|X)$ is a concave functional of $p(x)$ given $p(y|x)$.

Problem 15

Let X be a function of Y. Prove that $H(X) \leq H(Y)$. Interpret this result.

Proof

Since

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

When X is a function of Y $H(X|Y) = 0$, so the above expression will be:

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

Since $H(Y|X) \geq 0$, we will finally have:

$$H(X) \leq H(Y)$$

This inequality says that **we cannot extract more information from the original data using whatever sort of processing method**. The inequality also says that **any processing of the original data could only be achieved at the potential cost of losing some original information.**

Problem 16

Prove that for any $n \geq 2$, $H(X_1, \ldots, X_n) \geq \sum_{i=1}^{n} H(X_i|X_j, j \neq i)$

Proof

Since

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{1:i-1})$$

And conditioning will reduce entropy

$$H(X_1) \geq H(X_1|X_2) \geq \cdots \geq H(X_1|X_j), j \neq 1$$

$$\ldots$$

$$H(X_n) \geq H(X_n|X_1) \geq \cdots \geq H(X_n|X_j), j \neq n$$

Because

$$nH(X_1, \ldots, \quad X_n)$$

$$= \sum \left[H(X_1) + \cdots + H(X_1|X_j), j \neq 1\right] + \cdots + \left[H(X_1) + \cdots + H(X_1|X_j), j \neq n\right]$$

$$\geq n \sum_{i=1}^{n} H(X_i|X_{1:i-1})$$

We finally have

$$H(X_1, \ldots, X_n) \geq \sum_{i=1}^{n} H(X_i|X_{1:i-1})$$

Problem 17

Prove that $H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3) \geq 2H(X_1, X_2, X_3)$

Proof

Since

$$H(X_1, X_2, X_3) = H(X_1|X_2, X_3) + H(X_2, X_3)$$

$$H(X_1, X_2, X_3) = H(X_2|X_1, X_3) + H(X_1, X_3)$$

$$H(X_1, X_2, X_3) = H(X_3|X_1, X_2) + H(X_1, X_2)$$

Sum the above three equations up, we have:

$$H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3) + \sum_{i=1}^{3} H\left(X_I|X_J\right), j \neq i = 3H(X_1, X_2, X_3)$$

Since

$$\sum_{i=1}^{3} H\left(X_I|X_J\right), j \neq i \leq H(X_1, X_2, X_3)$$

We finally have:

$$H(X_1, X_2) + H(X_2, X_3) + H(X_1, X_3)$$

$$= 3H(X_1, X_2, X_3) - \sum_{i=1}^{3} H\left(X_I|X_J\right), j \neq i$$

$$\geq 2H(X_1, X_2, X_3)$$

References