20200929 Homework 石望华 120033910146

1. Prove that $Z=\max\{H(X|Y), H(Y|X)\}$ is a distance of discrete random variables X and Y.

① $Z(X,Y)=\max\{H(X|Y),H(Y|X)\}=\max\{H(X),H(Y)\}-I(X;Y)=Z(Y;X)$

② since $H(X)\geq 0$, so $Z(X,Y)\geq 0$

③ $Z(X,Y)=0$ ~~only~~ if and only if $H(X|Y)=H(Y|X)=0$

$\therefore H(X)-I(X;Y)=H(Y)-I(X;Y) \Leftrightarrow H(X)=H(Y)$

$\therefore X=Y$ or there is a one-to-one function mapping from X to Y.

④ $Z(X,Y)+Z(Y,T)=\max\{H(X),H(Y)\}+\max\{H(Y),H(T)\}-I(X;Y)-I(Y;T)$

$Z(X,T)=\max\{H(X),H(T)\}-I(X;T)$

(1) if $H(X)\geq H(Y)\geq H(T)$: $Z(X,Y)+Z(Y,T)=H(X|Y)+H(Y|T)\geq H(X|Y,T)+I(X;Y|T)=H(X|T)$
$=Z(X,T)$.

(2) if $H(X)\geq H(T)\geq H(Y)$: $Z(X,Y)+Z(Y,T)=H(X|Y)+H(T|Y)\geq H(X|Y,T)+H(Y|T)\geq H(X|Y,T)+I(X;Y|T)$
$=H(X|T)=Z(X,T)$.

(3) if $H(T)\geq H(X)\geq H(Y)$: $Z(X,Y)+Z(Y,T)=H(X|Y)+H(T|Y)\cancel{\geq H(X|Y,T)+H(Y|T)\geq H(X|Y,T)+I(X;Y|T)=Z(X,T)}$
$\geq H(Y|X)+H(T|X,Y)\geq I(T;Y|X)+H(T|X,Y)=H(T|X)=Z(X,T)$.

The other cases where $H(X)\leq H(Y)$ are similar.

$\therefore Z=\max\{H(X|Y),H(Y|X)\}$ is a metric for all X, Y.

2. Prove chain rules.

① $I(X_1,X_2,\cdots,X_n;Y)=H(X_1,\cdots,X_n)-H(X_1,X_2,\cdots,X_n|Y)$

$=\sum_{i=1}^{n}H(X_i|X_1,\cdots,X_{i-1})-\sum_{i=1}^{n}H(X_i|X_1,X_2,\cdots,X_{i-1},Y)$

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

② $I(X_1,X_2,\cdots,X_n;Y|Z)=H(X_1,X_2,\cdots,X_n|Z)-H(X_1,X_2,\cdots,X_n|Y,Z)$

$=\sum_{i=1}^{n}H(X_i|X_1,\cdots,X_{i-1},Z)-\sum_{i=1}^{n}H(X_i|X_1,X_2,\cdots,X_{i-1},Y,Z)$

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

3. $a_i>0, b_i>0$, prove Log-Sum inequality.

Let $a_i'=\frac{a_i}{\sum_{j=1}^{n}a_j}$, $b_i'=\frac{b_i}{\sum_{j=1}^{n}b_j}$. thus $0\leq D(a'||b')=\sum_{i=1}^{n}a_i'\log\frac{a_i'}{b_i'}=\sum_{i=1}^{n}\frac{a_i}{\sum_j a_j}\log\frac{a_i}{b_i}\cdot\frac{\sum_{j=1}^{n}b_j}{\sum_{j=1}^{n}a_j}$

$=\frac{1}{\sum_{j=1}^{n}a_j}\left[\sum_{i=1}^{n}a_i\log\frac{a_i}{b_i}-\sum_{i=1}^{n}a_i\cdot\log\frac{\sum_{j=1}^{n}a_j}{\sum_{j=1}^{n}b_j}\right]$

$\therefore \sum_{i=1}^{n}a_i\log\frac{a_i}{b_i}\geq\left(\sum_{i=1}^{n}a_i\right)\log\frac{\sum_{i=1}^{n}a_i}{\sum_{i=1}^{n}b_i}$

4. prove $D(p||q)$ is convex function of the pair (p,q).

Let $a_1=\lambda p_1$, $a_2=(1-\lambda)p_2$, $b_1=\lambda q_1$, $b_2=(1-\lambda)q_2$.

then $D(a_1+a_2||b_1+b_2)=\sum_x(a_1+a_2)\log\frac{a_1+a_2}{b_1+b_2}\leq\sum_x a_1\log\frac{a_1}{b_1}+\sum_x a_2\log\frac{a_2}{b_2}=\lambda\sum_x p_1\log\frac{p_1}{q_1}+(1-\lambda)\sum_x p_2\log\frac{p_2}{q_2}$
$=\lambda D(p_1||q_1)+(1-\lambda)D(p_2||q_2)$

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

5. prove. $H(p)$ is a concave function of p.

$$H(\lambda p_1+(1-\lambda)p_2)=\lambda\sum_{i=1}^{n}p_{1i}\log\frac{1}{\lambda p_{1i}+(1-\lambda)p_{2i}}+(1-\lambda)\sum_{i=1}^{n}p_{2i}\log\frac{1}{\lambda p_{1i}+(1-\lambda)p_{2i}}$$

let $q_i=\lambda p_{1i}+(1-\lambda)p_{2i}$. $\sum_{i=1}^{n}p_{1i}=1$, $\sum_{i=1}^{n}q_i\le 1$

$$\therefore H(\lambda p_1+(1-\lambda)p_2)=\lambda\sum_{i=1}^{n}p_{1i}\log\frac{1}{q_i}+(1-\lambda)\sum_{i=1}^{n}p_{2i}\log\frac{1}{q_i}$$

$\because \ln x\le x-1\quad(x>0)$

$\therefore \ln\frac{q_i}{p_{1i}}\le\frac{q_i}{p_{1i}}-1 \Leftrightarrow p_{1i}\ln\frac{1}{p_{1i}}\le p_{1i}\ln\frac{1}{q_i}+q_i-p_{1i}$, summing on i gives:

$$\sum_{i=1}^{n}p_{1i}\ln\frac{1}{p_{1i}}\le\sum_{i=1}^{n}p_{1i}\ln\frac{1}{q_i}+\sum_{i=1}^{n}q_i-\sum_{i=1}^{n}p_{1i}\le\sum_{i=1}^{n}p_{1i}\ln\frac{1}{q_i}$$

$$\therefore H(\lambda p_1+(1-\lambda)p_2)\ge\lambda\sum_{i=1}^{n}p_{1i}\log\frac{1}{p_{1i}}+(1-\lambda)\sum_{i=1}^{n}p_{2i}\log\frac{1}{p_{2i}}=\lambda H(p_1)+(1-\lambda)H(p_2)$$

6. Prove the Fano's inequality.

def: $P_e=P_r\{X\neq\hat{X}\}$, $Y=\begin{cases}0, & \text{if } X=\hat{X}\\ 1, & \text{if } X\neq\hat{X}\end{cases}$. $\quad P(Y=1)=P_e$, $H(Y|X,\hat{X})=0$, $H(X|\hat{X}=\hat{x},Y=0)=1$

$$H(X|\hat{X})=H(X|\hat{X})+H(Y|X,\hat{X})=H(X,Y|\hat{X})=H(Y|\hat{X})+H(X|Y,\hat{X})\le H(Y)+\sum_{\hat{x}\in\mathcal{X}}[P(Y=0,\hat{X}=\hat{x})\cdot$$
$$H(X|\hat{X}=\hat{x},Y=0)+P(Y=1,\hat{X}=\hat{x})H(X|\hat{X}=\hat{x},Y=1)]=H(Y)+\sum_{\hat{x}\in\mathcal{X}}(0+P(Y=1,\hat{X}=\hat{x})H(X|\hat{X}=\hat{x},Y=1))$$
$$\le H(Y)+P(Y=1)\log(|\mathcal{X}|-1)=h_b(P_e)+P_e\log(|\mathcal{X}|-1)$$

7. If $X_1\to\cdots\to X_6$ is a Markov chain, judge and show whether $(X_3,X_6)\to(X_2,X_5)\to(X_1,X_4)$ is also a Markov chain.

Only need to prove: $p(x_1,x_2,\cdots,x_6)p(x_2,x_5)=p(x_2,x_3,x_5,x_6)p(x_1,x_2,x_4,x_5)$ ------ ①

$X_1\to X_2\to\cdots\to X_6\Rightarrow p(x_1,x_2,\cdots,x_6)p(x_2)p(x_3)p(x_4)p(x_5)=p(x_1,x_2)p(x_2,x_3)p(x_3,x_4)p(x_4,x_5)p(x_5,x_6)$ -- ②

According to the proposition of Markov subchains, $X_2\to X_3\to X_5\to X_6$ and $X_1\to X_2\to X_4\to X_5$ form a Markov chain, thus:

$p(x_2,x_3,x_5,x_6)\,p(x_3)p(x_5)=p(x_2,x_3)\,p(x_3,x_5)\,p(x_5,x_6)$ ------------ ③

$p(x_1,x_2,x_4,x_5)\cdot p(x_2)\,p(x_4)=p(x_1,x_2)\,p(x_2,x_4)\,p(x_4,x_5)$ ------------ ④

Based on ①-④, we only need to prove ⑤:

$p(x_3,x_5)\cdot p(x_2,x_4)=p(x_2,x_5)\,p(x_3,x_4)$ ------------ ⑥

which is equivalent to prove ⑦:

$p(x_5|x_3)\,p(x_4|x_2)=p(x_4|x_2)\,p(x_5|x_2)$ when $p(x_2)>0$, $p(x_3)>0$ ---- ⑦

1. Distance of discrete random variables.

Prove that $\max\{H(X|Y), H(Y|X)\}$ is a distances of discrete random variables X and Y.

2. Prove the following two chain rules.

- **Proposition 2.26 (Chain Rule for Mutual Information).**

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

- **Proposition 2.27 (Chain Rule for Conditional Mutual Information)**

For random variables $X_1, X_2, \ldots, X_n, Y$ *and* Z ,

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

3. Log-Sum inequality: For non-negative numbers a_i and b_i, prove

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}.$$

4. Convex Relative Entropy

If (p_1, q_1) and (p_2, q_2) are pairs of probability mass functions then

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

for all $0 \leq \lambda \leq 1$. That is, $D(p||q)$ is **convex function** of the pair (p, q).

5. Concave Entropy

Let p be the probability mass function of discrete random variable X. Here $H(X)$ is denoted by $H(p)$. Prove that

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

That is, $H(p)$ is a **concave function** of p.

6. Prove the Fano's inequality.

7. If $X_1 \to X_2 \to X_3 \to X_4 \to X_5 \to X_6$ is a Markov chain, **judge and show** whether $(X_3, X_6) \to (X_2, X_5) \to (X_1, X_4)$ is also a Markov chain.

8. Remark. Convex and Concave mutual information

Mutual Information $I(X;Y)$ can be expressed by a function of input distribution $p(x)$ and transition distribution $p(y|x)$, i.e.,

$$I(X;Y) = f(p(x), p(y|x)).$$

A. For given input distribution $p(x)$, we say that $I(X;Y)$ is **convex** of transition distribution $p(y|x)$.

B. For given transition distribution$p(y|x)$, we say that $I(X;Y)$ is **concave** of input distribution $p(x)$.

Convex 8A can be explained by:

Let $p_1(x,y) = p(x)p_1(y|x)$ and $p_2(x,y) = p(x)p_2(y|x)$ be two joint distributions, we define

- $p(x,y) \triangleq \lambda p_1(x,y) + (1-\lambda)p_2(x,y) = \lambda p(x)p_1(y|x) + (1-\lambda)p(x)p_2(y|x) = p(x)\{\boldsymbol{\lambda p_1(y|x) + (1-\lambda)p_2(y|x)}\}$
- $q(x,y) \triangleq \lambda q_1(x,y) + (1-\lambda)q_2(x,y) \triangleq \lambda p(x)p_1(y) + (1-\lambda)p(x)p_2(y) = p(x)\{\boldsymbol{\lambda p_1(y) + (1-\lambda)p_2(y)}\}$
- $I_\lambda(X;Y) \triangleq D(p(x,y)||q(x,y))$
- $I_1(X;Y) \triangleq D(p_1(x,y)||q_1(x,y)) = D(p_1(x,y)||p(x)p_1(y))$
- $I_2(X;Y) \triangleq D(p_2(x,y)||q_2(x,y)) = D(p_2(x,y)||p(x)p_2(y))$

Then, for given input distribution $p(x)$, $I(X;Y)$ is convex of transition distribution $p(y|x)$, i.e.,

$$I_\lambda(X;Y) \leq \lambda I_1(X;Y) + (1-\lambda) I_2(X;Y) \text{ i.e.}$$

$$D(p(x,y)||q(x,y)) \leq \lambda D(p_1(x,y)||q_1(x,y)) + (1-\lambda) D(p_2(x,y)||q_2(x,y))$$