Lec 5: Consequences of log sum inequality.

Recap

- Consequences of Jensen's inequality
- log sum inequality (LSI)
  - used to show $D(p\|q)$ is convex in the pair (p, q)

$\sum_{i=1}^{k} a_i \log \frac{a_i}{b_i} \geq (\sum a_i) \log \frac{\sum a_i}{\sum b_i}$

$$D(\lambda p_1 + (1-\lambda) p_2 \,\|\, q_1 \lambda + (1-\lambda) q_2)$$
$$= \sum_x [\underbrace{p_1(x)\lambda}_{a_1} + \underbrace{(1-\lambda) p_2(x)}_{a_2}] \log \frac{\overbrace{p_1(x)\lambda}^{a_1} + \overbrace{(1-\lambda) p_2(x)}^{a_2}}{\underbrace{q_1(x)\lambda}_{b_1} + \underbrace{(1-\lambda) q_2(x)}_{b_2}}$$

applying LSI to the inside term for each x

$$\leq \sum_x \lambda p_1(x) \log \frac{p_1(x)}{q_1(x)} + (1-\lambda) \sum_x p_2(x) \log \frac{p_2(x)}{q_2(x)}$$
$$= \lambda D(p_1 \| q_1) + (1-\lambda) D(p_2 \| q_2).$$

Note: Here we treat $D(p(x) \| q(x))$ as a function of vector

$$\begin{pmatrix} p(x_1) \\ p(x_{|\mathcal{X}|}) \\ p(y_1) \\ p(y_{|\mathcal{X}|}) \end{pmatrix}$$ of size $(2|\mathcal{X}| \times 1)$

Today

1) $H(X)$ convex in $p(x)$
2) 2nd proof
3) $I(X;Y)$
   - concave in $p(x)$
   - convex in $p(y|x)$
4) Data processing inequality.

Theorem $H(X)$ is a concave function of $p(x)$

Pf: $D(p\|u) = \sum_x p(x) \log \frac{p(x)}{u(x)} = \log|\mathcal{X}| - H(X)$

$$\Rightarrow H(X) = -D(p\|u) + \log|\mathcal{X}| \qquad H(X) \triangleq H(p)$$

$$\therefore H(\lambda p_1 + (1-\lambda) p_2) = \log|\mathcal{X}| - D(\lambda p_1 + (1-\lambda) p_2 \,\|\, u)$$

But $u = \lambda u + (1-\lambda) u$

by convexity of $D(p\|q)$

$$\geq \log|\mathcal{X}| - \lambda D(p_1\|u) - (1-\lambda) D(p_2\|u)$$
$$= \lambda H(p_1) + (1-\lambda) H(p_2)$$

Alternate proof:

let $\theta = \begin{cases} 1 & \text{prob} = \lambda \\ 2 & \text{prob} = 1-\lambda \end{cases}$

X_1 has prob $p_1(x)$

X_2 has " $p_2(x)$.

Consider random variable X_θ. It has prob distribution
$\lambda p_1(x) + (1-\lambda) p_2(x)$.

$$H(X_\theta) \geq H(X_\theta | \theta)$$
$$= \lambda H(X_1) + (1-\lambda) H(X_2)$$
$$= \lambda H(p_1) + H(p_2)(1-\lambda)$$

<u>thm</u> Let $(X, Y) \sim p(x)\, p(y|x)$. the mutual information is concave in $p(x)$ for fixed $p(y|x)$ and convex in $p(y|x)$ for fixed $p(x)$.

<u>Pf</u>: Let $p(y|x)$ be fixed. To show that $I(X;Y)$ is concave in $p(x)$.

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

Note $p(y) = \sum_x p(x) p(y|x)$

$$H(Y|X) = \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)}$$
$$= \underline{b}^t \underline{p}_x$$

$\underline{p}_y = A \underline{p}_x$ where $A \Leftrightarrow p(y|x)$ fixed

$\therefore$ Can write
$$I(X;Y) = H(A\underline{p}_x) - \underline{b}^t \underline{p}_x$$
Suppose $\underline{p}_x = \lambda \underline{p}_{x,1} + (1-\lambda) \underline{p}_{x,2}$

$$I(X;Y) = H(\lambda A \underline{p}_{x,1} + (1-\lambda) A \underline{p}_{x,2}) - b^t \lambda p_{x_1} - b^t (1-\lambda) p_{x_2}$$
$$\geq \lambda H(A p_{x_1}) + (1-\lambda) H(A p_{x_2}) - \lambda b^t p_{x_1} - (1-\lambda) b^t p_{x_2}$$
$$= \lambda H(Y_1) + (1-\lambda) H(Y_2) - \lambda H(Y_1|X) - (1-\lambda) H(Y_2|X)$$
$$= \lambda I(X_1;Y_1) + (1-\lambda) I(X_2;Y_2)$$

here X_1 is a r.v with dist $\underline{p}_{x,1}$
& X_2 " $p_{x,2}$.

$P(Y_1 = y) = \sum_x p_{x_1}(x) p(y|x)$
$P(Y_2 = y) = \sum_x p_{x_2}(x) p(y|x)$

Next let $p(x)$ be fixed, Suppose
$$p(y|x) = \lambda p_1(y|x) + (1-\lambda) p_2(y|x)$$
Set
$$p_1(x,y) = p(x) p_1(y|x)$$
$$p_2(x,y) = p(x) p_2(y|x)$$
$$p_1(y) = \sum_x p(x) p_1(y|x)$$
$$p_2(y) = \sum_x p(x) p_2(y|x)$$

Then $\left[p(x,y) = p(x)\, p(y|x) \right]$, $p(y) = \sum_x p(x)\, p(y|x)$

$= \lambda p_1(y) + (1-\lambda) p_2(y)$

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

$$= D\big(\lambda p_1(x,y) + (1-\lambda) p_2(x,y) \,\|\, p(x)\,\lambda p_1(y) + p(x)\, p_2(y)(1-\lambda) \big)$$

by convexity of $D(p\|q)$

$$\le \lambda D\big(p_1(x,y) \,\|\, p(x) p_1(y) \big) + (1-\lambda) D\big(p_2(x,y) \,\|\, p(x) p_2(y) \big)$$

$$= \lambda\, I(X;Y)\Big|_{p(y|x) = p_1(y|x)} + (1-\lambda)\, I(X;Y)\Big|_{p(y|x) = p_2(y|x)}$$

Data processing inequality

<u>Defn</u>: X, Y, Z (Rvs) are said to form a Markov chain (written $X \to Y \to Z$) if $p(x,y,z) = p(x)\, p(y|x)\, p(z|y)$

<u>Claim</u>: $X \to Y \to Z$ iff X and Z are independant given Y.

$(\Leftarrow)$

$$p(x,z|y) = p(x|y)\, p(z|y).$$

$$\Rightarrow p(x,z,y) = p(x,z|y)\, p(y)$$

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

$$= p(x)\, p(y|x)\, p(z|y) \quad \therefore X \to Y \to Z$$

$(\Rightarrow)$

$$p(x,z|y) = \frac{p(x,z,y)}{p(y)} = \frac{p(x)\, p(y|x)\, p(z|y)}{p(y)}$$

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

$\therefore X, Z$ independant given Y.

<u>Corollary</u>: $X \to Y \to Z$ iff $Z \to Y \to X$.

Hence this is written sometimes as $X \leftrightarrow Y \leftrightarrow Z$

Note that: $X \to Y \to f(Y)$ for a deterministic function $f(\cdot)$

$$p(f(y)|y,x) = p(f(y)|y).$$

<u>Thm (DPI)</u> If $X \to Y \to Z$ then $I(X;Y) \ge I(X;Z)$

Consider $I(X;YZ) = I(X;Y) + \underbrace{I(X;Z|Y)}_{\to 0}$ through independance

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

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

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

In particular $I(X;Y) \geq I(X; f(Y)) \Rightarrow$ be careful about throwing away raw data