IT in CS - LECTURE 3

2/5/2019

TODAY: BASICS OF I.T. (contd.)

- Conditional Entropy, KL Divergence
- (In)equalities

* * *

Review of Entropy

- Suppose transmitting n i.i.d copies $X_1 \ldots X_n$, $X_i \sim P_X$ & $P_X = (p_1, \ldots p_m)$ $\left[\sum_{i=1}^{m} p_i = 1, \quad p_i \geq 0 \right]$
- "Budget" $\ell_i^* = \log \frac{1}{p_i}$ bits to transmit i
- Total Transmission cost will be $\sum_{i=1}^{m} p_i \ell_i^* = H(X)$!

- Could try to use any other $\{\ell_i\}$ $i=1 \ldots m$. Why $\ell_1^* \cdots \ell_m^*$?

- ① Not all ℓ_i achievable. E.g. $\ell_1 = \ell_2 = \ell_3 \cdots = \ell_m = 1$!

  Pset 1, Problem 4: ℓ_i achievable in prefix-free way $\Rightarrow$ $\sum 2^{-\ell_i} \leq 1$.

  So if we let $q_i = 2^{-\ell_i}$ then we can try to pretend $X_j \sim Q = (q_1 \ldots q_m)$ "sub-distribution".

  & transmit according to Q

But presumably $\sum p_i l_i = \underset{\substack{\uparrow \\ \text{sub-optimal}}}{\sum p_i \log \frac{1}{q_i}} \geq \underset{\substack{\uparrow \\ \text{Optimal}}}{\sum p_i \log \frac{1}{p_i}}$

Is the above really the case? Will see ....

Back to Basics

- Notation : $P_{XY}(\alpha,\beta) \triangleq \Pr\left[X=\alpha, Y=\beta\right]$

  "Distribution" : $\sum_{\alpha,\beta} P_{XY}(\alpha,\beta) = 1$, $P_{XY}(\alpha,\beta) \geq 0$.

- P_X , P_Y = marginals $P_X(\alpha) = \sum_{\beta} P_{XY}(\alpha,\beta)$.
- $P_{Y|X=\alpha}$ = Conditional Dist. $P_{Y|_{X=\alpha}}(\beta) = P_{XY}(\alpha,\beta) / P_X(\alpha)$
- Conditional Entropy = Expected Entropy after conditioning

$$H(Y|X) = \mathop{\mathbb{E}}_{\alpha}\left[H(Y|_{X=\alpha}) \right]$$

$$= \sum_{\alpha} P_X(\alpha) \cdot H(Y|_{X=\alpha})$$

$$= \sum_{\alpha} P_X(\alpha) \cdot \sum_{\beta} P_{Y|_{X=\alpha}}(\beta) \cdot \log \frac{1}{P_{Y|_{X=\alpha}}(\beta)}$$

$$= \sum_{\alpha,\beta} P_{XY}(\alpha,\beta) \cdot \log \frac{P_X(\alpha)}{P_{XY}(\alpha,\beta)}$$

AXIOMS OF ENTROPY

① $H(X) \leq \log|\Omega|$ if $X \in \Omega$

Equality iff $X \sim \mathrm{Unif}(\Omega)$

② $H(X,Y) = H(X) + H(Y|X)$

③ $H(Y|X) \leq H(Y)$ "Conditioning reduces entropy"

* * *

Will prove above today

* * *

②: $$H(X,Y) = \sum_{\alpha,\beta} P_{XY}(\alpha,\beta) \log \frac{1}{P_{XY}(\alpha,\beta)}$$

$$H(X) = \sum_{\alpha} P_X(\alpha) \log \frac{1}{P_X(\alpha)} = \sum_{\alpha,\beta} P_{XY}(\alpha,\beta) \log \frac{1}{P_X(\alpha)}$$

$$H(Y|X) = \sum_{\alpha,\beta} P_{X,Y}(\alpha,\beta) \log \frac{P_X(\alpha)}{P_{XY}(\alpha,\beta)}$$

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

$$\Leftarrow \log \frac{1}{P_{XY}(\alpha,\beta)} = \log \frac{1}{P_X(\alpha)} + \log \frac{P_X(\alpha)}{P_{XY}(\alpha,\beta)}$$

Corollary: $H(X_1 \ldots X_n)$ $\quad X_i$ i.i.d $\sim X$

$= n \cdot H(X)$.

The inequalities:

① $H(X) \leq \log|\Omega|$

② $H(X) < \log|\Omega|$ if $P_X \neq \text{Unif}$

③ $H(Y|X) \leq H(X)$

All follow from the "optimality" of entropy.

Theorem:

$\forall$ pair of distributions P, Q

$$\mathbb{E}_{X\sim P}\left[\log\frac{1}{P(X)}\right] \leq \mathbb{E}_{X\sim P}\left[\log\frac{1}{Q(X)}\right]$$

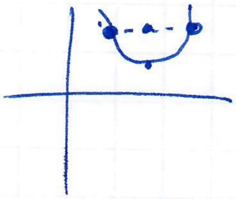
with equality iff $P = Q$.

$\uparrow$ compressing according to "correct" dist.

$\uparrow$ compressing according to "wrong" distribution

Key Ingredient in proof: JENSEN's INEQUALITY

if f is convex then



$$\mathbb{E}_{\substack{\cancel{X}\\ Z}}\left[f(\overset{\cancel{X}}{Z})\right] \geqslant f\left(\mathbb{E}_{Z}[\overset{\cancel{X}}{Z}]\right)$$

if f is concave then $\mathbb{E}_{Z}[f(Z)] \leq f(\mathbb{E}_{Z}[Z])$

- will apply to $f(z) = \log z$ (concave)

  & $z = \dfrac{Q(x)}{P(x)}$ $x \sim P$

- Get:

$$\mathbb{E}_{x \sim P}\left[\log \frac{Q(x)}{P(x)}\right] \leq \log\left(\mathbb{E}_{x \sim P}\left[\frac{Q(x)}{P(x)}\right]\right)$$

$$= \log\left(\sum_{\alpha} \frac{P(\alpha) \cdot Q(\alpha)}{P(\alpha)}\right)$$

$$= \log \sum_{\alpha} Q(\alpha)$$

$$= \log 1$$

$$= 0$$

Conclude:

$$\mathbb{E}_{x \sim P}\left[\log \frac{1}{P(x)}\right] \leq \mathbb{E}_{x \sim P}\left[\log \frac{1}{Q(x)}\right]$$

$\Uparrow$ compressing according to right dist.

$\Uparrow$ Compressing according to wrong dist.

Equality iff Equality in Jensen iff $z =$ constant (for strictly concave f)

iff $P(x) = Q(x)$ $\forall x$.

$\boxtimes$

- **KL Divergence**

$$D(P \| Q) \triangleq \underset{X\sim P}{\mathbb{E}}\left[\log \frac{P(X)}{Q(X)}\right]$$

- Theorem just proved says. $D(P\|Q) \geq 0$ & $D(P\|Q) > 0$ iff $P \neq Q$.

- Divergence
  - $\nearrow$ nice : $D(P^n \| Q^n) = n \cdot D(P\|Q)$.
  - $\searrow$ not so nice : Not a metric
    - $D(P\|Q) \neq D(Q\|P)$
    - $D(P\|Q)$ not necessarily finite

- Operational Meaning : loss per copy when compressing $X \sim P$ using mechanism for Q.

———— x ————

Using Divergence Theorem :

① $H(X) \leq \log|\Omega| = \underset{X\sim P}{\mathbb{E}}\left[\log|\Omega|\right]$

$$= \underset{X\sim P}{\mathbb{E}}\left[\log \frac{1}{Q(X)}\right] \qquad Q(x) = \mathrm{Unif}(\Omega)$$

② $H(Y|X) \leq H(Y) \Leftarrow D\left(P_{XY} \| P_X \times P_Y\right) \geq 0$

$$\Uparrow ③ \qquad\qquad ① \Downarrow$$

$$H(X,Y) \leq H(X) + H(Y) \quad \overset{②}{\Leftarrow} \quad \underset{(X,Y)\sim P_{XY}}{\mathbb{E}}\left[\log \frac{1}{P_{XY}(X,Y)}\right] \leq \underset{(X,Y)}{\mathbb{E}}\left[\log \frac{1}{P_X \cdot P_Y}\right]$$

$$= \mathbb{E}\left[\log \frac{1}{\ }\right] + \mathbb{E}\left[\log \frac{1}{P}\right]$$

Mutual Information

$I(Y;X)$ = information in X about Y

$$\triangleq H(Y) - H(Y|X)$$

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

$$= I(X;Y) \qquad [\text{symmetric}]$$

Chain Rule for information:

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

$$\Big[I(X;Y|Z) \triangleq \mathbb{E}_{\gamma}\big[I(X|_{z=\gamma} ; Y|_{z=\gamma}) \big]$$

$$\triangleq H(X|Z) - H(X|Y,Z) \Big]$$

Data Processing Inequality

$$X - Y - Z \qquad [(X \perp\!\!\!\perp Z) \mid Y]$$

$$\Rightarrow \quad I(X;Z) \le I(Y;Z).$$

Fano (In Pset 1): $e = Pr[X \neq g(X)]$.

$$\cancel{H(Y|X)}\ H(X|Y) \le h(\overset{e}{Pr[\cancel{X \neq g(x)}]}) + e \cdot \log|\Omega|$$