Recap

Mutual Information:

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

$$0 \leq I(X;Y) \leq \min\{H(X), H(Y)\}$$

$$I((X_1 \dots X_n); Y) = \sum_i I(X_i; Y | X_1 \dots X_{i-1})$$

Data Processing:

$$X \to Y \to Z$$

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

HW 1 on Gradescope (Due 1/24)

Graph Entropy

[Körner 1973]

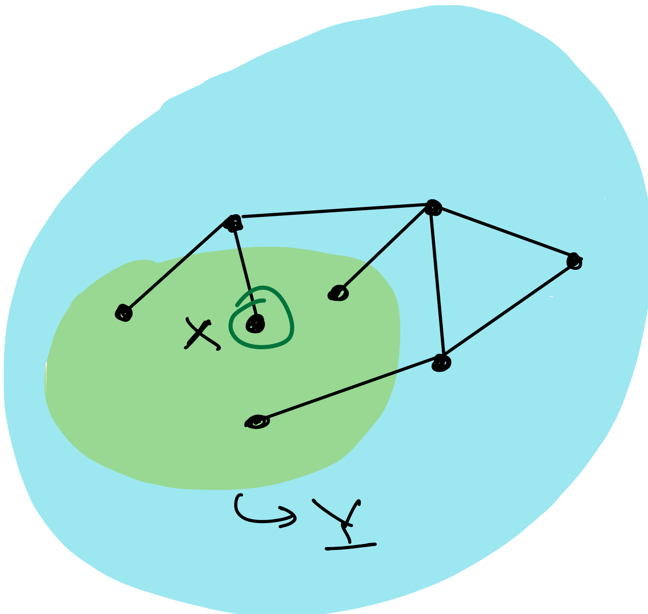
$G = (V, E)$

$$H(G) = \min I(X; Y)$$

s.t. $X \in V$ is uniform over V — can be replaced by any desired P

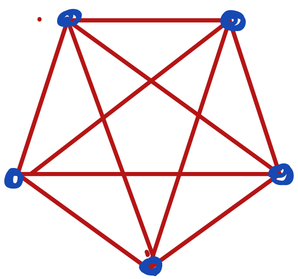
$Y \subseteq V$ contains X.

and has no edges
(Y is an independent set)



Confusion allowed
when $(u, v) \notin E$

$H(G) = \min I(X;Y)$ s.t. X is uniform on V
Y is an independent set containing X

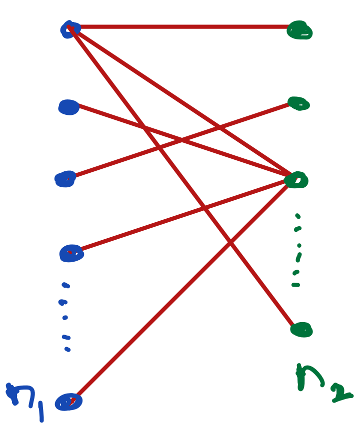


Complete graph : K_n

$H(K_n) =$ $\qquad I(X;Y) = H(X)$

$Y = \{X\}$

$= \log n$



Bipartite G

$H(G) = \quad Y = \{\bullet, \bullet\}$

$H(G) = I(X;Y) \le H(Y) \le 1$

$H_2\left(\frac{n_1}{n_1+n_2}\right)$

Ex: $H(G) \le \log(n/\alpha(G))$ $\quad \alpha(G) =$ max independent set

Subadditivity of graph entropy

- $G_1 = (V, E_1)$, $G_2 = (V, E_2)$, $G = (V, E_1 \cup E_2)$

Then $H(G) \le \underbrace{H(G_1)}_{X, Y_1} + \underbrace{H(G_2)}_{X, Y_2}$

<u>Proof:</u>

Consider sampling x, and then independently $Y_1 | X = x$ and $Y_2 | X = x$

$Y_1 \cap Y_2$ is independent set in G containing x

$$\therefore H(G) \le I(X ; Y_1 \cap Y_2)$$

$$\le I(X ; (Y_1, Y_2)) \quad \text{(data processing)}$$

$$= I(X ; Y_1) + I(X ; Y_2 | Y_1)$$

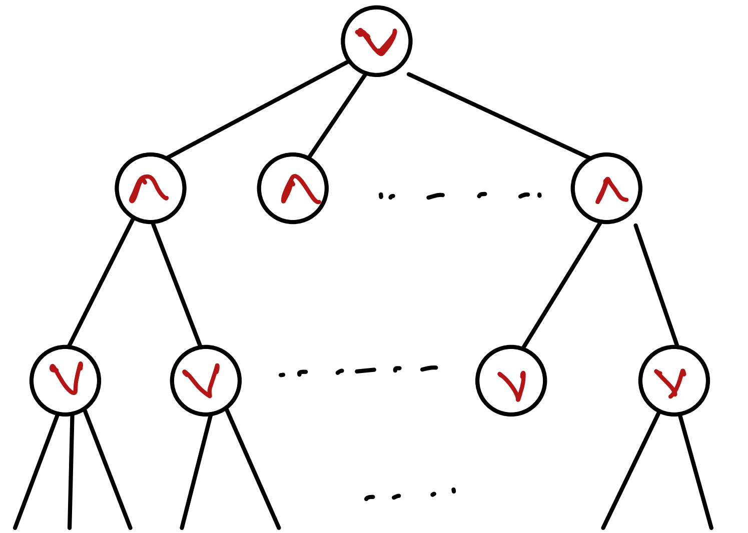
$$= I(X, Y_1) + \underbrace{H(X | Y_1)}_{\le H(X)} - \underbrace{H(Y_2 | X, Y_1)}_{= H(Y_2 | X)}$$

Covering K_n with bipartite graphs

$$K_n = G_1 \cup \cdots \cup G_r$$

$$\log n = H(K_n) \le \underbrace{H(G_1)}_{\le 1} + \cdots + H(G_r)$$

[Newman, Wigderson 95]



Min formula size for

$$f(z) = z_1 + \cdots + z_n \geq 2$$

Kullback-Leibler divergence (KL-divergence)

Measure of "distance" between distributions

$P(X)$ and $Q(X)$ for the same r.v. X

$$\underline{D(P \| Q)} = \sum_{x \in X} p(x) \cdot \log \frac{p(x)}{q(x)}$$

relative entropy

$$= \underbrace{\sum_{x \in X} p(x) \log \frac{1}{q(x)}}_{\text{Cross entropy}} - \underbrace{\sum_{x \in X} p(x) \log \frac{1}{p(x)}}_{H(X) \text{ for distribution } P(X)}$$

$$D(P \| Q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

Say $\mathcal{X} = \{a, b, c\}$

	a	b	c
P:	1/2	1/2	0
Q:	1/3	1/3	1/3

$$D(P \| Q) = \frac{1}{2} \log \frac{1/2}{1/3} + \frac{1}{2} \log \frac{1/2}{1/3} = 2 \times \frac{1}{2} (\log 3 - 1)$$

$$D(Q \| P) = \frac{1}{3} \log \frac{1/3}{1/2} + \frac{1}{3} \log \frac{1/3}{1/2} + \frac{1}{3} \cdot \log \frac{1/3}{0} = \infty$$

- $D(P\|Q) \geq 0$ $\quad$ ($=0$ if and only if $P=Q$)

Proof:

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

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

$Y = \frac{p(x)}{q(x)}$

w.p. $q(x)$

$$= \mathbb{E}\, Y \log Y$$

$$\geq (\mathbb{E}\, Y) \log (\mathbb{E}\, Y)$$

$$= 0$$

$\mathbb{E}\, Y = \sum \cancel{q(x)} \frac{p(x)}{\cancel{q(x)}}$

$= 1$

Joint Convexity

$$D(\alpha \cdot P_1 + (1-\alpha) \cdot P_2 \,\|\, \alpha \cdot Q_1 + (1-\alpha) \cdot Q_2) \leq \alpha \cdot D(P_1 \| Q_1) + (1-\alpha) \cdot D(P_2 \| Q_2)$$

Ex: $$(a_1 + a_2) \cdot \log \frac{a_1 + a_2}{b_1 + b_2} \leq a_1 \cdot \log \frac{a_1}{b_1} + a_2 \cdot \log \frac{a_2}{b_2}$$

Proof: $$\sum_x (\alpha \cdot p_1(x) + (1-\alpha) \cdot p_2(x)) \log \frac{(\overbrace{\alpha \cdot p_1(x)}^{a_1} + \overbrace{(1-\alpha) \cdot p_2(x)}^{a_2})}{(\underbrace{\alpha \cdot q_1(x)}_{b_1} + \underbrace{(1-\alpha) \cdot q_2(x)}_{b_2})}$$

$$\leq \sum_x \alpha \cdot p_1(x) \cdot \log \frac{\cancel{\alpha} \cdot p_1(x)}{\cancel{\alpha} \cdot q_1(x)} + \sum_x (1-\alpha) \cdot p_2(x) \cdot \log \frac{\cancel{(1-\alpha)} \cdot p_2(x)}{\cancel{(1-\alpha)} \cdot q_2(x)}$$

$$= \alpha \cdot D(P_1 \| Q_1) + (1-\alpha) \cdot D(P_2 \| Q_2)$$

Chain rule

$$D(P(X,Y) \| Q(X,Y)) = D(P(X) \| Q(X)) + D(P(Y|X) \| Q(Y|X))$$

<u>Proof:</u>

$$\sum_{x,y} p(x,y) \cdot \log \frac{p(x,y)}{q(x,y)} \quad \frac{p(x)}{q(x)} \cdot \frac{p(y|x)}{q(y|x)}$$

$$= \sum_{x} p(x) \cdot \Big(\underbrace{\sum_{y} p(y|x)}_{=1}\Big) \cdot \log \frac{p(x)}{q(x)} + \sum_{x} p(x) \cdot \sum_{y} p(y|x) \log \frac{p(y|x)}{q(y|x)}$$

$$= D(P(X) \| Q(X)) + \mathbb{E}_{x \sim p} D(P(Y|X=x) \| Q(Y|X=x))$$

Other distance measures

$$\delta_{TV}(P, Q) = \frac{1}{2}\|P - Q\|_1 = \frac{1}{2}\sum_x |p(x) - q(x)|$$

- Let $f : \mathcal{X} \to \{0, 1\}$ be any labeling function distinguishing P and Q. Then,

$$\left| \mathop{\mathbb{E}}_{x \sim P} f(x) - \mathop{\mathbb{E}}_{x \sim Q} f(x) \right| \leq \frac{1}{2}\|P - Q\|_1 = \delta_{TV}(P, Q)$$

<u>Proof</u>:

$$\left| \sum_x (p(x) - q(x)) \cdot \left(f(x) - \frac{1}{2}\right) + \cancelto{0}{\sum_x (p(x) - q(x)) \cdot \frac{1}{2}} \right|$$

$$\leq \sum_x |p(x) - q(x)| \cdot \underbrace{\left|f(x) - \frac{1}{2}\right|}_{= 1/2}$$