1

Tue 3/19

Last time:

✓ Bregman's Theorem on Permanent.

✓ Shearer's Lemma (generalization of sub-additivity of entropy)

✓ Application: # triangles in a graph with l edges.

Today: Graph Entropy.

* Source emitting symbols $X \in V$.

Source Coding Theorem achieves rate $H(X)$ (in the limit) and this is the best to hope for. (even with vanishing error prob.)

* It's possible to do better if limited "Confusion" is allowed.

* We model the allowed confusions as a graph over the alphabet symbols.

2

$G = (V, E)$, V = the alphabet,
$\{a, b\} \in E$ if a and b must be distinguished.

* Suppose one symbol is emitted by the source, uniformly at random.

* Encoder: $\mathrm{Enc}: V \longrightarrow \{0,1\}^R$

~~Decoder: Dec: $\{0,1\}^R \to V$~~

* Requirement: $\forall \{a,b\} \in E$, ~~$\mathrm{Dec}(\mathrm{Enc}(a)) \neq \mathrm{Dec}(\mathrm{Enc}(b))$~~ $\mathrm{Enc}(a) \neq \mathrm{Enc}(b)$.

Intuition: Best possible R for 1 source symbol = $\lceil \log \chi(G) \rceil$.

Complete graph: $G = K_k \Rightarrow R_{opt} = \lceil \log |V| \rceil$.

* More meaningful theory for $t \gg 1$ iid source symbols.

Def: $(X_1, \ldots, X_t)$ is distinguishable from $(Y_1, \ldots, Y_t)$ iff $\exists\, i \in [t]$ s.t. $\{X_i, Y_i\} \in E$.

* Formally, when $t > 1$, source is emitting a symbol from the t'th power of G.

③

* $G^t := (V^t, E^t)$, $V^t := \{ (v_1, \ldots, v_t) \text{ s.t. } v_i \in V \}$.

$$\{(v_1, \ldots, v_t), (w_1, \ldots, w_i)\} \in E^t \text{ iff}$$

$$\exists i \in [t] \text{ s.t. } \{v_i, w_i\} \in E.$$

* p^t = t-use distribution of the source:

$$p^t(v_1, \ldots, v_t) := \prod_{i \in [t]} p(v_i). \quad \left(p := \text{source distribution}\right)$$

* ~~Asymptotically, we allow $t \to \infty$ and also~~ allow a vanishable "error" ε.

That is, for an ε probability over the source output, the encoder is allowed to behave arbitrarily (e.g., output $\perp$).

* If $\varepsilon = 0$ (and source has full support), we are looking at the chromatic number of G^t. That is, the best achievable rate is $\lim_{t \to \infty} \frac{\log \chi(G^t)}{t}$.

④

* If $\varepsilon > 0$, we define ~~H(G,P)~~ the "entropy" of G as the best achievable rate.

That is,

$$H(G,P) := \lim_{t\to\infty} \min_{\substack{U \subseteq V^t, \\ P^t(U) \geq 1-\varepsilon}} \frac{1}{t} \log \chi(\underbrace{G^t(U)}_{\text{subgraph of } G \text{ induced by } U}).$$

* Introduced by Körner. He proved the limit exists and is independent of $\varepsilon \in (0,1)$.

More importantly, he proved:

Theorem (Körner): $H(G,P) = \min I(X;Y)$,

$$\begin{cases} X \in V \text{ drawn from } P, \\ Y \subset V \text{ s.t. } X \in V \text{ and } Y \text{ is an indep. set.} \\ \quad (\text{i.e., } \{u,v\} \in Y \Rightarrow \{u,v\} \notin E. \end{cases}$$

* From now on, $P :=$ uniform on G. (and use $H(G)$)

5

Examples.

* **Empty graph:**

$X \sim$ uniform on V, $\quad Y = V$

$$\Rightarrow H(G) \le I(X;Y) = 0.$$

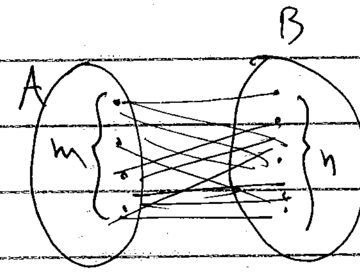
$$\Rightarrow H(G) = 0.$$

* Complete graph K_n:

$X \sim$ uniform on V, the only possibility for $Y = \{X\}$.

$$\Rightarrow I(X;Y) = H(X) = \log n.$$

* Complete bipartite $K_{m,n}$



$$Y = \begin{cases} A & \text{if } X \in A \\ B & \text{else} \end{cases}$$

$$H(G) \le I(X;Y) = H(X) - H(X|Y) = \log(m+n) - \left(\frac{m}{m+n}\log m + \frac{n}{m+n}\log n\right)$$

$$= h\left(\frac{n}{n+m}\right).$$

* ~~For~~ ~~$\le H(K_{n,n})$~~

On the other hand, Y is either a subset of A
or " " " B.

$$\Rightarrow H(X|Y) \cancel{\leq \log|A| + \log|B|}$$

$$\leq \Pr(X\in A)\log|A| + \Pr(X\in B)\log|B|$$

$$= \frac{m}{m+n}\log m + \frac{n}{m+n}\log n$$

$$\Rightarrow I(X;Y) \geq h\left(\frac{n}{n+m}\right).$$

$$\Rightarrow \begin{cases} H(G) = h\left(\frac{n}{n+m}\right). \\ H(K_{n,n}) = h(\frac{1}{2}) = 1. \end{cases}$$

★ <u>Subadditivity of Graph entropy</u> (most useful property)

$G_1 = (V, E_1)$, $G_2 = (V, E_2)$

$G = (V, E_1 \cup E_2) \Rightarrow H(G) \leq H(G_1) + H(G_2).$

<u>Proof:</u> Let $\begin{cases} H(G_1) = I(X; Y_1) \\ H(G_2) = I(X; Y_2) \\ Y_1, Y_2 \text{ independent. [conditioned on } X] \end{cases}$

$$\left[\text{i.e., } \quad p(x, y_1, y_2) = p(x) \cdot \underbrace{p_1(y_1 | x)}_{\text{dist. } Y_1} \cdot \underbrace{p_2(y_2 | x)}_{\text{dist. } Y_2} \right]$$

$Y_1 \cap Y_2$ is independent set

$$\Rightarrow H(G) \leq I(X; Y_1 \cap Y_2)$$

$$\leq I(X; Y_1, Y_2) \qquad \text{(data processing)}$$

$$= H(Y_1, Y_2) - \underbrace{H(Y_1, Y_2 | X)}$$

$$= H(Y_1, Y_2) - H(Y_1 | X) - H(Y_2 | X) \qquad (Y_1, Y_2 \text{ indep. given } X)$$

$$\leq H(Y_1) + H(Y_2) - H(Y_1 | X) - H(Y_2 | X) \qquad \text{(subadditivity of entropy)}$$

$$= H(G_1) + H(G_2).$$

$\square$

8

* **Monotonicity of $H(G)$:**

$G=(V,E)$, $F=(V,E')$, $E \subseteq E'$

$$\Rightarrow H(G) \leq H(F)$$

Proof: (X,Y) achieving $H(F)$ is feasible for G. $\square$

Corollary: $H(G_1) \leq H(G_1 \cup G_2) \leq H(G_1) + H(G_2)$ $\square$

* **Disjoint Union:** $G_1, \ldots, G_k$ connected components of G

$\rho_i := \dfrac{|V(G_i)|}{|V(G)|}$ (probability mass of the i^{th} component)

$$\Rightarrow H(G) = \sum_{i \in [k]} \rho_i \cdot H(G_i).$$

Proof: First, $H(G) \geq \sum p_i H(G_i)$:

Suppose $\begin{cases} H(G) = I(X; Y). \\ Y_i := Y \cap V(G_i). \end{cases}$

$$\ell(x): V(G) \to [k], \quad V(x) = i \text{ iff } x \in V(G_i).$$

$$H(G) = I(X; Y_1, \dots, Y_k) \underset{\text{data processing}}{=} I(X, \ell(X); Y_1, \dots, Y_k)$$

Chain rule

$$= I(\ell(x); Y_1, \dots, Y_k) + \underbrace{I(X; Y_1, \dots, Y_k \mid \ell(X))}$$

$(I(\cdot) \geq 0)$

$$\geq \sum_{i \in [k]} \Pr[\ell(X) = i] \cdot \underbrace{I(X; Y_1, \dots, Y_k \mid \ell(X) = i)}_{\downarrow \text{Chain}}$$

$$I(X; Y_i \mid \ell(x) = i) + I(X; Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_k \mid \ell(x) = i, Y_i)$$

$$\geq \sum_{i \in [k]} p_i \, I(X; Y_i \mid \ell(x) = i)$$

$$\geq \sum p_i \, H(G_i).$$

(10)

Next, $H(G) \leq \sum \rho_i H(G_i)$:

~~define~~ let $P_i(x, y_i)$ = minimizing dist. of $H(G_i)$.

define $p(x, y_1, \ldots, y_k) = \sum_i \rho_i \, P_1(y_1) \cdots P_k(y_k) \cdot P_i(x | y_i)$

[(i.e., pick $Y_1, \ldots, Y_k$ independently, a component $i \sim \rho_i$ and sample X from the i^{th} coordinate $\sim P_i(x | y_i)$]

Observe: 1) $I(\ell(X); Y_1, \ldots, Y_k) = 0$

Since $i = \ell(x)$ is chosen independent of $Y_1, \ldots, Y_k$

2) $I(X; Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k \,|\, \ell(X) = i, Y_i) = 0$

[Since x depends on Y_i only, if $\ell(x) = i$]

3) $I(X; Y_i) = H(G_i)$

$\Rightarrow$ all the inequalities in the lower bound can be achieved with equality

$\square$