Lecture 4

Lecturer: Anup Rao *Scribe: Punyashloka Biswal*

Last time we defined the *mutual information* $I(X \wedge Y) = H(X) - H(X \mid Y)$, and proved that it had the following properties:

- For all X and Y, $I(X \wedge Y) \geq 0$.
- For all X, Y, and Z, $I(X, Y \wedge Z) = I(X \wedge Z) + I(Y \wedge Z \mid X)$

To these we now add the following:

Fact 1. *For all random variables X, Y and functions f, $I(X, f(X) \wedge Y) = I(X \wedge Y)$.*

Proof This follows from the chain rule: $I(X, f(X) \wedge Y) = I(X \wedge Y) + I(f(X) \wedge Y \mid X) = I(X \wedge Y) + H(f(X)|X) + H(f(X)|X, Y) = I(X \wedge Y)$. ■

Fact 2. *For all X, Y and functions f, $I(f(X) \wedge Y) \leq I(X \wedge Y)$.*

Proof By Fact 1 and the chain rule, $I(f(X) \wedge Y) + I(X \wedge Y|f(X)) = I(f(X), X \wedge Y) = I(X \wedge Y)$. The proof follows from the fact that $I(X \wedge Y|f(X))$ is non-negative. ■

1 Graph Entropy

Today, we study a quantity called *graph entropy* associated with the graph, first considered by Körner [1]. The original motivation for this quantity was to characterize how much information can be communicated in a setting where pairs of symbols may be confused, though we shall see that it is very useful in a variety of settings.

A subset S of the vertices V of an undirected graph $G = (V, E)$ is *independent* if no edge in the graph has both endpoints in S. Given a graph G, define the graph entropy of G

$$H(G) = \min_{X,Y} I(X \wedge Y),$$

where the minimum is taken over all pairs of random variables X, Y such that

- X is a uniformly random vertex in G.
- Y is an independent set containing X.

Let us consider some examples:

1. Suppose G has no edges. Then if X is a uniformly random vertex and Y is fixed to be the vertex set V, we get $H(G) \leq I(X \wedge Y) = 0$. But $H(G) \geq 0$, so $H(G)$ must be 0 in this case.
2. Let G be the complete graph on n vertices. Then the only independent set containing a given vertex u is the singleton set $\{u\}$. Thus there is only one available choice for the distribution of X, Y, namely $\Pr[Y = \{X\}] = 1$. $H(G) = H(X) - H(X \mid Y) = \log n - 0$, because X is completely determined by $Y = \{X\}$.

3. Let G be the complete bipartite graph $K_{n,n}$. Call the two parts of the graph A and B. One possible choice of joint distribution for X and Y is to first pick X uniformly at random, and then to choose

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

   This gives us the upper bound

$$H(G) \leq I(X \wedge Y) \leq H(X) - H(X \mid Y) \leq \log(2n) - \log n = 1.$$

   On the other hand, we claim that any valid joint distribution must satisfy $H(X \mid Y) \leq \log n$. For if Y is an independent set, then it must be a subset of either A or B. Thus, $H(X|Y) \leq \log |Y| \leq \log n$. This implies that $H(G) \geq \log(2n) - \log n = 1$.

4. Let G be the unbalanced complete bipartite graph $K_{m,n}$. We choose X and Y exactly as before and get the bound

$$H(G) \leq \log(m+n) - \frac{m}{m+n} \log m - \frac{n}{m+n} \log n = H\left(\frac{n}{m+n}\right),$$

   where $H(\cdot)$ denotes the binary entropy function, or the entropy of a biased coin. As in the previous case, we have that $H(X|Y) \leq \frac{m}{m+n} \log m + \frac{n}{m+n} \log n$, proving that $H(G) = H(\frac{n}{m+n})$.

5. Let G be a complete r-partite graph, i.e., $V = [n] \times [r]$ and $E = \{\big((i,j),(k,l)\big) \mid j \neq l\}$. Then we can adapt the proofs from the last two examples to show that $H(G) = \log r$. In fact, we can show further that if G is r-partite with parts $S_1, \ldots, S_r$, the graph entropy of G is the same as $H(Z)$, where $\Pr[Z = i] = \Pr[X \in S_i]$ for uniform vertex X. In particular, $H(G) \leq \log r$ in this case.

2 Useful Properties of Graph Entropy

The power of graph entropy comes from the fact that it can be easily controlled even when the underlying graph is manipulated in natural ways.

Proposition 3 (Subadditivity)**.** *Let $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ be graphs on the same vertex set. Then their union $G = (V, E_1 \cup E_2)$ has entropy $H(G) \leq H(G_1) + H(G_2)$.*

Proof Let $p_1(x,y)$ and $p_2(x,y)$ be the distributions that minimize $I(X \wedge Y)$ for G_1 and G_2, respectively, and let us consider the distribution

$$p(x, y_1, y_2) = p(x) \cdot p_1(y_1 \mid x) \cdot p_2(y_2 \mid x).$$

In other words, we pick X uniformly at random, and conditioned on this choice of X we pick Y_1 and Y_2 independently according to each of the conditional distributions. For a given choice of X, observe that $Y_1 \cap Y_2$ contains X and is an independent set in G. Therefore,

$$\begin{aligned}
H(G) &\leq I(X \wedge (Y_1 \cap Y_2)) \\
&\leq I(X \wedge Y_1, Y_2) && \text{by Fact 2} \\
&= H(Y_1, Y_2) - H(Y_1, Y_2 \mid X) \\
&= H(Y_1, Y_2) - H(Y_1 \mid X) - H(Y_2 \mid X) && \text{since } Y_1, Y_2 \text{ are independent conditioned on any fixing of } X \\
&\leq H(Y_1) - H(Y_1 \mid X) + H(Y_2) - H(Y_2 \mid X) && \text{by subadditivity of entropy} \\
&= H(G_1) + H(G_2).
\end{aligned}$$

■

Proposition 4 (Monotonicity)**.** *If $G = (V, E)$ and $F = (V, E')$ are graphs on the same vertex set such that $E \subset E'$, then $H(G) \leq H(F)$.*

Proof If X, Y are random variables achieving $H(F)$, then Y is also an independent set in G, so $H(G) \leq I(X \wedge Y) = H(F)$. ■

The previous two propositions can be summarized as follows. If G_1, G_2 are graphs on the same vertex set, then $H(G_1) \leq H(G_1 \cup G_2) \leq H(G_1) + H(G_2)$.

Next, we consider what happens to the graph entropy when taking disjoint unions of graphs.

Proposition 5 (Disjoint union)**.** *If $G_1, \ldots, G_k$ are the connected components of G, and for each i, $\rho_i = |V(G_i)|/|V(G)|$ is the fraction of vertices in G_i, then*

$$H(G) = \sum_{i=1}^{k} \rho_i H(G_i).$$

Proof First, we shall show that $H(G) \geq \sum \rho_i H(G_i)$. Let X, Y be the random variables achieving $H(G)$. We can write $Y = Y_1, \ldots, Y_k$, where each Y_i is the intersection Y with the vertices of G_i. Define the function $l(x)$, where $l(x) = i$ if $x \in V(G_i)$. Then

$$\begin{aligned}
H(G) &= I(X \wedge Y_1, \ldots, Y_k) \\
&= I(X, l(X) \wedge Y_1, \ldots, Y_k) && \text{by Fact 1} \\
&= I(l(X) \wedge Y_1, \ldots, Y_k) + I(X \wedge Y_1, \ldots, Y_k \mid l(X)) \\
&\geq \textstyle\sum_i \Pr(l(X) = i)\ I(X \wedge Y_1, \ldots, Y_k \mid l(X) = i) && (I(\cdot) \geq 0) \\
&= \textstyle\sum_i \rho_i \left(I(X \wedge Y_i \mid l(X) = i) + I(X \wedge Y_1, Y_2, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k \mid l(X), Y_i)\right) \\
&\geq \textstyle\sum_i \rho_i I(X \wedge Y_i \mid l(X) = i) \\
&\geq \textstyle\sum_i \rho_i H(G_i),
\end{aligned}$$

where the last inequality follows from the fact that in $(X, Y_i)|l(X) = i$, X is a uniform vertex of $V(G_i)$, and Y_i is an independent set containing X.

Now we proceed to the upper bound. For $i = 1, \ldots, k$, let $p_i(x, y_i)$ be the minimizing distribution in the definition of $H(G_i)$. Then we can define the following joint distribution on $X, Y_1, \ldots, Y_k$:

$$p(x, y_1, \ldots, y_k) = \sum_i \rho_i \cdot p_1(y) \cdots p_k(y_k) \cdot p_i(x \mid y_i).$$

In words, we choose $Y_1, \ldots, Y_k$ independently according to the marginal distributions of $p_1, \ldots, p_k$, then pick a component i according to the distribution $\rho_1, \rho_2, \ldots, \rho_k$ and finally sample X from that component with conditional distribution $p_i(x|y_i)$. We can verify that for this choice, all the inequalities above hold with equality:

- We choose the component in which to put X according to the weights ρ_i, and independently choose the independent sets $Y_1, \ldots, Y_k$. Thus $I(l(X) \wedge Y_1, \ldots, Y_k) = 0$.
- Conditioned on $l(X) = i$ $Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k$ are independent of X, Y_i. Thus, $I(X \wedge Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots Y_k \mid l(X) = i, Y_i) = 0$.
- The last inequality is tight since conditioned on $l(X) = i$, the joint distribution $X, Y_i|l(X) = i$ is the minimizing distribution for the graph entropy.

■

References

[1] J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs, *in* "Transactions of of the 6th Prague Conference on Information Theory, etc.," 1971, Academia, Prague, (1973), 411–425.