CS 229r Information Theory in Computer Science March 01, 2016

Lecture 11

Lecturer: Madhu Sudan *Scribe: Tanay Mehta*

Today, we will finish proving the lower bound for the set disjointness problem. To do this, we will introduce new concepts such as embeddings and the Hellinger distance.

Information Complexity and Set Disjointness (Minjae)

Last time, we introduced the set disjointness problem for communication protocols. In order to prove lower bounds for this, we defined the concept of information cost as the mutual information between the inputs to the problem and communication protocol. The information cost of a protocol computing a function is a lower bound for the communication complexity. Therefore, we can prove lower bounds for the communication complexity of computing a function by showing lower bounds for its information cost. In particular, our goal is to prove the following theorem.

Theorem 1 $\mathrm{IC}_{\mu,f}(\mathrm{Disj}) = \Omega(n)$

We will show this theorem by proving the series of lemmas outlined in the previous lecture. Last time, we proved the following.

Lemma 2 $\mathrm{I}(X,Y;\Pi|Z) \geq \sum_{i=1}^{n} \mathrm{I}(X_i,Y_i;\Pi|Z)$

We will now prove the following.

Lemma 3 $\mathrm{I}(X_i,Y_i;\Pi|Z) \geq \mathrm{IC}_{\mu_i|Z_i}(\mathrm{Disj}^1)$

Recall that the distribution μ is defined as follows. Let $\tilde{X}_1,\ldots,\tilde{X}_n$, $\tilde{Y}_1,\ldots,\tilde{Y}_n$, and $Z_1,\ldots,Z_n$ be i.i.d. from the uniform distribution $U(\{0,1\})$. Define $X_i = \tilde{X}_i \wedge Z_i$ and $Y_i = \tilde{Y}_i \wedge \bar{Z}_i$. Then, μ is the distribution $\mu \to (X,Y,Z)$.

Note that Lemma 3 does not follow immediately from definitions because Π does not depend on one-bit inputs but n-bit inputs. We can get around this issue by using *embeddings*.

Embedding:

$$\begin{aligned} R &= \{Z_j\}_{j=1,j\neq i}^{n} \\ R_A &= \{\tilde{X}_j\}_{j=1;j\neq i}^{n} \\ R_B &= \{\tilde{Y}_j\}_{j=1;j\neq i}^{n} \\ \\ X_j &:= \tilde{X}_j \wedge Z_j \\ Y_j &:= \tilde{Y}_j \wedge \bar{Z}_j \\ \\ X &\leftarrow \{X_j\} \\ Y &\leftarrow \{Y_j\} \end{aligned}$$

Embeddings can be thought of as a zero-communication protocol with common randomness independent of the inputs X_i and Y_i. For this specific protocol, we claim the following.

Claim 4 $\mathrm{Disj}^n(X,Y) = \mathrm{Disj}^1(X_i,Y_i)$

Proof

$$\begin{aligned}
\mathrm{Disj}^n(X,Y) &= \bigvee_{j=1}^{n} X_j \wedge Y_j \\
&= \left(\bigvee_{j=1}^{n} X_j \wedge Y_j\right) \vee (X_i \wedge Y_i) \\
&= \left(\bigvee_{j=1}^{n} \tilde{X}_j \wedge \tilde{Y}_j \wedge Z_j \wedge \bar{Z}_j\right) \vee (X_i \wedge Y_i) \\
&= X_i \wedge Y_i
\end{aligned}$$

where the last equality follows since $Z_j \wedge \bar{Z}_j = 0$ from the previous line.
■

Proof [Lemma 3]
Define the following protocol making use of the above embedding.

$$\Pi^1 \quad := \quad \Pi(\mathrm{embed}(X_i), \mathrm{embed}(Y_i))$$

From definitions, we have that

$$\mathrm{I}(X_i, Y_i; \Pi^1 | Z) \quad \geq \quad \mathrm{IC}_{\mu_i|Z_i}(\mathrm{Disj}^1)$$

since IC is taken as the minimum over distributions. Furthermore, we have by construction that

$$\mathrm{I}(X_i, Y_i; \Pi | Z) \quad = \quad \mathrm{I}(X_i, Y_i; \Pi^1 | Z)$$

Combining these two, we have the statement of lemma 3.
■

Note that the embedding construction fully requires common randomness Z.

We will now prove one final lemma that implies Theorem 1.

Lemma 5 $\mathrm{I}(X_i, Y_i; \Pi | Z) > 0$ *where* $X_i, Y_i \in \{0, 1\}$

Since X_i, Y_i are one-bit values, we will suppress the subscripts for ease and write them as X, Y. We can also now restrict Z to one-bit. There is some protocol to account for throwing out $n-1$ bits. We will also define the notation $\Pi^{XY} := \Pi(X,Y)$. Therefore, we have 4 possible transcripts: $\Pi^{00}, \Pi^{01}, \Pi^{10}, \Pi^{11}$. Note that we have

$$\mathrm{I}(X, Y; \Pi | Z) = \frac{1}{2}(\mathrm{I}(Y; \Pi^{0Y}) + \mathrm{I}(X; \Pi^{X0}))$$

Now, we claim that all possible values of X, Y cannot be identically distributed. Suppose that $\Pi^{00} \equiv \Pi^{01}$ and $\Pi^{00} \equiv \Pi^{10}$. This implies $\mathrm{I}(Y; \Pi^{0Y}) = 0$ and $\mathrm{I}(X; \Pi^{X0})$ due to the *rectangle property.*

Rectangle property:
Suppose we have X, Y and X', Y' such that

$$\Pi(X, Y) = t = \Pi(X', Y')$$

then for deterministic protocols, we have that

$$\Pi(X', Y) = t = \Pi(X, Y')$$

We can extend this property to private randomized communication protocols by factoring out the private randomness and making it part of the input.

$$\begin{aligned} & \Pi((X,R_A),(Y,R_B)) && = t = \Pi((X',R'_A),(Y',R'_B)) \\ \implies & \Pi((X',R'_A),(Y,R_B)) && = t = \Pi((X,R_A),(Y',R'_B)) \end{aligned}$$

This statement implies the following by looking at the probability for X, Y to yield a specific transcript t.

$$\Pr[\Pi(X,Y)=t]\cdot\Pr[\Pi(X',Y')=t] = \Pr[\Pi(X,Y')=t]\cdot\Pr[\Pi(X',Y)=t]$$

The rectangle property for probabilistic protocols suggests that there is distance property between the input distributions. The correct distance measure for this case turns out to be the *Hellinger distance.*

Hellinger distance: For a finite universe, we define the Hellinger distance to be

$$\begin{aligned} h(P,Q) &= \frac{1}{\sqrt{2}}||\sqrt{P}-\sqrt{Q}||_2 \\ &= \sqrt{1-\sum_{w\in S}\sqrt{P(w)\cdot Q(w)}} \end{aligned}$$

where S is the support. The above sum is known as the *Bhattacharya coefficient.* The $\frac{1}{\sqrt{2}}$ factor scales the distance to be between 0 and 1. The Hellinger distance has the triangle property and other properties of distance measures. As a side note, the Hellinger distance can be generalized to continuous settings with the Lebesgue measure.

Recall the *total variation distance.*

$$\mathrm{TVD}(P,Q) = \frac{1}{2}||P-Q||_1$$

By the properties of the 1 and 2-norms, we have the following bounds.

$$h^2(P,Q) \le \mathrm{TVD}(P,Q) \le \sqrt{2}h(P,Q)$$

Ideally, we would like to find the total variation of the distributions in questions. However, this is difficult due to the rectangular property. The Hellinger distance has a property similar to the rectangular property.

Lemma 6 (Cut & Paste Lemma) $h(\Pi(X,Y),\Pi(X',Y')) = h(\Pi(X',Y),\Pi(X,Y'))$

Therefore, we have $h(\Pi^{01},\Pi^{10}) = 0 \implies h(\Pi^{00},\Pi^{11}) = 0$. Now, we can prove Lemma 5.

Proof [Lemma 5] Take $Z = 0$. Then, we can view Y as an indicator in $\mathrm{I}(Y;\Pi^{0Y})$ where

$$\begin{aligned} Y = 1 \to t\ \Pi^{01} \\ Y = 0 \to t\ \Pi^{00} \end{aligned}$$

Define the *Jensen-Shannon divergence* for arbitrary distributions P, Q to be

$$D_{JS}(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M)$$

where $M := \frac{1}{2}(P+Q)$ and D_{KL} is the KL-divergence. D_{JS} can be thought of as a symmetric (on the inputs) version of the KL-divergence.

Then, we have that

$$\begin{aligned} \mathrm{I}(Y;\Pi^{0Y}) &= D_{JS}(\Pi^{01}||\Pi 00) \\ &= \frac{1}{2}\left(D_{KL}(\Pi^{01}||\Pi^{0\frac{1}{2}}) + D_{KL}(\Pi^{00}||\Pi^{0\frac{1}{2}})\right) \end{aligned}$$

where $\Pi^{0\frac{1}{2}} = \frac{(\Pi^{00}+\Pi^{01})}{2}$. By Pinsker's inequality,

$$\begin{aligned} D_{KL}(\Pi^{01}||\Pi^{0\frac{1}{2}}) &\geq \frac{1}{2}\left(\text{TVD}(\Pi^{01}, \Pi^{0\frac{1}{2}})\right)^2 \\ &= \frac{1}{2}\left(\frac{\text{TVD}(\Pi^{01}, \Pi^{00})}{2}\right)^2 \end{aligned}$$

Note that $\text{TVD}(\Pi^{01}, \Pi^{10}) \geq \beta$ for some β. This implies that $\text{TVD}(\Pi^{01}, \Pi^{00}) \geq \beta/2$, or $\text{TVD}(\Pi^{00}, \Pi^{10}) \geq \beta/2$. Therefore we have a lower bound of $\frac{\beta^2}{16}$.

We want a lower bound for the following.

$$\begin{aligned} \text{TVD}(\Pi^{01}, \Pi^{10}) &\geq h^2(\Pi^{01}, \Pi^{10}) \\ &= h^2(\Pi^{00}, \Pi^{11}) \\ &\geq \frac{\text{TVD}(\Pi^{00}, \Pi^{11})^2}{2} \end{aligned}$$

For these distributions, note that Π^{11} accepts transcript t with $1-\epsilon$ probability and Π^{00} accepts with probability ϵ. Therefore, $\text{TVD}(\Pi^{00}, \Pi^{11}) \geq 1 - 2\epsilon$. This gives us a final, positive lower bound of $\frac{(1-2\epsilon)^4}{64}$.
■