CS 229r Information Theory in Computer Science	Feb 25, 2016
Lecture 10	
Lecturer: Madhu Sudan	*Scribe: Nathan Manohar*

Overview

Today:

1. Wrap Up Inner Product Lower Bound
2. Introduce Information Complexity Lower Bound for Set Disjointness

Review:

1. Communication Complexity (deterministic and probabilistic)
2. Rank Lower Bound (for deterministic)
3. Discrepancy (for randomized)

1 Inner Product (Madhu)

$$\mathrm{IP}(x,y) = \prod_{i=1}^{n} (-1)^{x_i \wedge y_i}$$

and

$$\mu = \mathrm{UNIF}(\{0,1\}^n \times \{0,1\}^n)$$

$$M_{\mu,\mathrm{IP}} = [M_{\mu,IP}(x,y)]_{x,y}$$

where $M_{\mu,IP}(x,y) = 4^{-n} \cdot IP(x,y)$.

We have that

$$\mathrm{Disc}(M,R) = |\sum_{(x,y)} R(x,y) \cdot M(x,y)|$$

We define

$$\mathrm{Disc}(M) = \max_{\text{rectangles } R} \{\mathrm{Disc}(M,R)\}$$

We have that

$$\mathrm{Disc}(M) \leq (1-\varepsilon) \cdot 2^{-k} \implies CL_{\varepsilon}(\mathrm{IP}) \geq k$$

If we rewrite $\mathrm{Disc}(M)$, we can express this as

$$\begin{aligned}
\mathrm{Disc}(M) &= \max_{S,T \subset \{0,1\}^n} |\sum_{x \in S, y \in T} M(x,y)| \\
&= \max_{U,V \in \{0,1\}^{2^n}} |U^T \cdot M \cdot V| \\
&\leq \max_{U,V \in \mathbb{R}^{2^n}, ||U||_2, ||V||_2 \leq 2^{n/2}} |U^T \cdot M \cdot V| \\
&= 2^n \lambda_{\max}(M)
\end{aligned}$$

$$M_{\mu_1,\mathrm{IP}_1} = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & -1/4 \end{pmatrix}$$

which has eigenvalues $\pm\frac{1}{\sqrt{8}}$. So, the maximum eigenvalue of $M_{\mu,\mathrm{IP}}$ is $(\frac{1}{\sqrt{8}})^n$, so we have that

$$\mathrm{Disc}(M) \leq 2^n \lambda_{\max}(M) = 2^n (\frac{1}{\sqrt{8}})^n = 2^{-n/2}$$

2 Information Complexity (Alex + Minjae)

Outline

1. Set disjointness
2. Information complexity
3. IC $\leq$ CC
4. choice of distribution
5. $IC^n \geq n \cdot IC^1$
6. $IC^1 = \Omega(1)$

Definition 1 (Set Disjointness Problem) *Let $x, y \in \{0,1\}^n$. We define $Disj(x, y) = \bigvee_{i=1}^{n}(x_i \wedge y_i)$. In particular, we have that*

$$Disj(x,y) = \begin{cases} 1 \text{ if } \exists i, x_i = y_i = 1 \\ 0 \text{ otherwise} \end{cases}$$

2.1 Notation

f is a function from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}$. μ is a distribution of $\{0,1\}^n \times \{0,1\}^n$. ε is the error parameter. π is the ε-error protocol for computing f (meaning π computes f correctly on every input with error probability $\leq \varepsilon$). $\pi(x, y, R, R_A, R_B)$ is the transcript of π with inputs x, y.

Definition 2 *Information Cost of π for computing f with respect to μ, ε is defined as*

$$IC^{\pi}_{\mu,\varepsilon}(f) = I(X, Y; \pi | R)$$

2.2 Ex. 1-bit AND

$x, y \in \{0,1\}$. Want to compute $f(x, y) = x \wedge y$.

Protocol:
δ error, $k >> \frac{1}{\delta}$. Repeat the following k times.

$$A \to B \begin{cases} 1 & \text{w.p. } 1-\delta \\ X & \text{w.p. } \delta \end{cases}$$

and

$$B \to A \begin{cases} 1 & \text{w.p. } 1-\delta \\ Y & \text{w.p. } \delta \end{cases}$$

If either is 0, output 0. Else, output 1. We see that the probability of error is

$$(1-\delta)^k \leq e^{\delta k} = \varepsilon$$

The information cost is

$$IC^{\pi}_{\mu,\varepsilon}(f) = I(X, Y; \pi) = H(X, Y) - H(X, Y|\pi) \leq \log_2 4 = 2$$

The bound is strictly less than 2 because in the case where $x \wedge y = 0$, we might only learn that one of the bits is 0 and be unsure about the other bit.

Lemma 3 $IC_{\mu,\varepsilon}(f) \leq CC_{\varepsilon}(f)\ \forall \mu, \varepsilon > 0$

Proof $CC_\varepsilon(f) = |\pi^*|$, the max length of π. $CC_\varepsilon(f) = |\pi^0| \geq H(\pi)$. However, $IC_{\mu,\varepsilon}(f) = I(X,Y;\pi) \leq H(\pi)$. ■

The main theorem we want to show is

Theorem 4 $IC_{\mu,f}(Disj) = \Omega(n)$

To prove this, we prove lemmas that roughly state the following

Lemma 5 *(Informal)* $IC(Disj^n) \geq n \cdot IC(Disj^1)$

Lemma 6 *(Informal)* $IC(Disj^1) \geq \Omega(1)$

Choice of Distribution:
To prove this, we will make use of the following distribution.

$$Z_i \sim \text{Unif}(\{0,1\})$$

If $Z_i = 0$, then $X_i = 0$ and $Y_i \sim \text{Unif}(\{0,1\})$. If $Z_i = 1$, then $Y_i = 0$ and $X_i \sim \text{Unif}(\{0,1\})$.

$$(X_i, Y_i, Z_i) \sim \eta$$
$$Z = (Z_1, \ldots, Z_n)$$
$$(X, Y, Z) \sim \zeta = \eta^n$$

Properties of ζ:

1. $\forall x, y \in \text{Supp}(\zeta),\ Disj(x,y) = 0$
2. $\forall Z \in \{0,1\}^n, X \perp Y | Z$
3. $(X_i, Y_i) \perp \{(X_j, Y_j)\}_{j \neq i}$

Definition 7 (Conditional Information Cost) $CIC^\pi_{\zeta,\varepsilon}(f) = I(X,Y;\pi|Z)$

Definition 8 (Conditional Information Complexity) $CIC_{\zeta,\varepsilon}(f) = \min_\pi\{CIC^\pi_{\zeta,\varepsilon}(f)\}$

Lemma 9 $CIC_{\zeta,\varepsilon}(f) \leq IC_{\mu,\varepsilon}(f)$ *where* $\mu = \zeta|Z$

Proof $CIC_{\zeta,\varepsilon}(f) = I(X,Y;\pi|Z) = H(\pi|Z) - H(\pi|X,Y,Z)$. $IC_{\mu^n,\varepsilon}(f) = I(X,Y;\pi) = H(\pi) - H(\pi|X,Y)$. Since $H(\pi|X,Y) = H(\pi|X,Y,Z)$ and $H(\pi) \geq H(\pi|Z)$, the result follows. ■

Lemma 10 $CIC^n \geq n \cdot CIC^1$

We will show the following two results.

1. $I(X,Y;\pi|Z) \geq \sum_{i=1}^n I(X_i,Y_i;\pi|Z)$
2. $I(X_i,Y_i;\pi|Z) \geq CIC_{\eta,\varepsilon}(Disj^1)$

Proof (of 1):

$$\begin{aligned}
I(X,Y;\pi|Z) &= H(X,Y|Z) - H(X,Y|\pi,Z) \\
H(X,Y|Z) &= \sum_{i=1}^n H(X_i,Y_i|Z,X_1,Y_1,\ldots,X_{i-1},Y_{i-1}) \\
&= \sum_{i=1}^n H(X_i,Y_i|Z) \\
H(X,Y|\pi,Z) &= \sum_{i=1}^n H(X_i,Y_i|\pi,Z,X_1,Y_1,\ldots,X_{i-1},Y_{i-1}) \\
I(X_i,Y_i;\pi|Z) &= H(X_i,Y_i|Z) - H(X_i,Y_i|\pi,Z)
\end{aligned}$$

Since $H(X_i, Y_i|\pi, Z) \geq H(X_i, Y_i|\pi, Z, X_1, Y_1, \ldots, X_{i-1}, Y_{i-1})$, we see that $I(X, Y; \pi|Z) \geq \sum_{i=1}^{n} I(X_i, Y_i; \pi|Z)$ as desired. ■