CS 229r Information Theory in Computer Science March 07, 2019

Lecture 12

Instructor: Madhu Sudan *Scribe: Alexandra Kieras*

1 Overview

1.1 Schedule

- The Set Disjointness Problem
- Information Complexity

1.2 References

The material covered in lecture today is discussed in the following:

- Bar-Yossef, Jayram, Kumar, Sivakumar

1.3 The Set Disjointness Problem

Recall the communication model discussed in previous lectures:

Alice and Bob are two players who have private values X and Y. Their goal is to compute some function f with inputs in $\{0,1\}^{2n}$; f is often (though not always) a Boolean function.

We define the set disjointness problem as the problem of determining whether two sets, X, Y, drawn from the integers $[1...n]$ are disjoint. Formally, the problem is defined as follows:

Let $X, Y \subseteq [n]$, then

$$\text{DISJ}^n(X,Y) = 1 \implies \exists_i \text{ s.t. } X_i = Y_i = 1$$
$$\text{DISJ}^n(X,Y) = 0 \text{ o/w}$$

Exercise 1. *Show that on any product distribution $\mu = \mu_x \times \mu_y$, there exists a protocol π to compute $DISJ^n(X,Y)$ with error ε and $O(\sqrt{n})$ communication complexity.*

2 Conditional Mutual Information

For (X, Y, Z) jointly distributed, we define $I(X, Y|Z)$ as the information gained about X from Y conditioned on Z.

Formally, this is defined as follows:

$$I(X,Y|Z) = E_{Z\sim P_Z}\left[I(X|_{Z=z}, Y|_{Z=z})\right]$$
$$= H(X|Z) - H(X|Y,Z)$$

Note that, unlike entropy, conditioning does **not**, in general, reduce mutual information.

Example 2. *Conditioning does not always reduce mutual information.*

Suppose we have $X \perp\!\!\!\perp Y$*,* $Z = X \oplus Y$*, for* $X, Y \in Unif(\{0,1\}^n)$*. Then:*

$$I(X;Y) = 0$$
$$I(X,Y|Z) = n$$

Example 3. *Conditional Mutual Information of a Markov Chain*

Suppose $X - Y - Z$ *is a Markov. Then, it follows that:*

$$I(X,Y) \geq I(X,Y|Z)$$

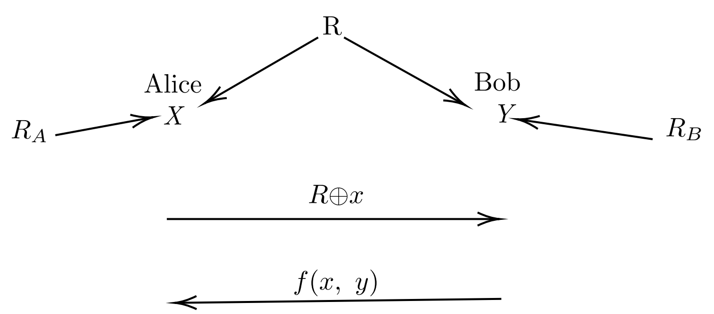
and further $I(X,Z|Y) = 0$

Exercise 4. *Prove the above claim.*

2.1 Motivation

Fix a protocol π, with error ε on all inputs while computing function f. Furthermore, fix underlying input distribution μ.

Question: How much does an observer learn about the inputs from watching the interaction?
We consider the following protocol:



Suppose the observer cannot view R (i.e, the observer cannot see the shared randomness that Alice and Bob use).

Then:

$$I(X,Y|R \oplus X, f(X,Y)) \leq H(f(X,Y))$$

That is, the observer learns little from watching the procedure, because they cannot observe the randomness R. So, we should condition on randomness R but not on R_A, R_B (the private randomness that Alice and Bob use, respectively).

2.2 Information Complexity

Definition 5. We define the *information complexity* for a protocol π over a distribution μ as:

$$IC_\mu(\pi) = I((X,Y),\pi|R)$$

The information complexity of a function f is the minimum over all protocols π which compute that function with small error.

In particular, if π is a $k-$bit protocol that $\varepsilon-$computes f, then $IC_\mu(f) \leq k$.

3 Proof

We seek to show:

$$\exists \mu_n \text{ s.t. } IC_\mu(\text{DISJ}^n = \Omega(n)$$

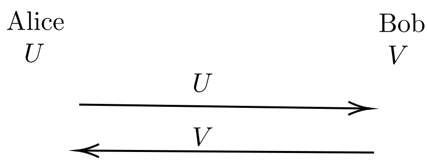
Additional intuition (which we will not prove):

$$IC_{\mu_n}(\text{DISJ}^n) \geq nIC_{\mu_1}(\text{DISJ}^1)$$

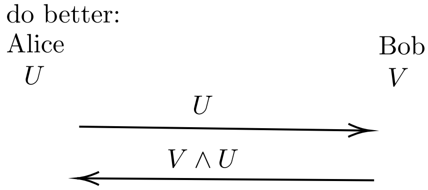
$$IC_{\mu_1}(\text{DISJ}^1) = \Omega(1)$$

3.1 Detour: the AND function

Let $U, V \in \{0, 1\}$. Suppose that Alice has U, Bob has V, and they would like to compute $f = U \wedge V$.
Alice and Bob directly exchange U, V.



This is reasonable, and Alice and Bob will exchange two bits per function computation. However, we can do better:



In this exchange, Alice sends U to Bob, and Bob sends the computed $U \wedge V$ back.

We see that if $U = 0$, we will reveal 1 bit and if $U = 1$. we will reveal both bits. So, we will reveal 3/2 bits in expectation.

We can come up with an even better protocol, however. Suppose Alice chooses a time $t_A \in [0, 1]$ and Bob a time $t_B \in [0, 1]$. At time t_A Alice sends U to Bob if $U = 0$. At time t_B, Bob sends 0 to Alice if $V = 0$. At time $t = 1$, Bob and Alice both exchange bits only if they have not already done so.

We consider the complexity of this operation:

$$\text{if } (UV) = \left.\begin{matrix} 00 \\ 01 \\ 10 \end{matrix}\right\} \text{ we reveal 1 bit}$$

$$\text{if } (UV) = 11 \text{ we reveal both}$$

So, we reveal $\frac{5}{4}$ bits on average.

3.2 Proof of DISJ bound

We return to the proof of the fact:

$$\exists \mu \text{ s.t. } IC_{\mu}(\text{DISJ}^n) = \Omega(n)$$

Let μ be a distribution over (X_i, Y_i) i.i.d such that:

$$(X_i, Y_i) = \begin{cases} 00 & p = \frac{1}{2} \\ 01 & p = \frac{1}{4} \\ 10 & p = \frac{1}{4} \end{cases}$$

We notice that X, Y are always disjoint! Rather than proving good performance of DISJ for this specific input, the procedure π we construct will have good performance on all inputs.

Let (X, Y, Z), with $Z \sim Unif(\{0,1\}^n)$. Then, to sample (X, Y, Z) with the distribution described above, we can execute:

For $i = 1$ to n do:

- if $Z_i = 0$ then $X_i = 0$ and $Y_i \sim Unif(\{0,1\})$
- $Z_i = 1$ then $Y_i = 0$ and $X_i \sim Unif(\{0,1\})$

3.3 Conditional Information Cost

Definition 6. We define the conditional information cost, CIC, as:

$$CIC_p(\pi) = I((X, Y), \pi | R, Z)$$

We then prove the claims:

1. $CIC_{\mu}(\text{DISJ}^n) \geq nCIC(\text{DISJ}^1)$
2. $CIC_{\mu}(\text{DISJ}^1 = \Omega(1)$

In particular, we prove Claim One in class today and Claim Two in the next lecture.

3.4 Proof of Claim One

Let M be some Markov chain $\pi \to (X, Y) \to Z$. It follows that $\pi | X, Y \perp Z | X, Y$. We then have:

$$I((X, Y), \pi | R) \geq I((X, Y), \pi | R, Z)$$

Considering the right hand side of the inequality:

$$I((X, Y), \pi | R, Z) = H(X, Y | R, X) - H(X, Y | \pi, R, Z)$$

$$H(X, Y | R, Z) - \sum_{i=1}^{n} H(X_i, Y_i | R, Z, X_{<i}, Y_{<i})$$

$$= \sum_{i=1}^{n} H(X_i, Y_i | Z_i)$$

$$= \sum_{i=1}^{n} H(X_i, Y_i | R, Z)$$

$$H(X, Y | \pi, R, Z) = \sum_{i=1}^{n} H(X_i, Y_i, \pi, R, Z, X_{<i}, Y_{<i})$$

$$\leq \sum_{i=1}^{n} H(X_i, Y_I | \pi, R, Z)$$

So:

$$I((X, Y), \pi | R, Z) \geq \sum_{i=1}^{n} \Big(H(X_i, y_i | R, Z) - H(X_i, Y_i | \pi, R, Z) \Big)$$

$$I((X, Y), \pi | R, Z) \geq \sum_{i=1}^{n} I(X_i, Y_i, \pi | R, Z)$$

We now show that $I((X_i, Y_i), \pi, R, Z) \geq CIC(\text{DISJ}^1)$ by considering the following two communication protocols:

Protocol One:

Suppose shared random variable $w \sim Bern(0.5)$ and private randomness R_A, R_B for communication protocol between Alice and Bob. Alice computes random variable U and Bob V as follows:

- $U = 0$ if $w = 0$; otherwise it is random
- $V = 0$ id $w = 1$; otherwise it is random

Alice and Bob then compute $U \wedge V$. We see that the information revealed by this protocol is

$$I((X, Y), \pi | R, w)$$

.

Protocol Two:

We define the protocol π as follows. Let Alice and Bob share R, Z, and suppose they use Z to compute $X_1, ..., X_i$ and $Y_1, ..., Y_i$ respectively, according to μ. Suppose the output of the communication procedure is $\text{DISJ}^n(X, Y)$. The information revealed by Protocol Two is then:

$$I((X_i, Y_i), \pi | R, Z)$$

Reduction:

We show that we can compute the reduction of Protocol Two to a single input using Protocol One.

Let $X_i = U$, $Y_i = V$, and suppose that we generate Z and R using the shared randomness R of Protocol One.

We see that Protocol One will output $X_i \wedge Y_i$, which is equivalent to $\text{DISJ}^1(X_i, Y_i)$ and so $I((X, Y), \pi | R, w) = CIC(\text{DISJ}^1)$. We conclude that $I((X_i, Y_i), \pi | R, Z) \geq CIC(\text{DISJ}^1)$.

Using the equation:

$$I((X,Y),\pi|R,Z) \geq \sum_{i=1}^{n} I(X_i, Y_i, \pi|R,Z)$$

We see that:

$$I((X,Y),\pi|R,Z) \geq CIC(\mathrm{DISJ}^1)$$
$$I((X,Y),\pi|R,Z) \geq nCIC_\mu(\mathrm{DISJ}^1)$$

$$CIC_\mu(\mathrm{DISJ}^n) \geq nCIC_\mu(\mathrm{DISJ}^1)$$