CS 229r Information Theory in Computer Science Feb 18, 2016

Lecture 8

Lecturer: Mustazee Rahman *Scribe: Sam Xi*

In our first lectures, we proved the Loomis-Whitney inequality, which provides an upper limit on the size of a set S in d dimensions in terms of the sizes of its projection onto each dimension. In this lecture, we will show that the Loomis-Whitney inequality is stable. The proof of stability is due to Ellis et al. [1].

We begin by defining notation:

- For a set $S \subset \mathbb{Z}^d$ and $A \subset [d]$, let $\pi_{[A]}(S)$ be the projection of S onto the subspace spanned by the vectors $\{e_i, i \in A\}$.

Theorem 1 (Loomis-Whitney Inequality) *If $S \subseteq \mathbb{Z}^d$ and $\pi_{[d]\setminus i}(S)$ is the projection of S onto the plane perpendicular to the axis x_i, then*

$$|S| \leq \prod_{i=1}^{d} |\pi_{[d]\setminus i}(S)|^{1/(d-1)}$$

In the case of equality, then $S = \pi_1(S) \times ... \times \pi_d(S)$.

Intuitively, this states that the size of a set is bounded by the size of a "box" formed by the projections of the set onto lower dimensional planes. The second statement says that if $|S|^{d-1}$ is equal to the size of the product of its $(d-1)$-dimensional projections, then S itself is the product of its one-dimensional projections.

Today, our goal is to prove that the Loomis-Whitney inequality is stable. If a set S is *close* to satisfying the Loomis-Whitney upper bound, is S then also almost a product of one-dimensional sets?

More formally, suppose we have $S \subseteq \mathbb{Z}^d$ and a box $B = B_1 \times ...B_d$ where $B_i \subseteq \mathbb{Z}$. Now suppose that

$$|S| \leq (1-\epsilon) \prod_{i=1}^{d} |\pi_{[d]\setminus i}(S)|^{1/(d-1)} \tag{1}$$

where $\epsilon \geq 0$. Then does there exist a constant c_d that satisfies the following inequality:

$$|S \triangle B| \leq c_d \epsilon |S| \tag{2}$$

In other words, can I pick a point from S that will also fall within B with probability $1-\epsilon$? It turns out the answer is yes, and we will show that the value of this constant c_d is at most $64d^4$.

1 Basic notions and notation

Let S be a set as defined in equation 1 and X be a random variable on $\mathbb{Z}^d$ where $X \sim Uni(S)$ and $\text{supp}(X) = \{x \in \mathbb{Z}^d : \Pr(X = x) > 0\}$. If $X = (X_1, ..., X_d)$, then the i-th marginal of X is the distribution of the random variable X_i.

The independent coupling of X, denoted $\hat{X} = (\hat{X}_1, ..., \hat{X}_d)$ is the random variable such that $\hat{X}_i \stackrel{d}{=} X_i$ (have the same distribution) but $\hat{X}_i$ are mutually independent of each other and from X. There are two properties of their KL-divergence $D(X \| \hat{X})$ which will be important in the proof of Loomis-Whitney stability.

The first property relates the divergence between X and $\hat{X}$ with the divergence of the distributions $X_{[i]}$ of X.

$$D(X \| \hat{X}) = \sum_{i=2}^{d} D(X_{[i]} \| X_{[i-1]}; \hat{X}_i) \tag{3}$$

The second property states that the KL-divergence is simply the total mutual information between the marginals of X:

$$D(X_{[i]} \| X_{[i-1]}; \hat{X}_i) = I(X_i, X_{[i-1]}) \tag{4}$$

Mutual information between two random variables is defined as

$$I(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y)$$

and applying a function on one of them reduces the total mutual information (referred to as monotonicity of mutual information).

$$I(X, Y) \geq I(f(X), Y)$$

2 Supporting Lemmas

To prove stability, we need the following two lemmas.

Lemma 2 *Suppose $S \subseteq \mathbb{Z}^d$ satisfies Equation 1. Let $X \sim Uni(S)$ and $\hat{X}$ is the independent coupling of X. Then for all $1 \leq i \leq d$,*

$$I(X_{[d] \setminus i}, X_i) \leq 2d\epsilon$$

Definition 3 *The "hole-weight" of S is defined as:*

$$Hole(S) = \Pr(\hat{X} \notin S)$$

Lemma 4 *Suppose $S \subseteq \mathbb{Z}^d$ and $0 < \alpha < 1$. Then $\exists S_1, ..., S_d \subseteq \mathbb{Z}$ with:*

$$\Pr(X_i \notin S_i) \leq \frac{2 * \mathit{Hole}(S)}{\alpha}$$

and further, for any $x_i \in S_i$,

$$\Pr(X_i = x_i) \geq \left(1 - \frac{\mathit{Hole}(S)}{\alpha}\right)\left(\frac{1-\alpha}{|S_i|}\right)$$

So, Lemma 4 states that if Hole(S) is small, then there is a set S_i in $\mathbb{Z}$ such that X_i places most of its mass on S_i and the distribution of X_i is nearly uniform on S_i. To connect Lemma 2 with Lemma 4, we have the following claim:

Claim 5

$$\mathit{Hole}(S) \leq D(X \| \hat{X})$$

Proof By the definition of KL divergence: $D(X\|\hat{X}) = \mathbb{E}[-\log(q(\hat{X})/p(X))]$, where p and q are the probability densities of X and $\hat{X}$, respectively. Because $x \to -\log(x)$ is convex, we can apply Jensen's inequality, so we get $D(X\|\hat{X}) \geq -\log(\mathbb{E}[q(\hat{X})/p(X)])$. However, $\mathbb{E}[q(\hat{X})/p(X)] = \Pr(\hat{X} \in \text{supp}(X))$. Then we have:

$$D(X\|\hat{X}) \geq -\log(\Pr(\hat{X} \in \text{supp}(X)))$$

Since $\Pr(\hat{X} \in \text{supp}(X)) = 1 - \Pr(\hat{X} \notin \text{supp}(X))$, and $-\log(1-x) \geq x$, we have:

$$D(X\|\hat{X}) \geq \Pr(\hat{X} \notin \text{supp}(X))$$

Note that $\text{supp}(X) = S$ because $X \sim Uni(S)$. Therefore

$$\begin{aligned} \Pr(\hat{X} \notin \text{supp}(X)) &= \text{Hole}(S) \\ D(X\|\hat{X}) &\geq \text{Hole}(S) \end{aligned}$$

■

3 Proof of LW stability

Suppose Equation 1 holds. We assume $\epsilon < 64d^3$ to avoid a trivial result. Apply Lemma 2 and the monotonicity of mutual information to Equation 4 and we get:

$$D(X\|\hat{X}) \leq 2d^2\epsilon$$

Now, apply Lemma 4, which states that $\exists S_1, ..., S_d \subseteq \mathbb{Z}$ such that

$$\Pr(x_i \notin S_i) \leq \frac{2\text{Hole}(S)}{\alpha}$$

Take $\alpha = 1/d$, and apply claim 5. Then this becomes:

$$\leq 2d\text{Hole}(S) \leq 4d^3\epsilon$$

Now, let box $B = S_1 \times ... \times S_d$. Then we have:

$$\Pr(X \notin B) = \frac{|S \setminus B|}{|S|} \leq \sum_i \Pr(X_i \notin S_i) \leq 4d^4\epsilon$$
$$|S \setminus B| \leq 4d^4\epsilon|S|$$

For $x = (x_1, ..., x_d) \in B \setminus S$, if we assume $\epsilon < 64d^3$ and apply some basic simplification, we get:

$$\Pr(X_i = x_i) \geq \frac{1}{|S_i|}\left(1 - \frac{1}{d}\right)^2$$

And so, by taking the product of this for all i, we get:

$$\Pr(X = x) \geq \frac{1}{|B|}\left(1 - \frac{1}{d}\right)^{2d}$$

By summing this over all x in $|B \setminus S|$:

$$\frac{|B \setminus S|}{|B|} \leq \left(1 - \frac{1}{d}\right)^{-2d} * \sum_{x \in B \setminus S} \Pr(\hat{X} = x)$$

This sum is the probability that $\hat{X}$ is in $B \setminus S$. Observe that

$$\Pr(\hat{X} \in B \setminus S) \leq \Pr(\hat{X} \notin S)$$

And since the right hand side is the definition of Hole(S), we have:

$$\frac{|B \setminus S|}{|B|} \leq \left(1 - \frac{1}{d}\right)^{-2d} * \text{Hole}(S)$$
$$\leq 32d^3\epsilon$$

We need to establish some upper bound on $|S|$ in terms of B.

$$|S| \geq |S \cap B| = |B| - |B \setminus S| \geq (1 - 32d^3\epsilon)|B|$$

So it follows that

$$\frac{|B/S|}{|S|} \leq \frac{32d^3\epsilon}{(1-32d^3\epsilon)} \leq 64d^3\epsilon.$$

Finally, we can compute $|S \triangle B|$.

$$\begin{aligned}|S \triangle B| &\leq |S \setminus B| + |B \setminus S| \\ &= (4d^4\epsilon + 64d^3\epsilon)|S| \\ &\leq 64d^4\epsilon|S|\end{aligned}$$

for $\epsilon < 64d^3$.

4 Proof of Lemma 2

Let $X \sim Uni(S)$. Without loss of generality, we may prove the lemma for the case $i = d$. If Equation 1 holds, then:

$$\log|S| \geq \frac{1}{d-1}\sum_{j=1}^{d} \log|\pi_{[d]\setminus j}(s)| + \log(1-\epsilon) \tag{5}$$

Since $\mathcal{H}(X) = \log|S|$ (because $X \sim Uni(S)$) and $\mathcal{H}(X_{[d]\setminus j]}) \leq \log|\pi_{[d]\setminus j]}(S)|$, this expression becomes

$$\frac{1}{d-1}\sum_j \mathcal{H}(X_{[d]\setminus j]}) - \mathcal{H}(X) \leq -\log(1-\epsilon) \tag{6}$$

Apply Taylor expansion on the righthand side.

$$-\log(1-x) = \sum_k \frac{x^k}{k} \leq 2x \text{ for } 0 \leq x \leq \frac{1}{2} \tag{7}$$

$$\frac{1}{d-1}\sum_j \mathcal{H}(X_{[d]\setminus j]}) - \mathcal{H}(X) \leq 2\epsilon \text{ for } \epsilon < 1/2 \tag{8}$$

We need to understand this difference on the lefthand side. Let's examine the summand. Applying the chain rule of entropy gives:

$$\mathcal{H}(X_{[d]\setminus j]}) = \mathcal{H}(X_d) + \mathcal{H}(X_{[d-1]\setminus j}|X_d), j \neq d \tag{9}$$

Now the lefthand side of this expression becomes:

$$\begin{aligned}
&\mathcal{H}(X_d) + \frac{\mathcal{H}(X_{[d-1]})}{d-1} + \frac{1}{d-1}\sum_{j<d} \mathcal{H}(X_{[d-1]\setminus j}|X_d) - H(X) \\
&\quad = \frac{\mathcal{H}(X_{[d-1]})}{d-1} + \frac{1}{d-1}\sum_{j<d} \mathcal{H}(X_{[d-1]\setminus j}|X_d) - \mathcal{H}(X_{[d-1]}|X_d) \qquad (10)
\end{aligned}$$

Once again, let's look at the summand here:

$$\mathcal{H}(X_{[d-1]\setminus j}|X_d) = \sum_{k<d, k\neq j} \mathcal{H}(X_k|X_{[k-1]\setminus j}; X_d) \qquad (11)$$

Because conditional entropy decreases when conditioning over a larger set of random variables:

$$\mathcal{H}(X_k|X_{[k-1]\setminus j}, X_d) \geq \mathcal{H}(X_k|X_{[k-1]}, X_d) \qquad (12)$$

So, substituting this into (11):

$$\begin{aligned}
\mathcal{H}(X_{[d-1]\setminus j]}|X_d) &\geq \sum_k \mathcal{H}(X_k|X_{[k-1]}; X_d) \\
&= \left(\sum_{k=1}^{d-1} \mathcal{H}(X_k|X_{[k-1]}, X_d)\right) - \mathcal{H}(X_j|X_{[j-1]}, X_d) \qquad (13) \\
&= \mathcal{H}(X_{[d-1]}|X_d) - H(X_j|X_{[j-1]}, X_d) \qquad (14)
\end{aligned}$$

This expression was part of a summation over j, so after summing over j, we conclude that:

$$\begin{aligned}
\sum_{j<d} \mathcal{H}(X_{[d-1]\setminus j}|X_d) &\geq (d-1)\mathcal{H}(X_{[d-1]}|X_d) - \sum_{j<d} \mathcal{H}(X_j|X_{[j-1]}, X_d) \\
&= (d-1)\mathcal{H}(X_{[d-1]}|X_d) - \mathcal{H}(X_{[d-1]}|X_d) \\
&= (d-2)\mathcal{H}(X_{[d-1]}|X_d) \qquad (15)
\end{aligned}$$

Now we substitute this back into (10), and we find:

$$\begin{aligned}
(10) &\geq \frac{\mathcal{H}(X_{[d-1]})}{d-1} - \mathcal{H}(X_{[d-1]}|X_d) + \frac{d-2}{d-1}\mathcal{H}(X_{[d-1]}|X_d) \\
&= \frac{\mathcal{H}(X_{[d-1]}) - \mathcal{H}(X_{[d-1]}|X_d)}{d-1} \\
&= \frac{I(X_{[d-1]}, X_d)}{d-1} \qquad (16)
\end{aligned}$$

So, from (8) and (16), we finally get:

$$\begin{aligned}
I(X_{[d-1]}, X_d) &\leq 2\epsilon(d-1) \\
&\leq 2d\epsilon \qquad (17)
\end{aligned}$$

References

[1] David Ellis and Ehud Friedgut and Guy Kindler and Amir Yehudayoff, *Geometric stability via information theory*, arXiv:1510.00258, 2015.