Lecture 12: Talagrand Inequality and Applications

Overview

- Today we shall see (without proof) a concentration inequality called the "Talagrand Inequality"
- This result shall help us prove the concentration of a large class of problems around its median
- As an application, we shall see a concentration result for the longest increasing subsequence

- Recall the definition of the Hamming distance between two elements $x, y \in \Omega := \Omega_1 \times \cdots \times \Omega_n$

$$\left|\{i \colon 1 \leqslant i \leqslant n \text{ and } x_i \neq y_i\}\right|$$

- Intuitively, the strings get penalized "1" for every index i where x_i and y_i are different
- We can consider a weighted variant of this distance where every index i has its own associated penalty α_i
- Before we proceed to develop this new notion of distance, let us first normalize the Hamming distance. Consider the following redefinition. Let $\alpha = (\alpha_1, \ldots, \alpha_n) = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$. We define

$$d_H(x, y) = \sum_{1 \leqslant i \leqslant n \colon x_i \neq y_i} \alpha_i$$

- For the sake of completeness, we write down the inequality that we saw on Hamming distance in this new form

$$\mathbb{P}\left[\mathbb{X} \in A\right] \cdot \mathbb{P}\left[d_H(\mathbb{X}, A) \geqslant E\right] \leqslant \exp(-E^2/2)$$

- Now, we are at a position to generalize the notion of distance to any vector α with norm 1. That is, consider $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that
  - $\alpha_1, \ldots, \alpha_n \geqslant 0$, and
  - $\sum_{i=1}^{n} \alpha_i^2 = 1$.
- We define the following distance between $x, y \in \Omega$ with respect to α as follows

$$d_\alpha(x, y) := \sum_{1 \leqslant i \leqslant n\colon\ x_i \neq y_i} \alpha_i$$

Intuitively, this captures the fact that every coordinate i could be penalized differently compared to other coordinates.

- Now, for a pair x, y we consider the "most severe penalty."

Definition (Convex Distance)

For $x, y \in \Omega$, we define the convex distance between x and y as follows

$$d_T(x,y) := \sup_{\alpha:\ \|\alpha\|_2=1} d_\alpha(x,y)$$

- Similar to the case of Hamming distance, we can define the distance of $x \in \Omega$ from a set $A \subseteq \Omega$

$$d_T(x,A) = \min_{y\in A} d_T(a,y)$$

So, if $d_T(x,A) \geqslant t$, then we have $d_T(x,y) \geqslant t$, for all $y \in A$.

Talagrand Inequality

- Let $\mathbb{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_n)$ be a random variable over Ω, such that each $\mathbb{X}_i$ is independent of the others and $\mathbb{X}_i \in \Omega_i$
- Let $f : \Omega \to \mathbb{R}$
- Talagrand inequality states that if any $A \subseteq \Omega$ is dense, then it is unlikely that $\mathbb{X}$ is far (w.r.t. the $d_T(\cdot, \cdot)$ distance) from A

Theorem (Talagrand Inequality)

For any $A \subseteq \Omega$, we have

$$\mathbb{P}\left[\mathbb{X} \in A\right] \cdot \mathbb{P}\left[d_T(\mathbb{X}, A) \geqslant E\right] \leqslant \exp(-E^2/4)$$

Application: Longest Increasing Subsequence

- Let us first formulate the longest increasing subsequence problem. Suppose $\mathbb{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_n)$, where each $\mathbb{X}_i$ is independent and uniformly distribution over $\Omega_i = [0, 1)$
- We are interested in $f(\mathbb{X})$, the length of the longest increasing subsequence in $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$
- Let us try to understand the expected value $\mathbb{E}\left[f(\mathbb{X})\right]$ and its concentration that we can conclude from the previous tools that we have studied
- Note that f is $(1, 1, \ldots, 1)$ bounded difference function, because changing one entry in $\mathbb{X}$ can change the longest increasing subsequence by at most 1. So, we can apply the independent bounded difference inequality to conclude the following

$$\mathbb{P}\left[f(\mathbb{X}) \geqslant \mathbb{E}\left[f(\mathbb{X})\right] + E\right] \leqslant \exp(-2E^2/n)$$

Note that the radius of concentration that we obtain from the inequality is (roughly) $\sqrt{n}$

- Although, this result is non-trivial, it is useless. Because we have $\mathbb{E}\left[f(\mathbb{X})\right] = \Theta(\sqrt{n})$. Students are highly encouraged to prove this result
- Our objective is to use the Talagrand inequality to prove a concentration of $f(\mathbb{X})$ around its median m with radius of concentration $\sqrt{m}$. Note that by the Markov inequality, we have $m \leqslant 2\mathbb{E}\left[f(\mathbb{X})\right]$, hence, m and $\mathbb{E}\left[f(\mathbb{X})\right]$ have the same order. Therefore, the radius of concentration is $\Theta(n^{1/4})$. Now, this result is useful

We aim to get a concentration inequality of $f(\mathbb{X})$.

- Define $B_a = \{y \colon y \in \Omega \text{ and } f(y) \leqslant a\}$
- Suppose we prove the following claim

Claim (A Technical Claim)

$$\mathbb{P}\left[f(\mathbb{X}) \leqslant a\right] \cdot \mathbb{P}\left[f(\mathbb{X}) \geqslant a + E\right] \leqslant \mathbb{P}\left[\mathbb{X} \in B_a\right] \cdot \mathbb{P}\left[d_T(\mathbb{X}, B_a) \geqslant \frac{E}{\sqrt{a+E}}\right].$$

- Using this technical claim, let us get our concentration inequalities for the distribution $f(\mathbb{X})$
- Note that Talagrand inequality applies to the right-hand side of the claim. Therefore, we get

$$\mathbb{P}\left[f(\mathbb{X}) \leqslant a\right] \cdot \mathbb{P}\left[f(\mathbb{X}) \geqslant a + E\right] \leqslant \mathbb{P}\left[\mathbb{X} \in B_a\right] \cdot \mathbb{P}\left[d_T(\mathbb{X}, B_a) \geqslant \frac{E}{\sqrt{a+E}}\right]$$
$$\leqslant \exp\left(-\frac{E^2}{4(a+E)}\right).$$

- **Bounding the upper tail.** Set $a = m$, the median of the distribution $f(\mathbb{X})$. Then, we have $\mathbb{P}\left[f(\mathbb{X}) \leqslant a\right] = \mathbb{P}\left[f(\mathbb{X}) \leqslant m\right] \geqslant 1/2$. Next, using the inequality, we get

$$\mathbb{P}\left[f(\mathbb{X}) \geqslant m + E\right] \leqslant \frac{\exp\left(-\frac{E^2}{4(m+E)}\right)}{\mathbb{P}\left[f(\mathbb{X}) \leqslant m\right]} \leqslant 2\exp\left(-\frac{E^2}{4(m+E)}\right).$$

- **Bounding the lower tail.** Set $a + E = m$, the median of the distribution $f(\mathbb{X})$. Then, we have $\mathbb{P}\left[f(\mathbb{X}) \geqslant a + E\right] = \mathbb{P}\left[f(\mathbb{X}) \geqslant m\right] \geqslant 1/2$. Next, using the inequality, we get

$$\begin{aligned}\mathbb{P}\left[f(\mathbb{X}) \leqslant a\right] &= \mathbb{P}\left[f(\mathbb{X}) \leqslant m - E\right] \\ &\leqslant \frac{\exp\left(-\frac{E^2}{4m}\right)}{\mathbb{P}\left[f(\mathbb{X}) \geqslant m\right]} \\ &\leqslant 2\exp\left(-\frac{E^2}{4m}\right).\end{aligned}$$

- Therefore, all that remains is to prove the technical claim.
- **Remark.** We did not use any "special property" of the function $f(\cdot)$. For a particular function $f(\cdot)$, if we can prove the technical claim, then we are done!

- **Remark.** This concentration is around the median (*not the mean*). However, by Markov inequality, we know that the median cannot be much larger than the mean.

In this part of the lecture, we will prove the technical claim for the particular function $f(\cdot)$ that outputs the length of the longest subsequence of its input bitstring

Proof outline.

- Recall that we need to prove

$$\mathbb{P}\left[f(\mathbb{X}) \leqslant a\right] \cdot \mathbb{P}\left[f(\mathbb{X}) \geqslant a+E\right] \leqslant \mathbb{P}\left[\mathbb{X} \in B_a\right] \cdot \mathbb{P}\left[d_T(\mathbb{X}, B_a) \geqslant \frac{E}{\sqrt{a+E}}\right].$$

- By definition, the event "$f(\mathbb{X}) \leqslant a$" is equivalent to the event "$\mathbb{X} \in B_a$." Therefore, proving the technical claim is equivalent to proving the inequality

$$\mathbb{P}\left[f(\mathbb{X}) \geqslant a+E\right] \leqslant \mathbb{P}\left[d_T(\mathbb{X}, B_a) \geqslant \frac{E}{\sqrt{a+E}}\right]$$

- Observe that if an event $\mathcal{A}$ implies an event $\mathcal{B}$, then $\mathbb{P}[\mathcal{A}] \leqslant \mathbb{P}[\mathcal{B}]$. Therefore, it suffices to prove that the event "$f(\mathbb{X}) \geqslant a + E$" implies the event "$d_T(\mathbb{X}, B_a) \geqslant \frac{E}{\sqrt{a+E}}$"
- In the rest of the lecture, we prove this implication

Proof.

- Suppose $\mathbb{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_n)$, where each $\mathbb{X}_i$ is independent and uniformly distributed over $\Omega_i = [0, 1)$
- We are interested in demonstrating a concentration bound for $f(\mathbb{X})$, where $f(\mathbb{X})$ is the longest increasing subsequence in $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$
- **Observation.** Consider any $x \in \Omega := \Omega_1 \times \cdots \times \Omega_n$. If $f(x) = k$ (i.e., the longest increased subsequence in x is k), then there is a set $K_x = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ such that K_x denotes the indices of the longest increasing subsequence in x

- **Observation.** Consider any $y \in \Omega$. Note that if y agrees with x at all the indices in K_x, then we have $f(y) \geqslant f(x)$ (it is possible that y has a longest increasing subsequence, but, definitely, it will not be shorter than the length of the longest increasing subsequence in x)
- **Observation.** Let us generalize the previous observation further. Consider any $y \in \Omega$. Note that if y agrees with x at all indices in K_x except at ℓ indices. Then, we have $f(y) \geqslant f(x) - \ell$. Formally, we can write this as follows

$$f(y) \geqslant f(x) - \left|\{i \colon i \in K_x \text{ and } x_i \neq y_i\}\right|$$

- Intuitively, we incur a penalty for every $i \in K_x$ where x and y differ. Let us fix $\alpha_x = (\alpha_1, \ldots, \alpha_n)$ such that

$$\alpha_i = \begin{cases} 0 & i \notin K_x \\ \frac{1}{\sqrt{|K_x|}} & i \in K_x \end{cases}$$

Note that $|K_x| = f(x)$. So, we conclude that

$$f(y) \geqslant f(x) - \sqrt{f(x)} d_{\alpha_x}(x, y)$$

- Rearranging, we get that

$$d_{\alpha_x}(x, y) \geqslant \frac{f(x) - f(y)}{\sqrt{f(x)}}$$

- Since, $d_T(\cdot,\cdot)$ is a supremum of $d_\alpha(\cdot,\cdot)$ over all α with norm 1, we get that
$$d_T(x,y) \geqslant \frac{f(x)-f(y)}{\sqrt{f(x)}}$$
- Define $B_a = \{y\colon y \in \Omega \text{ and } f(y) \leqslant a\}$. So, for all $y \in B_a$, we have $f(y) \leqslant a$. Therefore, for any $y \in B_a$, we get
$$d_T(x,y) \geqslant \frac{f(x)-a}{\sqrt{f(x)}}$$
- Since, the inequality holds for all $y \in B_a$, we conclude that
$$d_T(x,B_a) \geqslant \frac{f(x)-a}{\sqrt{f(x)}}$$

- **Observation.** If $f(x) \geqslant a + E$, then

$$d_T(x, B_a) \geqslant \frac{E}{\sqrt{a+E}}$$

- This observation concludes the proof of the technical claim.

Configuration Function

- The approach of applying the Talagrand inequality to the problem of longest increasing subsequence can be generalized to several problems.
- Consider the definition of c-configuration functions

Definition (Configuration Functions)

A function f is a c-configuration function, if for every x, y, there exists $\alpha_{x,y}$ such that the following holds

$$f(y) \geqslant f(x) - \sqrt{c \cdot f(x)}\, d_{\alpha_{x,y}}(x, y)$$

- Note that the longest increasing subsequence defines $f(\cdot)$ that is 1-configuration function. The derivation used above can be identically used for c-configuration functions