Lecture 11: Talagrand Inequality and Applications

Overview

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

Convex Distance I

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

$$\Big|\Big\{i \in [n] \colon x_i \neq y_i\Big\}\Big|$$

- Intuitively, we count "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 weight α_i
- Before, we proceed to developing 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

$$d_H(x, y) = \sum_{i \in [n] \colon x_i \neq y_i} \alpha_i$$

Convex Distance II

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

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

- Now, we 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

$$d_\alpha(x, y) := \sum_{i \in [n]:\ x_i \neq y_i} \alpha_i$$

Convex Distance III

- Now, for a pair x, y, we can consider the "worst direction" α that witnesses the highest distance

Definition (Convex Distance)

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

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

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

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

So, $d_T(x, A) \geqslant t$ implies that $d_T(x, y) \geqslant t$, for all $y \in A$. Further,

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
- Let $f : \Omega \to \mathbb{R}$
- Talagrand Inequality states the following

Theorem (Talagrand Inequality)

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

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

Application to Longest Increasing Subsequence I

- Suppose $\mathbb{X} = (\mathbb{X}_1, \dots \mathbb{X}_n)$, where each $\mathbb{X}_i$ is independent and uniformly distributed over $\Omega_i = [0, 1]$
- We are interested in $f(\mathbb{X})$, the length of the longest increasing subsequence in $(\mathbb{X}_1, \dots, \mathbb{X}_n)$
- **Observation.** Consider any $x \in \Omega$. If $f(x) = k$, then there is a set $K_x = \{i_1, \dots, i_k\} \subseteq [n]$ 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 longer increasing subsequence, but, definitely, it will not be shorter than the length of the longest increasing subsequence of x)

Application to Longest Increasing Subsequence II

- **Observation.** Consider any $y \in \Omega$. Note that if y agrees with x at all the indices in K_x except at ℓ indices, then we have $f(y) \geqslant f(x) - \ell$. Formally, we can write this as follows

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

- Let us fix $\alpha_x = (\alpha_1, \ldots, \alpha_n)$ such that

$$\alpha_i = \begin{cases} \frac{1}{\sqrt{K_x}}, & \text{if } i \in K_x \\ 0, & \text{otherwise.} \end{cases}$$

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

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

Application to Longest Increasing Subsequence III

- 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 $A_a = \{y \colon f(y) \leqslant a\}$. So, for all $y \in A_a$, we get

$$d_T(x, y) \geqslant \frac{f(x) - a}{\sqrt{f(x)}}$$

- Since, the inequality holds for all $y \in A_a$, we can conclude that

$$d_T(x, A_a) \geqslant \frac{f(x) - a}{\sqrt{f(x)}}$$

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

$$d_T(x, A_a) \geqslant \frac{t}{\sqrt{a+t}}$$

- So, we have

$$\mathbb{P}\left[f(\mathbb{X}) \geqslant a + t\right] \leqslant \mathbb{P}\left[d_t(\mathbb{X}, A_a) \geqslant \frac{t}{\sqrt{a+t}}\right]$$

- Multiplying both sides by $\mathbb{P}[\mathbb{X} \in A_a]$, we get

$$\begin{aligned}\mathbb{P}[\mathbb{X} \in A_a]\,\mathbb{P}\left[f(\mathbb{X}) \geqslant a + t\right] &\leqslant \mathbb{P}[\mathbb{X} \in A_a]\,\mathbb{P}\left[d_t(\mathbb{X}, A_a) \geqslant \frac{t}{\sqrt{a+t}}\right] \\ &\leqslant \exp\left(-\frac{t^2}{4(a+t)}\right)\end{aligned}$$

- Let m be the median of the random variable $f(\mathbb{X})$.

Application to Longest Increasing Subsequence V

- Suppose we use $a = m$. Then, we have $\mathbb{P}\left[\mathbb{X} \in A_a\right] \geqslant 1/2$. Therefore, we conclude that

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

- Suppose we use $a + t = m$. Then, we have $\mathbb{P}\left[f(\mathbb{X}) \geqslant a + t\right] \geqslant 1/2$. Then, we have

$$\mathbb{P}\left[\mathbb{X} \in A_a\right] = \mathbb{P}\left[f(\mathbb{X}) \leqslant m - t\right] \leqslant 2\exp\left(-\frac{t^2}{4m}\right)$$

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