Lecture 15: Talagrand Inequality

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, in the next lecture, we shall see a concentration result for the longest increasing subsequence

Convex Distance I

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

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

- Intuitively, we 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 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

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

Convex Distance II

- 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 possibly be penalized differently as compared to other coordinates.

Convex Distance III

- 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

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

Application: Longest Increasing Subsequence I

- 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

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

Application: Longest Increasing Subsequence II

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