Lecture 08: Applications of Talagrand Inequality
Convex Distance

- Last lecture we considered $d_H(x, y) := |\{i: x_i \neq y_i\}|$
- We generalize this notion of distance to $\alpha \in \mathbb{R}^n$ such that each of its coordinates are $\geq 0$ and $\|\alpha\| = 1$, i.e., $\sum_{i=1}^n \alpha_i^2 = 1$

**Definition**

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

- We define the convex distance as

**Definition (Convex Distance)**

$$d_{\mathcal{T}}(x, y) := \sup_{\alpha: \|\alpha\|=1} \sum_{1 \leq i \leq n, x_i \neq y_i} \alpha_i$$
We will not prove this inequality. We simply state it and shall use it to show the concentrated behavior of the longest increasing subsequence.

Theorem (Talagrand Inequality)

\[ \mathbb{P} [X \in A] \cdot \mathbb{P} \left[ d_T (X, A) \geq t \right] \leq \exp \left( -\frac{t^2}{4} \right) \]
Suppose $x = (x_1, \ldots, x_n)$ and each coordinate has been sampled independently and uniformly at random over $\mathbb{R}$.

Let $f(x)$ represent the longest increasing subsequence of $x$.

Suppose $f(x) = k$.

Note that there exists $K_x \subseteq \{1, \ldots, n\}$ such that the longest increasing subsequence is formed by

$$(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$$
Note that if $y$ matches $x$ everywhere in $K_x$, then we must have $f(y) \geq f(x)$

Similarly, if $y$ matches $x$ everywhere in $K_x$ except at $\ell$ positions, then $f(y) \geq f(x) - \ell$

In particular, we can write the following bound:

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

Let $\alpha$ be a vector such that $\alpha_i = 0$ if $i \notin K_x$, otherwise $\alpha_i = 1/\sqrt{f(x)}$. Note that $\|\alpha\| = 1$ and the above inequality can equivalently be written as:

$$f(y) \geq f(x) - \sqrt{f(x)}d_\alpha(x, y)$$
Longest Increasing Subsequence

Rearranging, we get $f(x) \leq f(y) + \sqrt{f(x)}d_\alpha(x, y)$

By the definition of $d_T(\cdot, \cdot)$, we can also write that:

$$f(x) \leq f(y) + \sqrt{f(x)}d_T(x, y)$$

Let $A_a = \{y : f(y) \leq a\}$

Consider $y \in A_a$. The above inequality becomes:

$$f(x) \leq a + \sqrt{f(x)}d_T(x, y)$$

Rearranging, we get:

$$d_T(x, y) \geq \frac{f(x) - a}{\sqrt{f(x)}}$$
Suppose $f(x) \geq a + t$. Then, it implies that

$$d_T(x, y) \geq \frac{t}{\sqrt{a + t}}$$

Therefore, we can conclude that

$$\mathbb{P}[f(X) \geq a + t] \leq \mathbb{P}\left[d_T(X, y) \geq \frac{t}{\sqrt{a + t}}\right]$$

Since, this statement is true for all $y \in A_a$, the following statement is also true

$$\mathbb{P}[f(X) \geq a + t] \leq \mathbb{P}\left[d_T(X, A_a) \geq \frac{t}{\sqrt{a + t}}\right]$$
Multiplying both sides by $\mathbb{P}[X \in A_a]$, we get

$$\mathbb{P}[X \in A_a] \cdot \mathbb{P}[f(X) \geq a + t] \leq \mathbb{P}[X \in A_a] \cdot \mathbb{P}\left[d_T(X, A_a) \geq \frac{t}{\sqrt{a + t}}\right] \leq \exp\left(-\frac{t^2}{4(a + t)}\right)$$
Suppose we choose $a = m$, where $m$ is the median of the distribution $f(X)$. So, we have $\mathbb{P}[f(X) \leq m] \geq 1/2$. We get:

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

Suppose we choose $a = m - t$. Then $\mathbb{P}[f(X) \geq a + t] \geq 1/2$. Now we have:

$$\mathbb{P}[X \in A_a] = \mathbb{P}[f(X) \leq m - t] \leq 2 \exp \left( -\frac{t^2}{4m} \right)$$
A function $f$ is a $c$-configuration function, if for every $x, y$, there exists $\alpha$ such that the following holds.

$$f(y) \geq f(x) - \sqrt{c \cdot f(x)d\alpha(x,y)}$$

Note that the longest increasing subsequence function is an $1$-configuration function. The derivation of the previous concentration bound on the longest increasing subsequence naturally generalizes to $c$-configuration functions.