Chapter 0 and 2 of HDP

Soumendu Mukherjee (notes by Nick Bhattacharya)

June 27, 2017

1 Chapter 0

First we recall Caratheodory's Theorem.

Caratheodory Theorem Let $T \subset \mathbb{R}^n$ be any subset and take a point x in the convex hull of T. Then there are $y_1, \cdots y_m$ in T such that

$$x = \sum_i \lambda_i y_i \text{ with } \sum_i \lambda_i = 1$$

Further m always satisfies $m \leq n+1$.

Vershynin motivates the book by pointing out that a remarkable "Approximate Caratheorodory Theorem" holds.

Approximate Caratheodory Theorem Let $T \subset \mathbb{R}^n$ be a set of bounded diameter and x be in the convex hull of T. For any $k \in \mathbb{N}$, we can find points $x_1, \cdots x_k$ such that

$$\|x - \frac{1}{k}\sum_{j=1}^{k} x_i\|_2 \leq \frac{1}{\sqrt{k}}$$

This is remarkable since it is independent of dimension n! The proof uses the probabilistic method and can be found in the book.

Application We now observe that this theorem lets us get an upper bound on the metric entropy of a polyedral set in $\mathbb{R}^n$. First some review.

"Review" of Metric Entropy: For any subset K in $\mathbb{R}^n$, recall that the covering number of K is given by

$$\text{cov}(\varepsilon, K) = \text{ the minimum number of balls of radius } \varepsilon \text{ that are needed to cover } K.$$

The log of this quantity is roughly what we call metric entropy.

The Application: Let P be a polyhedral set with N vertices. We may now use the theorem above, to get the bound

$$\text{cov}(\varepsilon, P) \leq N^{\lceil \frac{1}{\varepsilon^2} \rceil}$$

2 Chapter 2

We now move to Concentration Inequalities. The basic intuition follows from a simple example.

Example Let X_i be iid Bernoulli 0-1 random variables. Take S_N to be the number of head sin N tosses, $\sum_{i=1}^{N} X_i$. Using Chebyshev we get

$$\mathbb{P}\Big[S_N \geq \frac{3}{4}\Big] \leq \frac{4}{N}$$

However, the CLT suggests this tail should look like a Gaussian tail which goes like e^{-cN}. Can this argument be made rigorous? No, because the error in CLT decays like $\sqrt{N}$ -

$$\Big|\mathbb{P}[S_N^* \geq t] - \mathbb{P}[Z \geq t]\Big| \leq \frac{C}{\sqrt{N}}$$

The intuition *is* correct though and this is formalized in *Hoeffding's Inequality*: For any positive a_i,

$$\mathbb{P}\Big[\sum_i a_i X_i \geq t\Big] \leq \exp\Big(\frac{-t^2}{2\|a\|_2^2}\Big)$$

General Setup We discussed some large classes of random variables for which bounds like these exist. They are

Sub-Gaussian A random variable is sub-Gaussian with parameter σ if for all λ real

$$\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\lambda^2\sigma^2/2}.$$

Sub-Exponential A random variable is sub-exponential with parameters (σ, α) if for all λ with $|\lambda| \leq \alpha$

$$\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\lambda^2\sigma^2/2}$$

Concentration inequalities hold for these classes of random variables as well. For example

Theorem Let X_i be independent sub-Gaussian with parameter σ_i. Then for $S_n = \sum_{i=1}^N X_i$,

$$\mathbb{P}\Big[S_N - \mathbb{E}S_N \geq t] \leq \exp\left(\frac{-ct^2}{\sum_{i=1}^N \sigma_i^2}\right)$$

In broad strokes, we start with $\{\sum_i X_i > t\}$ and replace that with $\{e^{\sum_i X_i} > e^t\}$. Now we apply Markov's inequality and exploit independence + the **assumed** estimates to get the inequality. The sharpness of one's constants depends on how hard you work in bounding the generating function (as always).