Lecture 4. Applying concentration inequalities

Recap

- Finished up MD functionals
- Tail bounds $\longleftrightarrow$ resilience

Concentration inequalities

- Markov, Chebyshev, Chernoff, Rosenthal inequality (p^{th} moments)

Today:

Use concentration to analyze:

- max of independent variables $\leftarrow$ finite maximum
- maximum eigenvalue of random matrix

$\sup_{\|v\|_2 \le 1} v^T M v \leftarrow$ infinite maximum

- VC dimension
  - $\hookrightarrow \mathcal{H} \rightarrow vc(\mathcal{H}) = d \rightsquigarrow \sqrt{\frac{d}{n}}$ for avg.'s of n samples
  - $\hookrightarrow$ Symmetrization: "bringing your own randomness"

Warm-up Let $X_1, \dots, X_n$ are all independent zero-mean random variables, each sub-Gaussian w/ parameter σ.

Question. How large can $\max_{j=1}^{n} X_j$ get?

$\mathbb{P}[A \cup B] \le \mathbb{P}[A] + \mathbb{P}[B]$

$\sqrt{\log(n)}$

union bound

$\rightarrow \mathbb{P}\left[\max_{j=1}^{n} X_j \geq t\right]$

$= \mathbb{P}[X_1 \geq t \text{ or } X_2 \geq t \text{ or} \dots \text{ or } X_n \geq t]$

$\leq \sum_j \underbrace{\mathbb{P}[X_j \geq t]}_{\exp\left(-\frac{t^2}{2\sigma^2}\right)}$

$\leq \underbrace{n \exp\left(-\frac{t^2}{2\sigma^2}\right)}_{\text{Gaussian decay}} \ll 1/\delta$

$\exp\left(-\frac{t^2}{2\sigma^2}\right) \ll \frac{1}{n}$

$\frac{t^2}{2\sigma^2} \gg \log(n)$

$\rightarrow t \gg \sigma\sqrt{2(\log(n) + \log(1/\delta))}$

$t \geq \sigma\sqrt{2\log(n)} \rightarrow \text{prob.} \leq 1$

Bounded 2nd moment: $\max_{j=1}^{n} X_j \approx \sqrt{n}$

4th moment: $\max_{j=1}^{n} X_j \approx \sqrt[4]{n}$

⋮

etc.

$\mathbb{P}[|X| \geq t] \leq \frac{1}{t^k} \Rightarrow p^{th}$ moments for all $p < k$

Max eigenvalue of random matrix.

$\rightarrow X_1, \dots, X_n \sim \rho$ (sub-Gaussian w/ σ)

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T$$

$$\|M\|_{op} = \sup_{\|v\|_2 \le 1} v^T M v$$

Claim. $\|M\|_{op} \le O\left(\sqrt{\frac{d+\log(1/\delta)}{n}}\right)$ w.p. $1-\delta$.

$X \in \mathbb{R}^d$ is sub-Gaussian if $\langle X, v\rangle$ is sub-Gaussian for all unit vectors v

$$\mathbb{E}[\exp(\lambda\langle v, X\rangle)] \le \exp\left(\frac{\lambda^2\|v\|_2^2\sigma^2}{2}\right).$$

$$v^T M v = \frac{1}{n}\sum_{i=1}^{n} v^T(X_i X_i^T)v$$

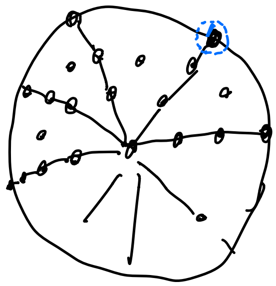
$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\langle X_i, v\rangle^2}_{\text{avg. of independent r.v.'s}}$$

$\langle X_i, v\rangle$: sub-Gaussian

$\langle X_i, v\rangle^2$: sub-Exponential

Idea. Approx. sup w/ finite max

$N_{1/4}$: maximal packing of radius $1/4$ $\rightarrow$ max set of points $v_1, \dots, v_m$ s.t. $\|v_i - v_j\|_2 \ge \frac{1}{4}$ $\forall i \ne j$.



① Bound $|N_{1/4}|$

Pf. Take balls of radius $1/8$ around each v_j.

- Balls disjoint.
- All contained in ball of radius $9/8$.

$\Rightarrow$ volume argument: $|N_{1/4}| \leq 9^d$

② Show $N_{1/4}$ approximates max

Claim. $\max_{u \in N_{1/4}} u^T M u \geq \frac{1}{2} \max_{\|v\|_2 = 1} v^T M v$

$$u \longrightarrow v$$

$$\begin{aligned} |\underbrace{v^T M v}_{\|M\|_{op}} - u^T M u| &= |v^T M (v-u) + u^T M (v-u)| \\ &\leq (\|v\|_2 + \|u\|_2) \|M(v-u)\|_2 \\ &\leq \underbrace{(\|v\|_2 + \|u\|_2)}_{\leq 2} \cdot \|M\|_{op} \cdot \underbrace{\|v-u\|_2}_{\leq 1/4} \\ &\leq \frac{1}{2} \|M\|_{op} \end{aligned}$$

$\Rightarrow u^T M u \geq \frac{1}{2} \|M\|_{op}$

$$\sup_{\|v\|_2 \le 1} \frac{1}{n}\sum_{i=1}^{n} \langle x_i, v\rangle^2$$

sub-exponential

$$\le 2 \max_{v \in N_{1/4}} \frac{1}{n}\sum_{i=1}^{n} \langle x_i, v\rangle^2$$

$$\to \frac{1}{n}\sum_{i=1}^{n} \underbrace{\|x_i\|_2^2}_{\approx\sqrt{d}} = \frac{1}{n}\sum_{i=1}^{n} d = d$$

$$\|\mathbb{E}[xx^T]\| = 1$$

$$\underbrace{O(\sigma^2)}+\underbrace{O\left(\sigma^2 \frac{d+\log(1/\delta)}{n}\right)}$$

$$\mathbb{E}\left[\underbrace{\exp\left(\frac{n}{\sigma^2}\cdot v^T M v\right)}\right]$$

$$= \mathbb{E}\left[\exp\left(\frac{\cancel{n}}{\sigma^2}\cdot\frac{1}{\cancel{n}}\sum_{j=1}^{n} |\langle x_i, v\rangle|^2\right)\right]$$

$$= \prod_{j=1}^{n} \exp\left(\frac{|\langle x_j, v\rangle|^2}{\sigma^2}\right) \le 2$$

$$\le \prod_{j=1}^{n} 2 = 2^n.$$

$$\mathbb{E}[\exp(\lambda X)] \le \exp\left(\frac{\lambda^2\sigma_A^2}{2}\right)$$

$$\mathbb{E}\left[\exp\left(\frac{X^2}{\sigma_B^2}\right)\right] \le 2$$

$$\sigma_A \le 10\cdot\sigma_B$$

$$\sigma_B \le 10\cdot\sigma_A$$

$$\boxed{\mathbb{P}\left[v^T M v \geq t\right]}$$

$$= \mathbb{P}\left[\exp\left(\frac{n}{\sigma^2}\cdot v^T M v\right) \geq \exp\left(\frac{n}{\sigma^2}\cdot t\right)\right]$$

$$\leq \boxed{2^n \cdot \exp\left(-\frac{n}{\sigma^2}\cdot t\right).}$$

$$\mathbb{P}\left[\max_{v\in N_{1/4}} v^T M v \geq t\right]$$

$$\leq 2^n \cdot |N_{1/4}| \cdot \exp\left(-\frac{n}{\sigma^2} t\right)$$

$$= \underbrace{2^n \cdot 9^d \cdot \exp\left(-\frac{n}{\sigma^2} t\right)}_{} \leq \delta$$

$$\to \exp\left(-\frac{n}{\sigma^2} t\right) \leq \frac{\delta}{2^n \cdot 9^d}.$$

$$\frac{n}{\sigma^2} t \geq \log\left(2^n \cdot 9^d \cdot \delta\right) = n\log(2) + d\log(9) + \log(1/\delta)$$

$$t = \sigma^2\left(\frac{n\log(2) + d\log(9) + \log(1/\delta)}{n}\right)$$

$$= O\left(\sigma^2\left(1 + \frac{d + \log(1/\delta)}{n}\right)\right).$$

Bounded $\|M\|_{op}$ via tail bound + $\underbrace{N_{1/4}}_{\text{"}\varepsilon\text{-net"}}$ + union bound

---

$\mathcal{H}$ of functions $f : \mathcal{X} \to \{0,1\}$

$$\nu(f) = \mathbb{P}[f(X) = 1] \quad \text{for } X \sim P$$

$$\nu_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad \text{for } X_1, \dots, X_n \sim P$$

Uniform convergence:

$$\sup_{f \in \mathcal{H}} |\nu(f) - \nu_n(f)| = O\left(\sqrt{\frac{d + \log(1/\delta)}{n}}\right) \quad \text{w.p. } 1-\delta$$

where d is some "dimension" of $\mathcal{H}$

VC dimension: $vc(\mathcal{H})$

max d s.t. $(f(x_1), \dots, f(x_d))$ can take on all 2^d distinct values for some $x_1, \dots, x_d \in \mathcal{X}$.

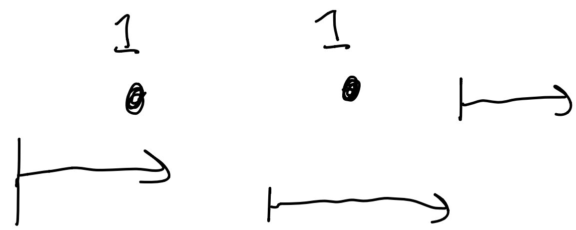
Examples:

$$\mathcal{H} = \{\mathbb{1}[x \geq t] \text{ for } t \in \mathbb{R}\}$$

$$vc(\mathcal{H}) = 1$$

0   1

0   0

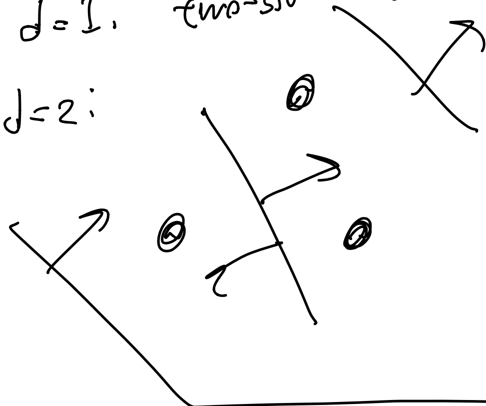


$\mathcal{H} = \{ \mathbb{I}[\langle v, x \rangle \geq \tau]$ for $v \in \mathbb{R}^d, \tau \in \mathbb{R} \}$

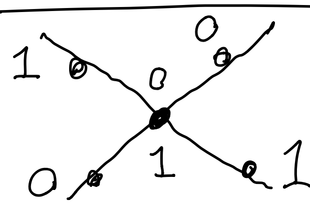
$vc(\mathcal{H}) = d+1$

$d=1$: two-sided thresholds

$d=2$:



all $8 = 2^3$



Facts about $vc(\mathcal{H})$

$V_{\mathcal{H}}(n)$ = # of ways a set of size n can be classified

$$V_{\mathcal{H}}(n) = 2^n \text{ for all } n \leq vc(\mathcal{H})$$

What about $n > vc(\mathcal{H})$?

Lemma. (Sauer-Shelah)

$$V_{\mathcal{H}}(n) \le \sum_{k=0}^{d} \binom{n}{k} \le 2 \cdot n^d$$

$\mathcal{H}$ of functions $f : \mathcal{X} \to \{0,1\}$

$$\nu(f) = \mathbb{P}[f(X) = 1] \quad \text{for } X \sim P$$

$$\nu_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad \text{for } X_1, \dots, X_n \sim P$$

Uniform convergence.

McDiarmid's inequality

True $\to$ $$\sup_{f \in \mathcal{H}} |\nu(f) - \nu_n(f)| = O\left(\sqrt{\frac{d + \log(1/\delta)}{n}}\right) \text{ w.p. } 1-\delta$$

where d is some "dimension" of $\mathcal{H}$

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} |\nu(f) - \nu_n(f)|\right] = O\left(\sqrt{\frac{d \log(n)}{n}}\right) \leftarrow \text{chaining}$$

noisy version of $\nu(f)$

$$\mathbb{E}\left[\sup_{f\in\mathcal{H}} |\nu_n(f) - \nu(f)|\right] \le \mathbb{E}\left[\sup_{f\in\mathcal{H}} |\nu_n(f) - \nu'_n(f)|\right]$$

$$= \frac{1}{n}\mathbb{E}_{X,X'}\left[\sup_{f\in\mathcal{H}}\left|\sum_{i=1}^{n} f(X_i) - \underbrace{f(X_i')}_{\equiv f(X_i') - f(X_i)}\right|\right]$$

$X_1', \dots, X_n' \sim P$

$\nu_n'(f)$

$$= \frac{1}{n}\mathbb{E}_{X,X',s}\left[\sup_{f\in\mathcal{H}}\left|\sum_{i=1}^{n} s_i (f(X_i) - f(X_i'))\right|\right]$$

random ± 1

$$\frac{1}{n}\mathbb{E}_{X,X',S}\left[\sup_{f\in\mathcal{H}}\left|\sum_{i=1}^{n} s_i(f(X_i)-f(X_i'))\right|\right]$$

$$\mathbb{E}_{X,X'}\left[\mathbb{E}_S\left[\sup_{f\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n} s_i(f(X_i)-f(X_i'))\right|\right]\right]$$

$\rightarrow f(X_i)-f(X_i')\in[-1,1]$ $-1, 0, 1$

$f(X_i)\rightarrow V_{\mathcal{H}}(n)$

$f(X_i')\rightarrow V_{\mathcal{H}}(n)$

$V_{\mathcal{H}}(2n)$

$\rightarrow$ across $f\in\mathcal{H}$: at most $V_{\mathcal{H}}(n)^2 \le 4n^{2d}$ values

$\rightarrow$ for fixed f, avg. of indep. bounded random variables

Tail bound: for fixed f,

* $\mathbb{P}_S\left[\frac{1}{n}\sum_{i=1}^{n} s_i(f(X_i)-f(X_i'))\ge t\right]\le \exp(-2nt^2)$

* Effectively at most $4n^{2d}$ distinct f

$\Rightarrow \mathbb{P}_S\left[\sup_f|\cdots|\ge t\right]\le 2\cdot 4n^{2d}\exp(-2nt^2)\le\delta$

$$\exp(-2nt^2)\le\frac{\delta}{8n^{2d}} \Rightarrow t\simeq\sqrt{\frac{\log(8n^{2d})+\log(1/\delta)}{n}}$$

$$=\Theta\left(\sqrt{\frac{d\log(n)+\log(1/\delta)}{n}}\right).$$