Lecture 5

Last time

- ▷ Degree of random
- ▷ graphs.
- ▷ Subexponentials
- ▷ Bernstein's Inequality

Today

- ▷ Johnson-Lindenstrauss Lemma.
- ▷ Orlicz norms

Dimension Reduction

Suppose we have a set of points $\mathcal{X} \to \{x_1, \dots, x_m\} \subseteq \mathbb{R}^d$ and we wanted to "compress" them by mapping them to $\mathbb{R}^n$ with $n < d$, while approximately maintaining their "geometry", i.e., we want $f : \mathbb{R}^d \to \mathbb{R}^n$ s.t. $\forall x, y \in \mathcal{X}$.

$$(1-\varepsilon)\|x-y\|_2^2 \le \|f(x)-f(y)\|_2^2 \le (1+\varepsilon)\|x-y\|_2^2$$ (☺)

How small can n be?

Theorem (Johnson-Lindenstrauss)

Fix $\varepsilon, \delta \in (0,1)$ a let $\mathcal{X} \subseteq \mathbb{R}^d$ be a set with n points. Take

$$n \geq \frac{8}{\varepsilon^2} \ln\left(\frac{m^2}{\delta}\right).$$

This is independent of d

and draw a matrix $M \in \mathbb{R}^{n \times d}$ with iid $N(0,1)$ entries. Then, with probability at least $1-\delta$, the map

$$f(x) = \frac{1}{\sqrt{n}} Mx$$

satisfies (☺) for all $x, y \in \mathcal{X}$ ⊣

Remark:

1) The embedding dimension depends on m, but not d. CRAZY!
2) This is an example of the probabilistic method.
3) This result is oblivious to the data.
4) What can you do when $\mathcal{X}$ is not finite? We'll come back to this question.

Proof: Given any $z = x - y$, we have

$$\frac{\|Mz\|^2}{\|z\|^2} = \sum_{i=1}^{n} \underbrace{\langle \overset{i\text{th row}}{M_i^T}, \frac{z}{\|z\|} \rangle}^2 \sim N(0,1) \text{ (why?)}$$

By the Sum Rule from Lecture 3

$\frac{\|Mz\|^2}{\|z\|^2}$ is $(2\sqrt{n}, 4)$-subexponential.

Thus, applying Bernstein's Ineq.

$$\mathbb{P}\left(\left|\frac{\|Mz\|^2}{n\|z\|^2} - 1\right| \geq \varepsilon\right) \leq 2\exp\left[-\left(\frac{n\varepsilon^2}{8} \wedge \frac{n\varepsilon}{8}\right)\right]$$

$$\uparrow$$

$$\frac{\|f(x)-f(y)\|^2}{\|x-y\|^2}$$

$$= 2\exp(-n\varepsilon^2/8).$$

This implies that for fixed $x, y \in X$

$\mathbb{P}((\star)$ does not hold for $x,y) \leq 2e^{-n\varepsilon^2/8}$.

Taking union bound over the $\binom{m}{2}$ pairs of points yields

$$\mathbb{P}((\star) \text{ holds for all } x,y \in X) \geq 1 - 2\binom{m}{2}e^{-\frac{n\varepsilon^2}{8}}$$

$$\geq 1 - m^2 e^{-\frac{n\varepsilon^2}{8}}$$

$$= 8. \qquad \square$$

Sub-Gaussian Norm

We gave one definition of

sub-Gaussians. Yet there many.

Proposition: Let X be a r.v. the following are equivalent (modulo const. factors:

1) $\exists K_1>0$ s.t. $\mathbb{P}(|X|>t)\le 2e^{-t^2/K_1}\quad \forall t\ge 0.$

2) $\exists K_2>0$ s.t. $\|X\|_{\mathcal{L}_p}:=(\mathbb{E}|X|^p)^{1/p}\le K_2\sqrt{p}\quad \forall p\ge 1.$

3) $\exists K_3>0$ s.t. $\mathbb{E}\exp(X^2/K_3^2)\le 2.$

Moreover, if $\mathbb{E}X=0$ then, these are quivalent to

4) $\exists K_4>0$ s.t. $\mathbb{E}\exp(\lambda X)\le\exp(K_4^2\lambda^2)\quad\forall\lambda\in\mathbb{R}.$

-1

Proof: We prove 1) ⇒ 2) ⇒ 3) ⇒ 4).

1) ⇒ 2) WLOG $K_1=1$ (why?). By HW1

$$\mathbb{E}|X|^p=\int_0^\infty p\,t^{p-1}\,\mathbb{P}(|X|\ge t)\,dt$$

$$\le p\int_0^\infty 2\,t^{p-1}e^{-t^2}\,dt$$

$t^2=s$

$$=\underbrace{p\int_0^\infty s^{p-1}e^{-s}\,ds}_{\text{Gamma function }\Gamma(p/2)}$$

$$\le p\,\Gamma(p/2)$$

$\Gamma(x)\le 3x^x$ for $x\ge\frac{1}{2}$. (why?)

$$\le 3p\,(p/2)^{p/2}$$

$$\Rightarrow \quad \|X\|_{L_p} \leq (3p)^{1/p} (p/2)^{1/2} \leq \frac{e^{(3/e)}}{\sqrt{2}} \sqrt{p}.$$

2) ⇒ 3) **WLOG** $K_2 = 1$.

$$\mathbb{E} \exp(\underset{\substack{\uparrow \\ \text{To be set}}}{\lambda^2} X^2) = 1 + \sum_{p \geq 1}^{\infty} \frac{\lambda^{2p} \, \mathbb{E} X^{2p}}{p!}$$

$$\leq 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p} (2p)^p}{p!}$$

$$\text{Sterling Approximation} \rightarrow \leq 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p} (2p)^p}{(p/e)^p}$$

$$= \sum_{p=0}^{\infty} (2e\lambda^2)^p = \frac{1}{1 - 2e\lambda^2} \overset{\substack{\frac{1}{2\sqrt{e}} \\ \downarrow}}{=} 2$$

3) ⇒ 1) Assume **WLOG** $K_3 = 1$.

$$\mathbb{P}(|X| > t) = \mathbb{P}(e^{X^2} \geq e^{t^2}) \leq e^{-t^2} \, \mathbb{E} e^{X^2} \leq 2 e^{-t^2}.$$

We proved 4) ⇒ 1) in Lecture 2. Prove the missing implication. □

A very slick consequence of this characterization is the following definition and result.

Def: The sub-Gaussian norm of a r.v. is

$$\|X\|_{\psi_2} := \inf\{K > 0 : \mathbb{E} \exp(X^2/K^2) \leq 2\}.$$

Lemma (HW): $\|\cdot\|_{\psi_2}$ is a norm on $\{X \mid \|X\|_{\psi_2} < \infty\}$. $\dashv$

We can restate the Hoeffding's ineq.

Theorem (Hoeffding via $\|\cdot\|_{\psi_2}$): Let $X_1, \dots, X_n$ independent r.v. and sub-Gaussian. Then,

$$\Big\|\sum_{i=1}^n X_i\Big\|_{\psi_2}^2 \le \overset{\text{universal.}}{C} \sum_{i=1}^n \|X_i\|_{\psi_2}^2. \qquad \dashv$$

Notice that $\|X\|_{\psi_2}$ does not require us to recenter X. In particular, we have

Lemma: $\|X - \mathbb{E}X\|_{\psi_2} \le C\|X\|_{\psi_2}$.

Proof: Apply triangle ineq.

$$\|X - \mathbb{E}X\|_{\psi_2} \le \|X\|_{\psi_2} + \|\mathbb{E}X\|_{\psi_2}.$$

To bound the second term

$$\|\mathbb{E}X\|_{\psi_2} = |\mathbb{E}X|\,\|1\|_{\psi_2} \underset{\text{Jensen's}}{\le} c\,\mathbb{E}|X| \overset{2)}{\le} c'\|X\|_{\psi_2}.$$

Different const. □

There is an analogous story for

subexponentials.

Proposition: Let X be a r.v. the following are equivalent (modulo const. factors):

1) $\exists K_1 > 0$ s.t. $\mathbb{P}(|X| > t) \le 2e^{-t/K_1} \quad \forall t \ge 0.$

2) $\exists K_2 > 0$ s.t. $\|X\|_{\mathcal{L}_p} := (\mathbb{E}|X|^p)^{1/p} \le K_2 \, p \quad \forall p \ge 1.$

3) $\exists K_3 > 0$ s.t. $\mathbb{E}\exp(|X|/K_3) \le 2.$

Moreover, if $\mathbb{E}X = 0$ then, these are quivalent to

4) $\exists K_4 > 0$ s.t. $\mathbb{E}\exp(\lambda X) \le \exp(K_4^2 \lambda^2) \quad \forall |\lambda| \le \frac{1}{K_4}.$

$\dashv$

This motivates the following.

Def: The subexponential norm of a r.v. is

$$\|X\|_{\psi_1} := \inf\{K > 0 : \mathbb{E}\exp(|X|/K) \le 2\}.$$

$\dashv$

Just as before $\|\cdot\|_{\psi_1}$ is a norm over the set of subexponentials. Moreover

$$\|X - \mathbb{E}X\|_{\psi_1} \le C\|X\|_{\psi_1}.$$

We motivated subexponential via χ^2 distributions, in turn products of sub-Gaussians are always subexponential.

Lemma: Suppose X, Y are sub-Gaussian, then

$$\|XY\|_{\psi_1} \le \|X\|_{\psi_2} \|Y\|_{\psi_2}.$$ ⊣

Proof: WLOG $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. Then

$$\mathbb{E}\exp(|XY|) \underset{\uparrow}{\le} \mathbb{E}\exp(X^2/2 + Y^2/2)$$

Young's ineq. $|ab| \le a^2/2 + b^2/2$

$$= \mathbb{E}\exp(X^2/2)\exp(Y^2/2)$$

$$\le \frac{1}{2}\left(\mathbb{E}\exp(X^2) + \mathbb{E}\exp(Y^2)\right)$$

$$\le \frac{1}{2}(2+2) = 2.$$ □

It is natural to wonder whether other functions besides exponentials define other norms capturing different growth/tails. Indeed, this is the case.

Def: Given a convex, nondecreasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ s.t. $\psi(0) = 0$ with $\psi(t) \xrightarrow{t\to\infty} \infty$, define the Orlicz norm of a r.v. X as

$$\|X\|_\psi = \inf\{K>0 \mid \mathbb{E}\,\psi(|X|/K) \le 1\}.$$ +

One can show that this defines a norm on $\{X \mid \|X\|_\psi < \infty\}$.

Example: For $\psi(t) = |t|^p$ with $p \ge 1$ defines $\mathcal{L}_p$. While $\psi_2(t) = e^{t^2} - 1$ and $\psi_1(t) = e^t - 1$ define sub-Gaussians and subexponentials, respectively. +