Dimension Reduction (JL)

Thm [JL84]: Given n points in $\mathbb{R}^D$ ~~can embed into~~ $\exists$ linear map $A: \mathbb{R}^D \to \mathbb{R}^k$ with $k = O\left(\frac{\log n}{\varepsilon^2}\right)$

st. $$\frac{\|Ax - Ay\|_2^2}{\|x-y\|_2^2} \in 1 \pm \varepsilon.$$

This is tight [Alon, Larsen-Nelson].

Thm: Given a unit vector $x \in \mathbb{R}^D$, $\exists$ a linear map $A: \mathbb{R}^D \to \mathbb{R}^k$ with $k = O(\varepsilon^{-2} \log 1/\delta)$ st.

$$\Pr\left[\|Ax\|^2 \in (1 \pm \varepsilon)\right] \geq 1 - \delta. \qquad \Pr\left[\|Ax\|^2 \in 1 \pm \varepsilon\right] \geq 1 - e^{-k\varepsilon^2/\text{const.}}$$

Pf of JL: $\exists \binom{n}{2}$ pairs of pts. So set $\delta = 1/n^2 \Rightarrow$ wp $1/2$ get $\|Ax_i - Ax_j\|^2 / \|x_i - x_j\|^2$ in $1 \pm \varepsilon$? ∎

this is tight by ~~Bees~~ Larsen-Nelson, but also an older result of ~~Nelson-~~ Jayaram Woodruff, Kane Meka Nelson.

Algo: $A = \frac{1}{\sqrt{k}}$ Matrix (${k \times D}$), $k = \frac{\log n}{\varepsilon^2}$, full of Gaussians $N(0,1)$.

Recall
- $c N(\mu, \sigma^2) \sim N(c\mu, c^2\sigma^2)$
- $N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
- $\bar{v} \cdot \bar{g} = \sum_i v_i N(0,1) \sim N(0, \sum v_i^2) = N(0, \|v\|^2)$

• **Intuition**: $x = e_1$

$$\Rightarrow Ax = \frac{1}{\sqrt{k}} \bar{g} \Rightarrow \|Ax\|^2 = \frac{1}{k}\|\bar{g}\|^2 = \frac{1}{k}\sum_{i=1}^{k} N(0,1)^2$$

$\Rightarrow E[\|Ax\|^2] = 1$, sum of several, so should concentrate.
But unbounded, can't apply Chernoff.

• **Formally**: $E[\|Ax\|^2]$ = correct since each entry of $Ax \sim N(0, \|v\|^2) = N(0,1)$.

• **Concentration**: $\Pr[\underbrace{\|Ax\|^2}_{\sum_{i=1}^{k} g_i} \geq 1+\varepsilon] \leq \frac{E[e^{t\sum g_i^2}]}{e^{t(1+\varepsilon)}} = \frac{\prod E[e^{tg_i^2}]}{e^{t(1+\varepsilon)}} \rightsquigarrow$ calculation $\frac{1}{\sqrt{1-2t}}$

then optimize $\left(\frac{1}{\sqrt{1-2t}\, e^{t(1+\varepsilon)}}\right)^k$ to get $e^{-k\varepsilon^2/8}$.

$= e^{-k(t + \varepsilon t - \frac{1}{2}\ln(1-2t))}$.

Easy as ABC.

Extends to other distributions

M_{ij} = gaussian $\quad A = \frac{1}{\sqrt{k}} M$. $\quad \|Ax\|^2 = \frac{1}{k}\sum g_i^2$

in gen. $M_{ij} = \cancel{V}$ iid $\to$ $\quad \|Ax\|^2 = \frac{1}{k}\sum$ subexponentials.

subgaussian $\mu=0$, var $=1$.

Mean: $\mathbb{E}\|Ax\|^2 = \mathbb{E}\frac{1}{k}\sum_{i=1}^{k}\underbrace{\left(\sum_{j=1}^{D} V_{ij}x_j\right)^2}$

$\mathbb{E}\sum V_{ij}^2 x_j^2$ + cross terms $\to$ Exp $=0$.

$= \sum x_j^2 = \|x\|^2$.

{variance 1, mean 0.

Concentration?

Subgaussian and Subexponential r.v.s.

Def: V is sub~~exp~~gaussian (with parameter c) if $\exists$ constant c s.t. $\mathbb{E}\left[e^{s(V-\mathbb{E}(V))}\right] \le e^{c s^2/2} \quad \forall s \in \mathbb{R}$

$\Updownarrow$

$Pr\left[|V - EV| \ge \lambda\right] \le 2e^{-\lambda^2/2c^2}$

Examples: Gaussian. $\quad Pr[|g| \ge s] \le e^{-s^2/2} \quad c=1$

: Rademacher ± 1 w.p. $\frac{1}{2}$ $\quad E[e^{sV}] = \frac{e^s + e^{-s}}{2} \sim e^{-s^2/2} \quad c=1 \quad$ Any Bounded r.v. $[a,b]$: $c = \frac{b-a}{2}$.

: Sums of subgaussians = subgaussian.

V_i subgaussian w/ c_i, $\Rightarrow \sum a_i V_i$ subg. w/ param $\sum a_i^2 c_i$.

Def: W is subexponential if $\exists$ constants $\nu, b \ge 0$ if $\mathbb{E}\left[e^{s(W-\mu)}\right] \le e^{\nu^2 s^2/2}$ $\boxed{\forall |s| \le \frac{1}{b}.}$

$\Leftrightarrow \exists$ constants c_1, c_2 st $Pr\left[|V - EV| \ge \lambda\right] \le c_1 e^{-c_2\lambda} \quad \forall \lambda \ge 0$.

Examples: Exponential

: square of subgaussian

: sums of subexponentials. W_1 (ν_1, b_1), W_2 (ν_2, b_2) $\Rightarrow a_1W_1 + a_2W_2 \sim \text{subexp}(a_1^2\nu_1 + a_2^2\nu_2)$ for $|sa_1| \le \frac{1}{b_1}$, $|sa_2| \le \frac{1}{b_2}$.

: χ^2 (i.e. square of $N(0,1)$) : MGF $\mathbb{E}[e^{s(x-1)}] = \frac{e^{-s}}{\sqrt{1-2s}} \Rightarrow$ holds only for $\boxed{s \le \frac{1}{2}}$

Subexponential tail bound:

if X is subexp(ν, b) then $\quad \forall t \geq 0 \quad Pr[X \geq \mu + t] \leq \begin{cases} e^{-t^2/2\nu^2} & 0 \leq t \leq \nu^2/b \\ e^{-t/2b} & t > \nu^2/b. \end{cases}$

i.e. sub~~exp~~gaussian at the begining, and then exponential after that. $\Big\} \rightarrow \exp\left(-\frac{t^2}{2(\nu^2 + tb)}\right)$

"Bernstein" type bound.

$\Rightarrow$ Prove JL for subgaussian ensembles. $\|Ax\|^2 \sim \frac{1}{K} \sum$ subgaussian^2 $\sim \frac{1}{K} \sum$ subexp.

use above bound $\Rightarrow$ done!

BTW: Any bounded rvs $[a,b]$ are subgaussian with param $c = \frac{b-a}{2}$. Sum will have param: $\frac{n(b-a)}{2}$

$\Rightarrow$ get ~~Chernoffs~~ Hoeffdings bound $\quad Pr[|S_n - \mu| > t] \leq e^{-\frac{2t^2}{n(b-a)^2}}$

But could be that these are subexponential with different parameters too:-

Maybe it has a much ~~better~~ smaller ν for which

$$E[e^{s(W-\mu)}] \leq e^{\nu^2 s^2/2} \quad \text{for } \forall |s| \leq \frac{1}{b}$$ ↙ but for a smaller range

Then we can get a "Bernstein"-type bound that trades off this ν vs. b.

As in the previous lecture.

JL and compressive sensing

Given $x \in \mathbb{R}^D$, $A \in \mathbb{R}^{k \times D}$ st $Ax = b$ and $\underset{\text{supp}(x)}{\|x\|_0} \le r$, can we find such an x?

sparse solution to linear systems.

NP-hard in general.

- Can we design A matrices st when we get b obtained by A(sparse vector) $= b$ we can reconstruct x?

- One algo: "Basis Pursuit" $\boxed{\begin{array}{l} \min \|x\|_1 \\ \text{st } Ax = b. \end{array}}$

Can we design matrices A that if $Ax = b$ has unique r-sparse sol^n x^* then BP finds unique min x^*? } BP-exact.

Thm: [Donoho, Candes-Tao, Rudelson-Vershynin]

- Can do this if $K \ge \Omega(r \log D/r)$, A drawn from distribution -JL ensemble (i.e. $\forall x$, $\Pr[\|Ax\| \in 1 \pm \varepsilon \|x\|] \ge 1 - e^{-c\varepsilon^2 k}$)

$\Pr[A \text{ is BP exact}] \ge 1 - e^{-ck}$

[Note: just to figure out which subset is supp(x), need $\log \binom{D}{r} \approx r \log(D/r)$ bits of info. So this is pretty good].

Main Idea: RIP (restricted isometry ppty).

Def: A is t-restricted εRIP if $\forall x$ with $\text{supp}(x) \le t$, $\frac{\|Ax\|}{\|x\|} \in 1 \pm \varepsilon$.

Lemma: if A is $3r$-restricted ε-RIP $\underset{\exists \varepsilon > 0 \text{ st}}{}$ $\Rightarrow$ A is BP exact with sparsity r.

Intuition: ("every $3r$ columns are almost orthogonal" $\Rightarrow$ "no nontrivial linear combs").

Lemma: Distributional JL $\Rightarrow$ t-restricted εRIP.

Intuition: "counting: $\binom{n}{t}$ columns, for each ε-net of size $\left(\frac{1}{\varepsilon}\right)^t$. All good.

See [Matousek notes], [Baraniuk, Davenport, de Vore, Wakin], etc.