Lecture 7 — 2/2/21

Last time: $f \in \mathbb{R}_+^n$

goal: estimate $F_2 = \|f\|_2^2$.

Toward algos: for $i = 1..k$: $k = O\left(\frac{1}{\varepsilon^2}\right)$

picked $\sigma_{ij} \in \{\pm 1\}$, $j \in [n]$.

sketches $z_i = \sum_{j=1}^{n} \sigma_{ij} f_j$.

est: $\frac{1}{k} \sum z_i^2 \stackrel{\Delta}{=} Z^2$.

Proved:

$$\left.\begin{array}{l} \mathbb{E}[Z^2] = F_2 \\ \mathrm{Var}[Z^2] \le \frac{4F_2}{k}. \end{array}\right\} \stackrel{\text{by Cheb.}}{\Rightarrow} \quad Z^2 = (1 \pm \varepsilon) F_2$$

with prob $\ge 90\%$

if $k \ge \Omega\left(\frac{1}{\varepsilon^2}\right)$.

Also: what if Prob. fail $\le \delta$ param.

$Z^2 = (1 \pm \varepsilon) \cdot F_2$ with prob. $\ge 1 - \delta$,

as long as $k \ge \Omega\left(\frac{1}{\delta} \cdot \frac{1}{\varepsilon^2}\right)$.

(by Chebyshev).

A question today: can we get better dep. on δ?

Yes, in fact $k = O\left(\frac{\lg 1/\delta}{\varepsilon^2}\right)$ is enough.

Tow+ is a linear sketch:

$$(z_1, z_2, \dots z_k)^T = \frac{1}{\sqrt{k}} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{k1} & \dots & & \sigma_{kn} \end{bmatrix} \cdot \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

each $\in \{\pm 1\}$ i.i.d.

Est: $\sum_{i=1}^{k} z_i^2$.

Corollary: sketch of sum of f^1 & f^2
= sum of sketches of f^1 & f^2.

Can see sketch: $\varphi : \mathbb{R}^n \to \mathbb{R}^k$:

$$\varphi(f) = \boxed{\sigma\text{'s}} \cdot \boxed{f}.$$

Est: $\sum_i (\varphi(f)_i)^2 \approx F_2$

$$\|\varphi(f)\|_2^2 = (1\pm\varepsilon)\cdot F_2 = (1\pm\varepsilon)\cdot\|f\|_2^2$$

ℓ_2 norm of a vector of dim. $k \approx O\left(\frac{\lg 1/\delta}{\varepsilon^2}\right)$.

ℓ_2 norm of a vector of dim. n

An example of Dimension Reduction

Lemma [Johnson-Lindenstrauss '84]:

$\forall \varepsilon > 0$, there is a distribution over ~~random~~ $\varphi: \mathbb{R}^n \to \mathbb{R}^k$

s.t. $\forall x, y \in \mathbb{R}^n$:

$$\Pr_\varphi\left[\|\varphi(x)-\varphi(y)\| = (1\pm\varepsilon)\cdot\|x-y\|\right] \geq 1 - e^{-\varepsilon^2 k/9}.$$

Proof: If φ is linear:

$\varphi(x) - \varphi(y)$ approximates $\|x-y\|$

equivalent:

$\varphi(x) - \varphi(y) = \varphi(x-y)$ approximates $\|z\|$, $z \triangleq x-y$.

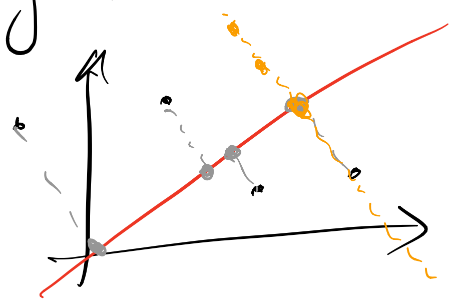
Enough to show $\|\varphi(z)\| \approx \|z\|$,
$\forall z \in \mathbb{R}^n$.

Which φ?

1) Toy + actually works.

$$\varphi(z) = \begin{array}{c c} & n \\ k & \boxed{\pm 1} \end{array} \cdot z.$$

2) φ: pick a random k-dim. linear subspace of $\mathbb{R}^n$, project argument onto subspace.

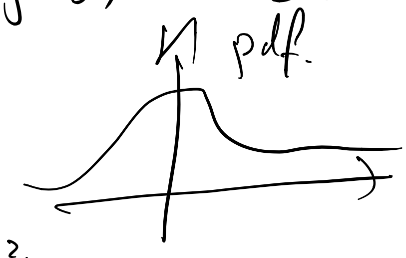


3) $$\varphi(z) = \boxed{g_{ij}} \cdot z.$$

a random iid Gaussian/Normal.

Obs: if no randomness, then $\forall \varphi : \mathbb{R}^n \to \mathbb{R}^{n-1}$ $\exists\, x, y$ will collapse.

$$\varphi(z) = \frac{1}{\sqrt{k}}\left(\sum_{j=1}^{n} g_{1j} z_j \,,\ \sum g_{2j} z_j, \ \ldots\ \sum g_{kj} z_j\right).$$



$g_{ij} \sim$ Gaussian $N(0,1)$.

g has $\mathrm{pdf}(g) = \frac{1}{\sqrt{2\pi}} e^{-g^2/2}$.

Property: spherically symmetric.

$(g_1, \ldots, g_n) \leftarrow$ pdf is s.s.

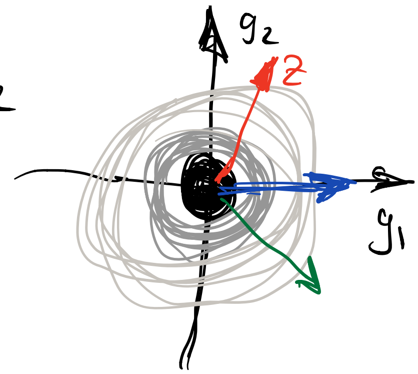
iid $\in N(0,1)$ $\quad$ $\mathrm{pdf}(g_1, \ldots, g_n)$ depends only on $\|g\|$.

$$\mathrm{pdf}(g_1, g_2, \ldots g_n) = \prod_{i=1}^{n} \frac{1}{(\sqrt{2\pi})} \cdot e^{-g_i^2/2}$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot e^{-\sum_{i=1}^{n} g_i^2/2}$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot e^{-\|g\|_2^2/2}.$$

Corollary: $z \cdot g$ has same distribution as if $z = (\|z\|, 0, 0, \dots 0)$.



$\Rightarrow$ $z \cdot g$ has distribution $\|z\| \cdot g'$,
$g' \in N(0,1)$

Why Gaussian distrib. can replace ± 1 in ToW?

1) $\mathbb{E}[\sigma_{ij}] = 0$

2) $\mathbb{E}[\sigma_{ij}^2] = 1$.

3) $\mathbb{E}[\sigma_{ij}^4] = 1$. ok if $O(1)$.

To get guarantees of ToW+, enough to replace $\sigma_{ij} \in \{\pm\}$ with # r.v. from dist. sat $1 \div 3$.

Eg: $N(0,1)$ satisfies them.

So far: $\varphi(z) = \frac{1}{\sqrt{k}} \Big(\underbrace{\sum g_{1j} z_j}_{\text{distributed as } g^{(1)} \cdot \|z\|}, \underbrace{\sum g_{2j} z_j}_{g^{(2)} \cdot \|z\|}, \dots \Big)$

$\varphi(z)$ is distributed as

$$\frac{1}{\sqrt{k}} \left(\|z\| \cdot g^{(1)}, \|z\| \cdot g^{(2)}, \dots, \|z\| \cdot g^{(k)} \right)$$

iid $\in N(0,1)$.

$$= \|z\| \cdot \frac{1}{\sqrt{k}} \cdot \underbrace{\left(g^{(1)}, g^{(2)}, \dots g^{(k)} \right)}_{k\text{-dim } N(0,1).}$$

$$\|\varphi(z)\|_2^2 = \|z\|_2^2 \cdot \frac{1}{k} \cdot \underbrace{\sum_{i=1}^{k} \left(g^{(i)}\right)^2}_{\chi^2_k}.$$

chi-squared dist. w/ k degrees of fr.

Fact: $\Pr\left[\chi^2_k \notin (1 \pm \varepsilon) k \right] \le e^{-\varepsilon^2 k/8}$ for $\varepsilon < 1/2$.

$\Rightarrow \Pr_\varphi [\, \|\varphi(z)\|_2^2 = (1 \pm \varepsilon)\|z\|^2] \geq 1 - e^{-\varepsilon^2 k/9}$. $\boxtimes_{JL}$.

$\sqrt{1-\varepsilon} \approx 1 - \varepsilon/2$.

$\alpha \in (1 \pm \varepsilon)^{1/2} \cdot \alpha' \Rightarrow \alpha \in (1 \pm \varepsilon)\alpha'$.

<u>Corollary</u> [of JL]: fix N vectors $x_1, \ldots x_N \in \mathbb{R}^n$. Pick φ as in JL, with $k = \Theta\left(\frac{\lg N}{\varepsilon^2}\right)$. Then

$$\Pr_\varphi [\, \forall i,j \in [N],\ \|\varphi(x_i) - \varphi(x_j)\| = (1 \pm \varepsilon)\|x_i - x_j\|]$$
$$\geq 1 - 1/N.$$

<u>proof</u>: fix $k = \frac{3 \cdot 9 \cdot \ln N}{\varepsilon^2}$.

Then, by JL, $\forall\, i,j \in [N]$:

$$\Pr[\, \|\varphi(x_i) - \varphi(x_j)\| \notin (1 \pm \varepsilon)\|x_i - x_j\|] \leq e^{-\varepsilon^2 k/9} = 1/N^3.$$

By union bound over all pairs $i,j \in [N]$:

$\Rightarrow \Pr[\exists\, i,j \in [N]$ where $\ldots \notin \ldots]$

$$\leq N^2 \cdot 1/N^3 = 1/N.$$

$\boxtimes$

<u>Rephrase:</u> for $\forall N$ vectors in $\mathbb{R}^n$, can map them into $\mathbb{R}^k$,

$$k = O\left(\frac{\lg N}{\varepsilon^2}\right),$$

while preserving dist $1 \pm \varepsilon$.

<u>Remark:</u> is k tight? yes, even for non-linear maps φ. (even if φ depends on $x_1 \ldots x_N$).

<u>Remark:</u> what about dim-reduction in other spaces?

…

Can we do the same in ℓ_1?

$\varphi: \ell_1^n \to \ell_1^k$ s.t.

distances preserved with "good" prob.

or map N vectors in ℓ_1 into lower-dim. ℓ_1?

Basically no: min dim is

[Brinkman, Charikar ~2003] $k = N^{\Omega(1/\alpha)}$ for

aprox. $\alpha > 1$.

[Naor, Johnson 2009]: good dim-red. $\Rightarrow$ must be ℓ_2.