14. (Aside) Johnson-Lindenstrauss

Friday, February 9, 2024 3:16 PM

Note: these notes use different notation

Fact If $X_i \sim N(0,1)$ are independent, then $Z = \sum_{i=1}^{M} X_i^2 \sim \chi_M^2$ ("chi-squared")

and if $Z \sim \chi_M^2$ then $\mathbb{E}[Z] = M$ and $\forall \varepsilon \in (0, \tfrac{1}{2})$

$$\mathbb{P}[\,|Z - M| \geq \varepsilon \cdot M\,] \leq 2 \cdot \exp\left(-\tfrac{M}{4}(\varepsilon^2 - \varepsilon^3)\right)$$

could also use Bernstein's ineq. for sub-exponential r.v.

Lemma Fix $x \in \mathbb{R}^N$, let $A \in \mathbb{R}^{M \times N}$ be random w/ $A_{ij} \sim N(0,1)$ iid

Let $y = Ax$, then w.p. $\geq 1 - 2\exp\left(\frac{-M}{4}(\varepsilon^2 - \varepsilon^3)\right)$,

$$(1-\varepsilon)\|x\|^2 \leq \left\|\tfrac{1}{\sqrt{M}} y\right\|^2 \leq (1+\varepsilon)\|x\|^2$$

Facts: $\vec{a} \sim N(\vec{\mu}, \Sigma)$

$\Rightarrow C \cdot \vec{a} \sim N(C \cdot \vec{\mu}, C \Sigma C^T)$

proof wlog let $\|x\| = 1$.

- y_i is also a Gaussian, as it is a weighted sum of Gaussians

  (Recall: if $X \sim \mathcal{L}_1$ and $Z \sim \mathcal{L}_2$ then $X + Y \sim \mathcal{L}_1 * \mathcal{L}_2$ convolution

  The sum of normal distributions (i.e. multimodal) isn't Gaussian, it's a mixture model)

- so it's completely characterized by its mean + variance

$$\mathbb{E}[y_i] = \sum_j x_j \cdot \mathbb{E}[a_{ij}] = 0$$

$$\mathrm{Var}[y_i] = \mathbb{E}[y_i^2] = \mathbb{E}\Big[\Big(\sum_{j=1}^{N} a_{ij}\, x_j\Big)^2\Big] + \mathbb{E}\Big[\sum_{j=1}^{N} \sum_{\substack{j'=1 \\ j' \neq j}}^{N} a_{ij} a_{ij'}\, x_j x_{j'}\Big]$$

and $y_i \perp y_j \; (i \neq j)$ $\quad = \|x\|^2 + 0 = 1$

$\Rightarrow \|y\|^2 \sim \chi_M^2$

and $\big|\|y\|^2 - M\big| \geq \varepsilon \cdot M \Leftrightarrow \left|\frac{1}{M}\|y\|^2 - \underbrace{1}_{=\|x\|^2}\right| \geq \varepsilon$ so use Fact above.

Thm (Johnson-Lindenstrauss 1984) [one of many variants]

Let $X = \{x_1, \dots, x_K\} \subseteq \mathbb{R}^N$ and $\varepsilon \in (0, \tfrac{1}{2})$. If $M \geq \frac{16}{\varepsilon^2}\log(K)$

then $\exists$ a Lipschitz continuous map $f: \mathbb{R}^N \to \mathbb{R}^M$ s.t. $\forall x, y \in X$

$$(1-\varepsilon)\|x-y\|^2 \leq \|f(x) - f(y)\|^2 \leq (1+\varepsilon)\|x-y\|^2 \quad (*)$$

proof Draw $A \in \mathbb{R}^{M \times N}$ as before, think of $f(x) = \frac{1}{\sqrt{M}} \cdot Ax$

For a fixed x, y then (via linearity) $(*)$ holds w.p. $\geq 1 - 2 \cdot \exp\left(\frac{-M}{4}(\varepsilon^2 - \varepsilon^3)\right)$

and we have $\binom{K}{2} = \frac{K(K-1)}{2}$ pairs, so

$$\mathbb{P}[\text{any pair } x,y \text{ fails } (*)] \leq \sum_{\text{all pairs}} \mathbb{P}[\text{fixed pair } x,y \text{ fails } (*)]$$

$$\leq \frac{K(K-1)}{2} \cdot 2 \cdot \exp\left(\frac{-M}{4}(\varepsilon^2 - \varepsilon^3)\right)$$

So choose $M = \frac{16}{\varepsilon^2}\log(K)$ ...

$$= K(K-1)\exp\left(\frac{-16}{4\varepsilon^2}(\varepsilon^2 - \varepsilon^3)\log(K)\right)$$

14a. Johnson-Lindenstrauss

Monday, February 12, 2024   11:04 AM

$$= k\cdot(k-1)\cdot k^{\left(-4(1-\frac{1}{2}\varepsilon)\right)} \qquad \text{and} \quad \varepsilon<\tfrac12 \Rightarrow 1-\tfrac{\varepsilon}{2} > -\tfrac12$$

$$< k\cdot(k-1)\cdot k^{-2} = \frac{k-1}{k} < 1 \qquad \text{or } -4(1-\tfrac{\varepsilon}{2}) < -2$$

So $\mathbb{P}[\,(*) \text{ holds } \forall \text{ pairs}\,] > 0$

via the probabalistic method this means such a map f must exist. $\square$

Discussion

- In practice, for a specific A that we draw, we want it to be very likely, so make M larger in that case

- $O(1/\varepsilon^2)$ dependence grows quickly so it's best for low accuracy
  ... but analysis is tight up to a $\log(1/\varepsilon)$ factor
  $\log(K)$ dependence is "correct"

- Independent of original dimension N!] Big deal!
- Works if A_{ij} sub-Gaussian
- Faster "JL-inspired" transforms exist... ask me about them?

- See David Woodruff's 2014 monograph for an example of a chaining argument to use classical JL to apply to a whole subspace.

- Preserving $\|\Phi(x-y)\|^2 = (1\pm\varepsilon)\|x-y\|^2 \;\; \forall x,y\in V$
also preserves inner products. Assume $0\in V$ so $\|\Phi x\|^2 \approx \|x\|^2 \; \forall x\in V$
Then the polarization identity (for $\mathbb{R}$ scalars) is
$$\langle x,y\rangle = \tfrac12\left(\|x\|^2+\|y\|^2-\|x-y\|^2\right)$$
so
$$\langle \Phi x, \Phi y\rangle = \tfrac12\left(\|\Phi x\|^2+\|\Phi y\|^2-\|\Phi(x-y)\|^2\right)$$
$$\approx \tfrac12\left(\|x\|^2+\|y\|^2-\|x-y\|^2\right) = \langle x,y\rangle .$$

14b. Johnson-Lindenstrauss + chaining

Sunday, February 18, 2024 8:30 PM

$a = (1 \pm \varepsilon) b$ shorthand for $(1-\varepsilon) b \le a \le (1+\varepsilon) b$

Recap: before we talked about preserving distance $\|\Phi x - \Phi y\|^2 = (1 \pm \varepsilon)\|x-y\|^2$ and if this is for k points (ie. $k \cdot (k-1)/2$ pairs), can do it w/ random matrix w/ $m \ge O(\frac{1}{\varepsilon^2} \log(k/\delta))$

To simplify, look only at $\|\Phi x\|^2 = (1 \pm \varepsilon)\|x\|^2$ for k points
(to go from k to $\tilde{k} = O(k^2)$ pairs, it's in the log term so not a big deal)

What if we want $\|\Phi x\|^2 = (1 \pm \varepsilon)\|x\|^2$ for all $\underbrace{x \text{ in a subspace}}_{\text{an } \infty \text{ \# of points!}}$ V

In particular, $A = V \cap S^{n-1}$ ← unit sphere
↖ d-dim subspace of $\mathbb{R}^n$

Idea #1: Cover A w/ an $\varepsilon/2$-net, let this be "B". Then find a $\varepsilon/2$ J-L matrix for $|B|$ points (ie. $N_2(\varepsilon/2, A)$ points). Use triangle-ineq.
That works but it gives a large number...
$N_2(\varepsilon/2, A) \le (1 + 2/\varepsilon)^d$ ... nasty ε-dependence.

Idea #2 Clever chaining argument (cf. Woodruff book)

let B be a $\varepsilon = \frac{1}{2}$ net for A (and $B \subseteq A$, so $\forall b \in B$, $\|b\| \le 1$)

Let $\|\Phi b\|^2 = (1 \pm \varepsilon)\|b\|^2$ $\forall b \in B$
actually assume $\langle \Phi b, \Phi b' \rangle = (1 \pm \varepsilon)\langle b, b' \rangle$ $\forall b, b' \in B$ which is similar

Fix $a \in A$, goal is $\|\Phi a\|^2 = (1 \pm \varepsilon)\|a\|^2$. Note $\|a\| = 1$

Write $a = b^0 + a^0$ w/ $b^0 \in B$ being closest pt. in B, so $\|a^0\| \le \frac{1}{2}$ since B is a $\frac{1}{2}$-net

Then $B \subseteq A \subseteq V$ (subspace)

so $a^0 \in V$ also. So $\frac{a^0}{\|a^0\|} \in A = V \cap S^{n-1}$

Write $a^0/\|a^0\| = b^1 + a^1$ w/ $b^1 \in B$, $\|a^1\| \le \frac{1}{2}$, etc.

So, continuing,

$\|b^0\| \le 1 \quad \forall k$
$\|a^k\| \le \frac{1}{2} \quad \forall k$

$$a = b^0 + a^0 = b^0 + \|a^0\|(b^1 + a^1)$$
$$= b^0 + \|a^0\|(b^1 + \|a^1\|(b^2 + a^2))$$
$$= \sum_{k=0}^{\infty} c^k \quad \text{w/ } c^0 = b^0,\ c^1 = \|a^0\| \cdot b^1,\ c^2 = \|a^0\| \cdot \|a^1\| \cdot b^2 \ldots$$

so $\|c^k\| \le 2^{-k}$

Idea: we re-use the same b^i sometimes, just w/ different coefficients!

being a bit "loose" w/ details

So

$$\|\Phi a\|^2 = \|\Phi \textstyle\sum c^k\|^2 = \displaystyle\sum_{0 < i \le j < \infty} \|\Phi c^i\|^2 + 2\langle \Phi c^i, \Phi c^j \rangle$$

did Cauchy-Schwarz

$$= \Big(\underbrace{\sum_{0 < i < j < \infty} \|c^i\|^2 + 2\langle c^i, c^j \rangle}_{= \|\sum c^k\|^2}\Big) \pm 2\varepsilon \Big(\underbrace{\sum_{i \le j} \underbrace{\|c^i\|}_{\le 2^{-i}} \cdot \underbrace{\|c^j\|}_{\le 2^{-j}}}_{O(1)}\Big)$$

$$= \|a\|^2 + O(\varepsilon)$$