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 $\forall \varepsilon \in (0, \frac{1}{2})$

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

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. $\geq 1 - 2\exp\left(\frac{-M}{4}(\varepsilon^2 - \varepsilon^3)\right)$,

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

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 $X + Y \sim \underbrace{\mathcal{L}_1 * \mathcal{L}_2}_{\text{convolution}}$

  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

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

$\text{Var}[y_i] = \mathbb{E}[y_i^2] = \mathbb{E}\left[\sum_{j=1}^{N} a_{ij}^2 x_j^2\right] + \mathbb{E}\left[\sum_{j=1}^{N} \sum_{\substack{j'=1 \\ j \neq j'}}^{N} a_{ij} a_{ij'} x_j x_{j'}\right]$

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

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

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

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

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

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

14a. Johnson-Lindenstrauss

Monday, February 12, 2024 11:04 AM

$$= k\cdot(k-1)\cdot k^{(-4(1-1/\varepsilon))}$$ and $\varepsilon < 1/2 \Rightarrow 1-1/\varepsilon > -1/2$

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

So $\mathbb{P}[\,(*)$ holds $\forall$ 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.