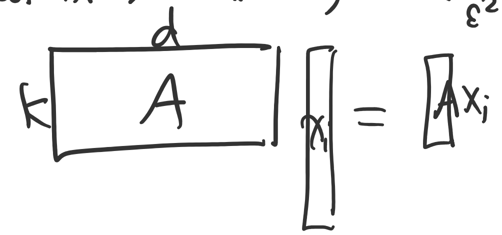
Gaussian r.v.'s, Chernoff Bounds, Dimension Reduction

Motivation (Data Science): Suppose $x_1, \dots, x_n \in \mathbb{R}^d$ data points in high dimensions (dn space). Can we compress this data while keeping pairwise relations e.g. distances $|x_i - x_j|$, inner product $\langle x_i, x_j \rangle$

Thm (Johnson-Lindenstrauss): $\exists$ projection matrix $A \in \mathbb{R}^{k \times d}$, $k = O\left(\frac{\log n}{\varepsilon^2}\right)$, s.t. $1 - \varepsilon \le \frac{\|Ax_i - Ax_j\|_2^2}{\|x_i - x_j\|_2^2} \le 1 + \varepsilon \quad \forall i, j \in [n]$.



In fact, taking each entry as $\frac{N(0,1)}{\sqrt{k}}$ (Gaussian $\to N(0,1)$) succeeds with prob $1 - \frac{1}{\text{poly}(n)}$.

Remarks: ① Fast algo: sample $kd = O(\varepsilon^{-2} d \log n)$ $N(0,1)$'s.

② Data-oblivious: doesn't depend on $x_1 \cdots x_n$!

<u>Main Lemma</u> (Distributional J-L):

If each entry of $A \sim \frac{N(0,1)}{\sqrt{k}}$ with $k = O(\varepsilon^{-2} \log \delta^{-1})$ then for any unit vector $x \in \mathbb{R}^d$,

$$\Pr\left[\|Ax\|_2^2 \in 1 \pm \varepsilon\right] \ge 1 - \delta.$$

Proof of Thm using Lemma: Set $\delta = 1/n^3$, $k = O(\varepsilon^{-2} \log n)$.

For each x_i, x_j, apply Lemma on $\frac{x_i - x_j}{\|x_i - x_j\|^2}$:

$$\left\|A\left(\frac{x_i - x_j}{\|x_i - x_j\|}\right)\right\|^2 \in 1 \pm \varepsilon \quad \text{w.p.} \ge 1 - \frac{1}{n^3}.$$

$$\left\|A\left(\frac{x_i - x_j}{\|x_i - x_j\|}\right)\right\| \in 1 \pm \varepsilon \quad \text{w.p.} \geq 1 - \ldots n^3.$$

$$\hookrightarrow = \left\| \frac{A(x_i - x_j)}{\|x_i - x_j\|} \right\|^2 = \frac{\|A(x_i - x_j)\|^2}{\|x_i - x_j\|^2} = \frac{\|Ax_i - Ax_j\|^2}{\|x_i - x_j\|^2}.$$

Gaussian r.v.'s

Def: The Gaussian with mean μ and variance σ^2 has p.d.f.

$$f(x) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma}}_{\text{normalization factor s.t. } \int_{-\infty}^{\infty} f(x) = 1} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

E.g. if $\mu = 0$ and $\sigma^2 = 1$, then $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

Prop: If $G_1 \sim N(\mu_1, \sigma_1^2)$, $G_2 \sim N(\mu_2, \sigma_2^2)$ indep, then

① [Scaling] $cG \sim N(c\mu_1, c^2\sigma_1^2)$

③ [Addition] $G_1 + G_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

expected. Surprising thing is it's still Gaussian

Intuition of Lemma: what does "$\|Ax\|^2 \in 1 \pm \varepsilon$" mean?

Suppose $x = e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Then

$$Ax = \begin{bmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{k1} \end{bmatrix} \sim \frac{1}{\sqrt{k}} \begin{bmatrix} N(0,1) \\ N(0,1) \\ \vdots \\ N(0,1) \end{bmatrix} \sim \begin{bmatrix} N(0,1/k) \\ N(0,1/k) \\ \vdots \\ N(0,1/k) \end{bmatrix}.$$

So $\|Ax\|^2 \sim \underbrace{N(0,1/k)^2 + \cdots + N(0,1/k)^2}_{k \text{ of them}}$

$$\mathbb{E}\|Ax\|^2 \sim k \cdot \mathbb{E}\, N(0,1/k)^2$$
$$= k \cdot \mathrm{Var}(0,1/k) = k \cdot 1/k = 1.$$

$= k \cdot \mathrm{Var}(0, 1/k) = k \cdot 1/k = 1.$

So expectation is correct. Concentration?

$\sim \frac{1}{k} \mathbb{E}\, N(0,1)^2.$

Hoeffding: Sum of k coinflips is within $\sqrt{k \log n}$

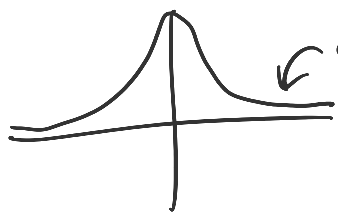
$\Rightarrow$ average is within $\sqrt{\log n / k}$

$k \sim \varepsilon^{-2} \log n$: within ε.

Unfortunately $N(0,1)^2$ unbounded, not in $[0,1]$ so can't use Hoeffding.

But $N(0,1)$ has "thin tails"

so does $N(0,1)^2$.



$\sim e^{-x^2/2}$, faster than exponential decay

Proof of Lemma:

Step 1: rotational symmetry of Gaussian

Write $Ax = \frac{1}{\sqrt{k}} \begin{bmatrix} — N(0,1) — \\ \vdots \\ — N(0,1) — \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} = \frac{1}{\sqrt{k}} \begin{bmatrix} ? \\ \\ \end{bmatrix}$ — $x_1 N(0,1) + x_2 N(0,1) + \cdots$

$$\begin{aligned} \text{Each entry} &\sim x_1 N(0,1) + x_2 N(0,1) + \cdots + x_d N(0,1) \\ &\sim N(0, x_1^2) + N(0, x_2^2) \quad \cdots \quad N(0, x_d)^2 \quad \text{[scaling]} \\ &\sim N(0, x_1^2 + \cdots + x_d^2) \quad \text{[addition]} \\ &\sim N(0, 1). \end{aligned}$$

So $Ax \sim \frac{1}{\sqrt{k}} \begin{bmatrix} N(0,1) \\ \vdots \\ N(0,1) \end{bmatrix}$

concentrates

$\sqrt{k} \begin{bmatrix} \vdots \\ N(0,1) \end{bmatrix}$

To show: $\|Ax\|^2 = \frac{1}{k}\left(N(0,1)^2 + \dots + N(0,1)^2\right)$ concentrates $\in 1 \pm \varepsilon$.

Step 2: Markov on $\mathbb{E}e^{tZ}$.

Let $Z = \frac{1}{k} Y_i^2$ where each $Y_i \sim N(0,1)$ indep.

$$\Pr[Z \geq 1+\varepsilon] = \Pr\left[e^{tkZ} \geq e^{tk(1+\varepsilon)}\right]$$

$$\leq \mathbb{E}e^{tkZ} / e^{tk(1+\varepsilon)} \quad \text{[Markov]}$$

$$= \mathbb{E}\prod_i e^{tY_i^2} / e^{tk(1+\varepsilon)}$$

$$= \prod_i \mathbb{E}e^{tY_i^2} / e^{tk(1+\varepsilon)}$$

Calculate $\mathbb{E}e^{tY_i^2} = \int_{g=-\infty}^{\infty} e^{tg^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-g^2/2}\, dg$

$$= \frac{1}{\sqrt{2\pi}} \int_{g=-\infty}^{\infty} e^{tg^2 - g^2/2}\, dg$$

$$= \frac{1}{\sqrt{2\pi}} \int_{g=-\infty}^{\infty} e^{-\frac{1-2t}{2} g^2}\, dg$$

Substitute $z = \sqrt{1-2t}\, g$, $dz = \sqrt{1-2t}\, dg$:

$$= \underbrace{\frac{1}{\sqrt{2\pi}} \int_{z=-\infty}^{\infty} e^{-z^2/2}}_{=1} \frac{dz}{\sqrt{1-2t}} = \frac{1}{\sqrt{1-2t}}$$

$$\frac{1}{\sqrt{2\pi}} \underbrace{\int_{z=-\infty}}_{=1} \qquad \sqrt{1-2t} \qquad \sqrt{1-2t}$$

$$= \prod_i \frac{1}{\sqrt{1-2t}} \Big/ e^{tk(1+\varepsilon)}$$

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

Focus on $\underbrace{e^{t(1+\varepsilon)}}_{\geq 1} \underbrace{\sqrt{1-2t}}_{\leq 1}$, want > 1.

$$e^{t(1+\varepsilon)} = 1 + t(1+\varepsilon) + O(t^2)$$
$$= 1 + t + t\varepsilon + O(t^2)$$

$$\sqrt{1-2t} = 1 - t - O(t^2)$$

$$e^{t(1+\varepsilon)}\sqrt{1-2t} = (1+t+t\varepsilon+O(t^2))(1-t-O(t^2))$$
$$= 1 - t^2 + \underbrace{t\varepsilon}_{\text{dominate?}} \pm O(t^2).$$

Set $t = (\text{small const}) \cdot \varepsilon$: $e^{t(1+\varepsilon)}\sqrt{1-2t} = 1 + \Theta(\varepsilon^2)$

In fact, set $t = \varepsilon/4$: $\geq e^{\varepsilon^2/8}$

$$\leq e^{-\varepsilon^2 k/8}.$$

Set $k = \frac{8}{\varepsilon^2} \ln \frac{2}{\delta}$:

$$\Pr[Z \geq 1+\varepsilon] \leq e^{-\ln \delta/2} = \frac{\delta}{2}.$$

Lower tail: $\Pr[Z \leq 1-\varepsilon] \leq \frac{\delta}{2}$

Lower tail: $\Pr[Z \le 1-\varepsilon] \le \frac{\delta}{2}$

Union bound $\Pr[Z \in 1 \pm \varepsilon] \le \delta.$

What about other entries in A?

Turns out $\frac{\pm 1}{\sqrt{k}}$ entries also work! (Hoeffding)

Can also adapt this proof.

Def: a r.v. X is (1-) subgaussian if $\mathbb{E}\, e^{tX} \le e^{t^2/2}$ $\forall t > 0.$

Ex: $N(0,1)$ is subgaussian:

$$\mathbb{E}\, e^{tN(0,1)} = \frac{1}{\sqrt{2\pi}} \int_{g=-\infty}^{\infty} e^{tg} e^{-g^2/2}\, dg$$

$$= \frac{1}{\sqrt{2\pi}} \int_{g=-\infty}^{\infty} e^{-(g-t)^2/2 + t^2/2}\, dg$$

$$= \frac{1}{\sqrt{2\pi}} e^{t^2/2} \underbrace{\int_{g=-\infty}^{\infty} e^{-(g-t)^2/2}\, dg}_{=\sqrt{2\pi}} = e^{t^2/2}.$$

Ex: ± 1 is subgaussian:

$$\mathbb{E}\, e^{t(\pm 1)} = \frac{e^{+t} + e^{-t}}{2} = \cosh t = 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \dots \le e^{t^2/2}$$

Lemma: If X is subgaussian and $Y \sim N(0,1)$, then
$\mathbb{E}\, e^{tX^2} \le \mathbb{E}\, e^{tY^2}$ $\forall t > 0.$