Today

1) Metric embeddings

2) Random Projections

3) Johnson-Lindenstrauss Lemma

4) Application of (3)

Recall

The Metric Framework → Only for NN Con onit

i) Find a metric in your Problem

ii) Find Structure in your metric

iii) Use metric structure to solve problem

Concentration Framework

a) Show (*) true if all RVs near $\mathbb{E}$

b) Concentration: one RV near $\mathbb{E}$ w/ high Pr

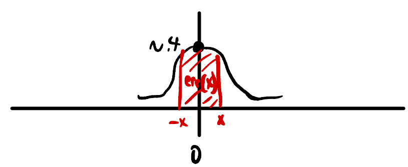
c) Union Bound: all RVs at $\mathbb{E}$ w/ high Pr

Fact: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Gaussians RV

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\underset{\mathbb{E}\ \ \mathrm{var}}{Z \sim N(0,1)} \rightarrow \text{"Standard Gaussian"}$$



Dfn. $N(\mu, \sigma^2) = \mu + \sigma Z$ where $Z \sim N(0,1)$ is a "non-Standard Gaussian"

Fact: $(N(0,1), N(0,1), \ldots)$ in a uniformly random direction ("rotational symmetry")

Fact: $X \sim N(0, a^2)$, $Y \sim N(0, b^2)$, independent $\rightarrow X + Y \sim N(0, a^2 + b^2)$

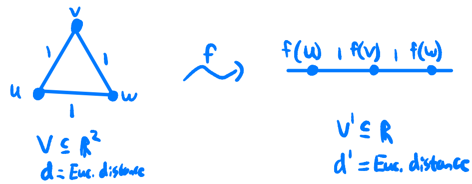
("Σ_i of Gaussians is Gaussian")

Metric Embeddings: how can we approximate a complex metric w/ a simple one ((ii) today)

An **embedding** of metric space (V,d) into metric space (V',d') is a function $f: V \to V'$

(V,d) **embeds isometrically** into (V',d') iff $\exists$ embedding f s.t. $d(u,v) = d'(f(u), f(v))$ $\forall u,v \in V$

Isometry not always possible, e.g.



$d(u,v), d(v,w) = 1$
so $d(u,w) = 2 \neq 1$

f has **distortion** α if $d(u,v) \leq d'(f(u), f(v)) \leq \alpha \cdot d(u,v)$ $\forall u,v \in V$

E.g. $\alpha = 2$ above

Ideally: α small + (V',d') simpler/more structured than (V,d)

Theorem: given m points $V \subseteq \mathbb{R}^n$, $\varepsilon > 0$, $\exists$ a linear f that embeds (V,d) into (V',d) for $V' \subseteq \mathbb{R}^k$ w/ distortion $\alpha = \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$ where $k = O(\log m / \varepsilon^2)$ $\ll n$

$$\overset{??}{\sqrt{\frac{e^{\varepsilon}}{e^{-\varepsilon}}}} = \sqrt{e^{2\varepsilon}} = e^{\varepsilon} \approx 1+\varepsilon$$

JL Lemma: $\exists$ random linear $\tilde{F}: \mathbb{R}^n \to \mathbb{R}^k$ for $k = 24 \cdot \frac{\ln m}{\varepsilon^2}$ s.t. for any m^2 unit ($\|\cdot\| = 1$) vectors $\bar{W}$ it holds that

$$\|\tilde{F}(w)\|^2 \in [1-\varepsilon, 1+\varepsilon] \quad \forall w \in \bar{W}$$

except w/ $\Pr \leq \frac{2}{m}$

Claim: JL Lemma $\to$ theorem

Proof idea: let $\bar{W} := \left\{\frac{x-y}{\|x-y\|} : x, y \in V\right\}$

Skip, come back if time

Let $\tilde{F}$ be as in JL, $\alpha = \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$ and $f = \frac{\tilde{F}}{\sqrt{1-\varepsilon}}$

f has distortion α

$\Updownarrow$

$\|u-v\| \leq \|f(u) - f(v)\| \leq \alpha \|u-v\| \quad \forall u,v \in V$

$\Updownarrow$

$\|u-v\| \leq \|f(u-v)\| \leq \alpha \cdot \|u-v\|$

$\Updownarrow$

$\|z\| \leq \|f(z)\| \leq \alpha \cdot \|z\| \quad \forall z \in V-V$

$\Updownarrow$

$1 \leq \left\|f\left(\frac{z}{\|z\|}\right)\right\| \leq \alpha \quad \forall z \in V-V$

$\Updownarrow$

$1 \leq \|f(w)\|^2 \leq \alpha^2 \quad \forall w = \frac{z}{\|z\|} \; \forall z \in V-V$

$\Updownarrow$

$1-\varepsilon \leq (1-\varepsilon)\|f(w)\|^2 \leq 1+\varepsilon \quad \forall w = \frac{z}{\|z\|} \; \forall z \in V-V$

$\Updownarrow$

$1-\varepsilon \leq \|\sqrt{1-\varepsilon} \cdot f(w)\|^2 \leq 1+\varepsilon \quad \forall w = \frac{z}{\|z\|} \; \forall z \in V-V$

$\Updownarrow$

$1-\varepsilon \leq \|\tilde{F}(w)\|^2 \leq 1+\varepsilon \quad \forall w = \frac{z}{\|z\|} \; \forall z \in V-V$

$\to$ holds except w/ $\Pr \leq \frac{2}{m}$ so thm. follows by probabilistic method

Will also see all poly-time computable

χ^2 Distributions + Concentration

X is a **Chi-squared RV** w/ k degrees of freedom if

$$X = \sum_{i=1}^{k} z_i^2$$

where each $z_i \sim N(0,1)$ independently

Notated $X \sim \chi_k^2$

Claim: $\mathbb{E}[X] = k$ for $X \sim \chi_k^2$

Have $\mathbb{E}[z_i^2] = Var(z_i) + \mathbb{E}[z_i]^2 = 1 + 0$

Claim follows by LoE

Can't apply Chernoff for concentration b/c not $\in \{0,1\}$ (or even bounded)

Nonetheless, similar proof works

Can get $-O(\varepsilon)$ in exponent instead of $-O(\varepsilon^2)$ $\forall \varepsilon$

Skip, come back if time

Claim: $\Pr(|X - \mathbb{E}[X]| \geq \varepsilon \cdot k) \leq 2 \cdot \exp(-\varepsilon^2 \cdot k/8)$ for $X \sim \chi_k^2$ $(\varepsilon \in (0,1))$

Will prove upper tail; lower tail symmetric; Claim follows by union bound

Let $t = \varepsilon/4$ so $t \in (0, \frac{1}{4})$ and $3t + \varepsilon t = \frac{3}{4}\varepsilon + \frac{1}{4}\varepsilon^2 \geq \varepsilon^2/8 \;\; \forall \varepsilon \in (0,1)$

$$\Pr(X - \mathbb{E}[X] \geq \varepsilon \cdot k) = \Pr(X \geq (1+\varepsilon)\cdot k) \underset{e^{tx} \text{ non-decreasing}}{\leq} \Pr\left(e^{t\cdot X} \geq e^{(1+\varepsilon)\cdot tk}\right) \underset{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{t \cdot X}]}{e^{(1+\varepsilon)\cdot tk}} \underset{z_i \text{ ind.}}{=} \frac{\prod_i \mathbb{E}[e^{t \cdot z_i^2}]}{e^{(1+\varepsilon)\cdot tk}}$$

Now calculate $\mathbb{E}[e^{t \cdot z_i^2}] = \frac{1}{\sqrt{2\pi}} \int_a e^{ta^2} \cdot e^{-a^2/2} \, da = \frac{1}{\sqrt{2\pi}} \int_a e^{-a(\frac{1}{2}-t)} = \frac{1}{\sqrt{2(\frac{1}{2}-t)}}$

$= \frac{1}{\sqrt{1-2t}}$

Gaussian integral: $\left(\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}\right)$

Plugging $\mathbb{E}[e^{t \cdot z_i^2}]$ in, get $\Pr(X \geq (1+\varepsilon)\cdot k) \leq \left(\frac{1}{e^{(1+\varepsilon)\cdot t}\sqrt{1-2t}}\right)^k$

But $\frac{1}{e^{(1+\varepsilon)t}\sqrt{1-2t}} = \exp(-t - \frac{1}{2}\ln(1-2t) - \varepsilon t) \leq \exp(-3t - \varepsilon t) \leq \exp(-\varepsilon^2/8)$

$e^{-x} \leq 1 - .5x$ for $x \in (0,1)$ so $-\ln(1-2t) \leq -4t$

Above choice of t

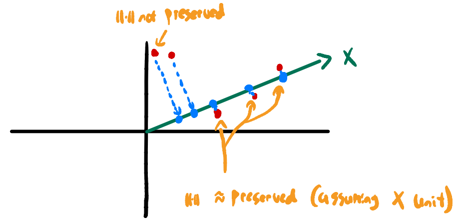
Combining above gives result

Random Projection

Simplest case: $k=1$

Let $X = (X_1, X_2, \dots, X_m)$ where $X_i \sim N(0,1)$ $\forall i$ (a vector in a uniformly random direction)

Let $f : \mathbb{R}^m \to \mathbb{R}$ be $f(w) = \langle w, X \rangle$



Pick random direction X

w mapped to how far along X $\text{Proj}(w \to x)$ is

Claim: $f(w)^2 \sim \chi_1^2$ $\forall$ unit w

$\|w\| = 1 \to \|w\|^2 = 1 \to \sum_i w_i^2 = 1$

So $f(w) = \underbrace{\sum_i \underbrace{w_i X_i}_{\sim N(0, w_i^2) \text{ by non-standard Gaussian definition}}}_{\sim N(0, w_1^2 + w_2^2 + \dots + w_m^2) \text{ by } \sum \text{Gaussians} = \text{Gaussian}}$

$= N(0,1)$ b/c $\|w\|=1$

So $(f(w))^2 \sim \chi_1^2$

Corollaries

1) $\mathbb{E}[f(w)^2] = 1$ $\forall$ unit w → "Random Projection preserves unit vector length in $\mathbb{E}$"

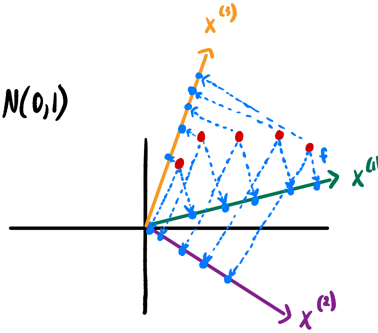
2) $\Pr(|f(w)^2 - 1| \geq \varepsilon) \leq 2 \cdot \exp(-\varepsilon^2/8)$ $\forall$ unit w → Vector length even concentrates under random projection

Problem: Want to UB over m^2 vectors so need $\exp(-\varepsilon^2) \approx \frac{1}{m^2} \leftrightarrow \varepsilon \approx \sqrt{\log m}$

but then would know $f(w)^2 \in [1 - \sqrt{\log n}, 1 + \sqrt{\log n}] \neq [1-\varepsilon, 1+\varepsilon]$

Solution: increase $\mathbb{E}$ to $\approx \log m$ to get better concentration by repeated trials

Johnson-Lindenstrauss Lemma



$w \rightsquigarrow (\langle w, x^{(1)}\rangle, \langle w, x^{(2)}\rangle, \langle w, x^{(3)}\rangle, \dots)$

Let $X^{(i)} \sim (X_1^{(i)}, X_2^{(i)}, \dots, X_m^{(i)})$ where each $X_j^{(i)} \sim N(0,1)$

Let $f_i(w) := \langle x^{(i)}, w\rangle$

Let $F(w) := (f_1(w), f_2(w), \dots, f_k(w))$

$F(w) = Aw$ where $A = \begin{pmatrix} - x^{(1)} - \\ \vdots \\ - x^{(k)} - \end{pmatrix}$ so F linear

Claim: $\|F(w)\|^2 \sim \chi_k^2$ $\forall$ unit w

$\|F(w)\|^2 = \sum_i (f_i(w))^2$ and $f_i(w)^2 \sim \chi_1^2$ by previous Claim

Corollaries

1) $\mathbb{E}\left[\|F(w)\|^2\right] = k$ $\forall$ unit w

2) $\Pr\left(\left|\|F(w)\|^2 - k\right| \geq \varepsilon \cdot k\right) \leq 2 \cdot \exp(-\varepsilon^2 \cdot k / 8)$ $\forall$ unit w

$\to$ $k \approx \log m \to$ good Concentration

Let $\tilde{F}(w) := \frac{1}{\sqrt{k}} \cdot F(w)$

$\to$ Scale down so unit vector mapped to unit vector

$\to$ Trivially Poly-time computable (even w/o knowing Points)

$\hookrightarrow$ Linear since F linear

Proof of JL Lemma

Note $\mathbb{E}\left[\|\tilde{F}(w)\|^2\right] = \mathbb{E}\left[\|\frac{1}{\sqrt{k}} F(w)\|^2\right] = \frac{1}{k} \mathbb{E}\left[\|F(w)\|^2\right] = 1$ (a)

$\|\tilde{F}(w)\|^2 \notin [1-\varepsilon, 1+\varepsilon]$ only if $\|F(w)\|^2 \notin [(1-\varepsilon)\cdot k, (1+\varepsilon)k]$ only if $\left|\|F(w)\|^2 - k\right| \geq \varepsilon \cdot k$

But $\Pr\left(\left|\|F(w)\|^2 - k\right| \geq \varepsilon k\right) \leq 2 \cdot \exp(-3 \ln m) = \frac{2}{m^3}$ for $k = 24 \cdot \frac{\ln m}{\varepsilon^2}$ (b)

By union bound, $\tilde{F}(w) \in [1-\varepsilon, 1+\varepsilon]$ $\forall$ m^2 w except w/ $\Pr \leq \frac{2}{m}$ (c)

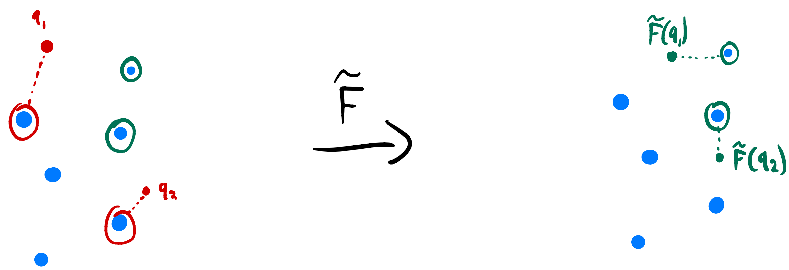
Application of JL

$\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$ - NN - Search: initially given $X \subseteq \mathbb{R}^n$, w/ $|X| = m$ (i)

repeatedly given queries $q_1, q_2, \dots$

$\forall\, q_i$, return $x \in X$ s.t. $d(q_i, x) \le \sqrt{\frac{1+\varepsilon}{1-\varepsilon}} \cdot \overbrace{d(q_i, X)}^{:= \min_{x \in X} d(q_i, x)}$

as fast as possible



Naive solution: $O(n \cdot m)$ time per query

JL Solution: $O\left(\frac{\log m}{\varepsilon^2} \cdot m\right)$ per query

let $\tilde{F} : \mathbb{R}^n \to \mathbb{R}^k$ be linear embedding for X

for $k = O\left(\frac{\log m}{\varepsilon^2}\right)$ so w/ distortion $\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$ from JL

let $Y := \{\tilde{F}(x) : x \in X\}$ (ii)

For query q_i,

return $x \in X$ s.t. $\tilde{F}(x) = y$ where $y = \arg\min_{y \in Y} d(y, \tilde{F}(q_i))$ (iii)

Easy to verify result is (approximately) correct

Takes $O\left(\frac{\log m}{\varepsilon^2} \cdot (m + n)\right)$ per query

↳ Can even reduce to $\tilde{O}\left(\frac{1}{\varepsilon^2}\right)$ w/ a little more work!