Recap:

ℓ_2 regression: $\underset{x}{\text{argmin}} \|Ax-b\|_2^2$ A: $n \times d$ matrix, $b \in \mathbb{R}^n$, $n \gg d$

$O(nd^2)$ via normal eq^n

relaxed problem $\|Ax-b\|_2 \leq (1+\varepsilon) \|Ax^*-b\|_2$

failure prob. δ

1. Sample $r \times n$ matrix S $r \ll n$
2. Compute SA, Sb
3. Solve $\underbrace{\underset{x}{\text{argmin}} \|(SA)x - (Sb)\|_2}_{\text{smaller regression}}$

ℓ_2- subspace embedding:

$(1 \pm \varepsilon)$ ℓ_2 subspace embedding for col. space of $n \times d$ matrix A is a matrix S s.t. $\forall x \in \mathbb{R}^d$ $\|SAx\|_2^2 = (1 \pm \varepsilon) \|Ax\|_2^2$

U: orthonormal basis for cols of A

$\|SUy\|_2^2 = (1 \pm \varepsilon) \|Uy\|_2^2 = (1 \pm \varepsilon) \|y\|_2^2$

operator norm $\|I_d - U^T S^T S U\|_2 \leq \varepsilon$

Oblivious ℓ_2- subspace embedding

distrib^n on $r \times n$ matrices S

w. prob $1-\delta$ for any fixed $n \times d$ matrix A

$S \sim \Pi$ is a $(1 \pm \varepsilon)$ ℓ_2 subspace embedding for A

Def^n Random matrix $S \in \mathbb{R}^{k \times n}$ is a JL-transform JLT(ε, δ, f) if w. prob $1-\delta$ for any f element subset $V \subset \mathbb{R}^n$,

$\forall v, v' \in V$, $|\langle Sv, Sv' \rangle - \langle v, v' \rangle| \leq \varepsilon \|v\|_2 \|v'\|_2$

Thm: $S = \frac{1}{\sqrt{k}} R \in \mathbb{R}^{k \times n}$ R_{ij} indpt normal, $k = \Omega\left(\frac{\log(f/\delta)}{\varepsilon^2}\right)$
then S is a JLT(ε, δ, f)

Subspace Embedding Property

$T = \{ y \in \mathbb{R}^n \mid y = Ax \text{ for some } x \in \mathbb{R}^d, \|y\|_2 = 1 \}$

Need : $\|Sy\|_2^2 = (1 \pm \varepsilon)\|y\|_2^2 = 1 \pm \varepsilon \quad \forall y \in T$

ε-net : Finite subset $N \subset T$ so that $\forall y \in T$

$\exists\, w \in N, \; \|y - w\|_2 \le \frac{1}{2}$

If $\forall\, w, w' \in N, \langle Sw, Sw' \rangle = \langle w, w' \rangle \pm \varepsilon$

then $\forall\, y \in T \quad \|Sy\|_2^2 = 1 \pm O(\varepsilon)$

<u>Lemma</u> : $\exists$ γ-net N of T, $|N| \le \left(1 + \frac{2}{\gamma}\right)^d$

Pf : $t = \operatorname{rank}(A)$

$T = \{ y \in \mathbb{R}^n \mid y = Ux, \; x \in \mathbb{R}^t, \; \|y\|_2 = 1 \}$

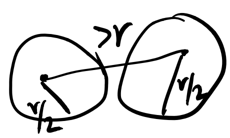
U: orthonormal basis for A

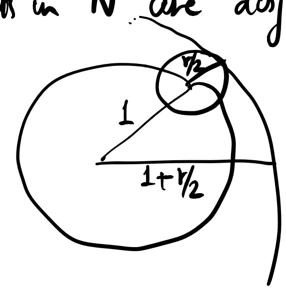
(Intuition) Net for S^{t-1} will give net for T
image of S^{t-1} under U

Choose maximal set N' of points in S^{t-1}

so that no 2 points within dist γ of each other.

Balls of radius $\gamma/2$ centered at points in N' are disjoint





$$|N'| \le \frac{\left(1 + \frac{\gamma}{2}\right)^t}{\left(\frac{\gamma}{2}\right)^t} = \left(1 + \frac{2}{\gamma}\right)^t$$

$N = \{ y \in \mathbb{R}^n \mid y = Ux \text{ for some } x \in N' \}$

If $\exists\, Ux \in T$ s.t. $\forall\, y = Uz \in N, \; z \in N'$

$$\|y - Ux\|_2 = \|Uz - Ux\|_2 > \gamma$$

$\Rightarrow x \in S^{t-1}$, s.t. $\forall z \in N', \|x - z\|_2 > \gamma$ contradiction

Set $V = N$, $f = 5^d$ in JLT theorem

$S = \frac{1}{\sqrt{k}} R$ $\quad R: k \times n \quad R_{ij}$: indpt std. normal

$k = \Theta\left((d + \log(1/\delta)) / \varepsilon^2 \right)$

Oblivious ℓ_2 - subspace embedding

Optimal # rows : d/ε^2

bottleneck : time to compute SA $\quad O(kdn) = O(nd^2/\varepsilon^2)$

$\downarrow$

$nnz(A)$ $\quad O\left(\frac{d}{\varepsilon^2} \cdot nnz(A)\right)$

[Dasgupta, Kumar, Sarlos] [Kane, Nelson]

$\frac{\log(f/\delta)}{\varepsilon}$ non-zero entries per col

$O\left(nnz(A) \cdot \frac{d}{\varepsilon}\right)$

near-optimal.

[Clarkson-Woodruff] $S.A$ in $O(nnz(A))$

Sparse Embedding Matrix

S: CountSketch matrix $\quad r \times n$

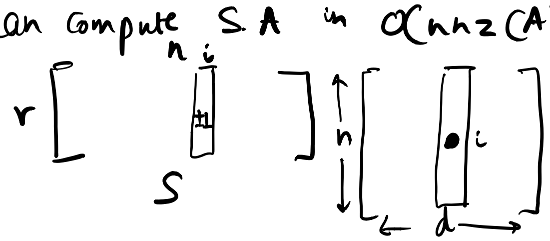
For each of n cols S_{*i}

Choose uniform random row $h(i) \in \{1, 2 .. r\}$

Choose random elt of $\{\pm 1\}$: $\sigma(i)$

$$S_{h(i),i} = \sigma(i)$$
$$S_{ji} = 0 \quad \forall j \neq h(i)$$

Can compute $S.A$ in $O(nnz(A))$



Thm S: count sketch matrix

$r = O\left(\frac{d^2}{\varepsilon^2 \delta}\right)$ rows

For any fixed A, S is $(1 \pm \varepsilon)$-ℓ_2 subspace embedding for A w. prob $1-\delta$

h: pairwise indpt

σ: 4-wise indpt.

Proof Sketch:

$$\Pr\left[\| I_d - U^T S^T S U\|_2 \geq \varepsilon\right] = \Pr\left[\| I_d - U^T S^T S U\|_2^{\ell} \geq \varepsilon^{\ell}\right] \quad \ell \text{ even}$$

$$\leq \frac{1}{\varepsilon^{\ell}} E\left[\| I_d - U^T S^T S U\|_2^{\ell}\right]$$

$$\leq \frac{1}{\varepsilon^{\ell}} E\left[tr\left((I_d - U^T S^T S U)^{\ell}\right)\right]$$

Simpler Proof [Nguyen]

Approx. Matrix Multiplication

Defn C is ε-approx matrix product of A, B if

$$\|A^T B - C\|_F \leq \varepsilon \|A\|_F \|B\|_F$$

Maintain sketch SA, SB, S: $r \times n$ matrix

Want $\mathbb{E}[A^T S^T S B] = A^T B$

Defn: Distrib^n D on $S \in \mathbb{R}^{k \times d}$ satisfies $(\varepsilon, \delta, \ell)$-JL moment property if

$\forall x \in \mathbb{R}^d$, $\|x\|_2 = 1$ $\quad \mathbb{E}\left[\left(\|Sx\|_2^2 - 1\right)^{\ell}\right] \leq \varepsilon^{\ell} \delta$

Defn: For scalar random var X

$\|X\|_p = \mathbb{E}\left[|X|^p\right]^{1/p}$

$\|\cdot\|_p$ is a metric: $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$

Lemma: Let $\ell \geq 2$, $\varepsilon, \delta \in (0, 1/2)$

D: distrib^n that satisfies $(\varepsilon, \delta, \ell)$- JL moment property

For A, B with d rows

$$P_{S\sim D}\left[\|A^T S^T S B - A^T B\|_F > 3\varepsilon \|A\|_F \|B\|_F \right] \leq \delta$$

Pf: For unit vectors $x, y \in \mathbb{R}^d$

$$\langle Sx, Sy\rangle = \frac{1}{2}\left(\|Sx\|_2^2 + \|Sy\|_2^2 - \|S(x-y)\|_2^2\right)$$

$$\|\langle Sx, Sy\rangle - \langle x, y\rangle\|_\ell = \frac{1}{2}\left\| \left(\|Sx\|_2^2 - 1\right) + \left(\|Sy\|_2^2 - 1\right) - \left(\|S(x-y)\|_2^2 - \|x-y\|_2^2\right)\right\|_\ell$$

$$\leq \frac{1}{2}\left(\|\|Sx\|_2^2 - 1\|_\ell + \|\|Sy\|_2^2 - 1\|_\ell + \|\|S(x-y)\|_2^2 - \|x-y\|_2^2\|_\ell \right)$$

$$\leq \frac{1}{2}\left(\varepsilon\delta^{1/\ell} + \varepsilon\delta^{1/\ell} + \|x-y\|_2^2 \varepsilon \delta^{1/\ell}\right)$$

$$\leq 3\varepsilon\delta^{1/\ell}$$

For arbitrary x, y

$$\|\langle Sx, Sy\rangle - \langle x, y\rangle\|_\ell \leq 3\varepsilon\delta^{1/\ell} \|x\|_2 \|y\|_2$$

$(i,j)^{th}$ entry of $A^T B$ is $(\underset{i^{th}\text{ col}}{A^i}, \underset{j^{th}\text{ col}}{B^j})$

$$\left\|\|A^T S^T S B - A^T B\|_F^2\right\|_{\ell/2} \leq \sum_{i,j} \left\|\left(\langle SA^i, SB^j\rangle - \langle A^i, B^j\rangle\right)^2\right\|_{\ell/2}$$

$$\leq (3\varepsilon\delta^{1/\ell})^2 \sum_{i,j} \|A^i\|_2^2 \|B^j\|_2^2$$

$$= (3\varepsilon\delta^{1/\ell})^2 \|A\|_F^2 \|B\|_F^2$$

$$\mathbb{E}\left[\|A^T S^T S B - A^T B\|_F^\ell\right] \leq (3\varepsilon\delta^{1/\ell})^\ell \|A\|_F^\ell \|B\|_F^\ell$$

$$P\left[\|A^T S^T S B - A^T B\|_F^\ell > (3\varepsilon)^\ell \|A\|_F^\ell \|B\|_F^\ell\right]$$

$$\leq \frac{1}{(3\varepsilon\|A\|_F\|B\|_F)^\ell} \mathbb{E}\left[\|A^T S^T S B - A^T B\|_F^\ell\right]$$

$$\leq \delta$$

$(\varepsilon, \delta, \ell)$ JL moment $\xrightarrow{\checkmark}$ Approx matrix mult^{n} $\longrightarrow$ $(1 \pm \varepsilon)$ subspace embedding

Count Sketch

to prove

TODO

S: Count Sketch with $\frac{2}{\varepsilon^2 \delta}$ rows.

<u>Claim</u>: S satisfies $(\varepsilon, \delta, 2)$ JL moment property

$h : [d] \to [r]$ 2wise indpt

$\sigma : [d] \to \{\pm 1\}$ 4-wise indpt.

<u>Pf</u>: For any unit vector $x \in \mathbb{R}^d$

to prove

$$E_S\left[\left(\|Sx\|_2^2 - 1\right)^2\right] = \mathbb{E}_S\left[\|Sx\|_2^4 - 2\, E_S\left[\|Sx\|_2^2\right] + 1 \leq \varepsilon^2 \delta\right.$$

$$\mathbb{E}\left[\|Sx\|_2^2\right] = \sum_{i \in [r]} E\left[\left(\sum_{j \in [d]} I_{h(j)=i}\, x_j\, \sigma(j)\right)^2\right]$$

$$= \sum_{i \in [r]} \sum_{j, j' \in [d]} x_j x_{j'}\, \mathbb{E}\left[I_{h(j)=i}\, I_{h(j')=i}\right] \mathbb{E}\left[\sigma(j)\, \sigma(j')\right]$$

$$= \sum_{i \in [r]} \sum_{j \in [d]} \frac{x_j^2}{r}$$

$$= \|x\|_2^2 = 1$$

$$\mathbb{E}\left[\|Sx\|_2^4\right] \leq 1 + \frac{2}{r}$$

Need $r \geq \frac{2}{\varepsilon^2 \delta}$

<u>Proof</u>: (Count Sketch gives ℓ_2-subspace embedding)

S satisfies $(\varepsilon, \delta, 2)$ JL moment property

U orthonormal basis for cols of A

$$P\left[\|U^T S^T S U - \underbrace{U^T U}_{I_d}\|_F > 3\varepsilon \underbrace{\|U\|_F \|U\|_F}_{d}\right] \leq \delta$$

Apply with $\varepsilon' = \varepsilon/d$ $\quad r = \frac{2}{(\varepsilon')^2 \delta} = \frac{2d^2}{\varepsilon^2 \delta}$

$$P\left[\| U^T S^T S U - I_d \|_F > 3\varepsilon \right] \leq \delta$$

$$\Rightarrow P\left[\| U^T S^T S U - I_d \|_2 > 3\varepsilon \right] \leq \delta$$

Woodruff: Sketching As a Tool for Numerical Linear Algebra