JL Lemma, Dimensionality Reduction, and Subspace Embeddings

Lecture 11
September 29, 2020

F_2 estimation in turnstile setting

AMS-ℓ_2-Estimate:

```
Let Y_1, Y_2, ..., Y_n be {-1, +1} random variables that are
    4-wise independent
z ← 0
While (stream is not empty) do
    a_j = (i_j, Δ_j) is current update
    z ← z + Δ_j Y_{i_j}
endWhile
Output z^2
```

Claim: Output estimates $||x||_2^2$ where x is the vector at end of stream of updates.

$$\underset{\pm 1\ \ \pm 1\ \ \cdots\ \ \pm 1}{[Y_1\ Y_2\ \cdots\ Y_n]} \overset{\bar{X}}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}} = z \qquad z^2$$

Analysis

$Z = \sum_{i=1}^{n} x_i Y_i$ and output is Z^2

$$Z^2 = \sum_i x_i^2 Y_i^2 + 2 \sum_{i \neq j} x_i x_j Y_i Y_j$$

and hence

$$\mathbf{E}\left[Z^2\right] = \sum_i x_i^2 = ||x||_2^2.$$

One can show that $Var(Z^2) \leq 2(\mathbf{E}\left[Z^2\right])^2$.

Linear Sketching View

Recall that we take average of independent estimators and take median to reduce error. Can we view all this as a sketch?

AMS-ℓ_2-Sketch:

```
    k = c log(1/δ)/ε²
    Let M be a ℓ × n matrix with entries in {−1, 1} s.t
        (i) rows are independent and
        (ii) in each row entries are 4-wise independent
    z is a ℓ × 1 vector initialized to 0
    While (stream is not empty) do
        a_j = (i_j, Δ_j) is current update
        z ← z + Δ_j M e_{i_j}
    endWhile
    Output vector z as sketch.
```

M is compactly represented via k hash functions, one per row, independently chosen from 4-wise independent hash family.

Geometric Interpretation

Given vector $x \in \mathbb{R}^n$ let M the random map $z = Mx$ has the following features

- $\mathbf{E}[z_i] = 0$ and $\mathbf{E}\left[z_i^2\right] = \|x\|_2^2$ for each $1 \leq i \leq k$ where k is number of rows of M
- Thus each z_i^2 is an estimate of length of x in Euclidean norm
- When $k = \Theta(\frac{1}{\epsilon^2}\log(1/\delta))$ one can obtain an $(1 \pm \epsilon)$ estimate of $\|x\|_2$ by averaging and median ideas

Thus we are able to compress x into k-dimensional vector z such that z contains information to estimate $\|x\|_2$ accurately

$$M: \quad k\left\{\begin{bmatrix} +1 & \pm1 & \cdots & \pm1 \\ \vdots & & & \vdots \end{bmatrix}\right. \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}$$

$$\frac{1}{\sqrt{k}}\|z\|_2 \approx \|x_2\|$$

z_1^2, z_2^2 } [illegible]

z^2 }

$\vdots$

z_k^2 }

Geometric Interpretation

Given vector $x \in \mathbb{R}^n$ let M the random map $z = Mx$ has the following features

- $\mathbf{E}[z_i] = 0$ and $\mathbf{E}\left[z_i^2\right] = \|x\|_2^2$ for each $1 \leq i \leq k$ where k is number of rows of M
- Thus each z_i^2 is an estimate of length of x in Euclidean norm
- When $k = \Theta(\frac{1}{\epsilon^2}\log(1/\delta))$ one can obtain an $(1 \pm \epsilon)$ estimate of $\|x\|_2$ by averaging and median ideas

Thus we are able to compress x into k-dimensional vector z such that z contains information to estimate $\|x\|_2$ accurately

Question: Do we need median trick? Will averaging do?

Distributional JL Lemma

Lemma (Distributional JL Lemma)

Fix vector $x \in \mathbb{R}^d$ *and let* $\Pi \in \mathbb{R}^{k \times d}$ *matrix where each entry* Π_{ij} *is chosen independently according to standard normal distribution* $\mathcal{N}(0,1)$ *distribution. If* $k = \Omega(\frac{1}{\epsilon^2}\log(1/\delta))$*, then with probability* $(1-\delta)$

$$(1-\epsilon)\|x\|_2 \le \left\|\frac{1}{\sqrt{k}}\Pi x\right\|_2 \le (1 + \epsilon)\|x\|_2.$$

Can choose entries from $\{-1, 1\}$ as well.
Note: unlike ℓ_2 estimation, entries of Π are independent.

Letting $z = \frac{1}{\sqrt{k}}\Pi x$ we have projected x from d dimensions to $k = O(\frac{1}{\epsilon^2}\log(1/\delta))$ dimensions while preserving length to within $(1 \pm \epsilon)$-factor.

$$\Pi \begin{bmatrix} - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix} \quad x \in \mathbb{R}^n \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$k \times n$

$\Pi_{ij} \sim \underline{\underline{N}}(0,1)$

$\Pi x = z$

$z \in \mathbb{R}^k$

$\frac{1}{\sqrt{k}} \|z\|_2$

$\simeq (1 \pm \varepsilon) \|x\|_2$

with prob

$\geq (1-\delta)$.

Dimensionality reduction

Theorem (Metric JL Lemma)

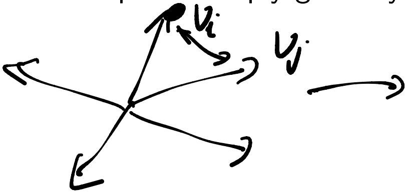
Let $v_1, v_2, \ldots, v_n$ be any n points/vectors in $\mathbb{R}^d$. For any $\epsilon \in (0, 1/2)$, there is linear map $f : \mathbb{R}^d \to \mathbb{R}^k$ where $k \le 8 \ln n/\epsilon^2$ such that for all $1 \le i < j \le n$,

$(1+\varepsilon)$

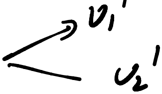
$$(1-\epsilon)||v_i - v_j||_2 \le ||f(v_i) - f(v_j)||_2 \le ||v_i - v_j||_2.$$

Moreover f can be obtained in randomized polynomial-time.

Linear map f is simply given by random matrix Π: $f(v) = \Pi v$.



low dimen



$||\bar{v}_i - \bar{v}_j||_2$
$= ||v_i' - v_j'||_2$

$v_1, v_2, \ldots, \bar{v}_n \quad \in R^d$.

DJL Lemma: If you choose $\Pi \in R^{k \times d}$ for $k = \frac{1}{\varepsilon^2} \ln \frac{1}{\delta}$ then for any fixed vector $\bar{x} \in R^d$

$$\frac{1}{\sqrt{k}} \|\Pi x\|_2 \approx (1 \pm \varepsilon) \|x\|_2$$

$\binom{n}{2}$ vectors $\underline{\bar{v}_i - \bar{v}_j}$ $\quad i \neq j$

If we choose $\delta = \frac{1}{n^3} \Rightarrow k = \frac{c}{\varepsilon^2} \ln \frac{1}{n^3}$
$\approx \frac{c}{\varepsilon^2} \ln n$.

then with prob $\left(1 - \frac{1}{n^3}\right)$

$$\frac{1}{\sqrt{k}} \|\Pi(\bar{v}_i - \bar{v}_j)\|_2 \approx (1 \pm \varepsilon) \|\bar{v}_i - \bar{v}_j\|_2$$

By union bound all $\bar{v}_i - v_j$ vectors are preserved with prob $1 - \binom{n}{2} \cdot \frac{1}{n^3}$
$\geq 1 - \frac{1}{n}$.

Dimensionality reduction

Theorem (Metric JL Lemma)

Let $v_1, v_2, \ldots, v_n$ *be any* n *points/vectors in* $\mathbb{R}^d$*. For any* $\epsilon \in (0, 1/2)$*, there is linear map* $f : \mathbb{R}^d \to \mathbb{R}^k$ *where* $k \le 8 \ln n/\epsilon^2$ *such that for all* $1 \le i < j \le n$*,*

$$(1-\epsilon)||v_i - v_j||_2 \le ||f(v_i) - f(v_j)||_2 \le ||v_i - v_j||_2.$$

Moreover f *can be obtained in randomized polynomial-time.*

Linear map f is simply given by random matrix Π: $f(v) = \Pi v$.

Proof.

Apply DJL with $\delta = 1/n^2$ and apply union bound to $\binom{n}{2}$ vectors $(v_i - v_j)$, $i \neq j$. $\square$

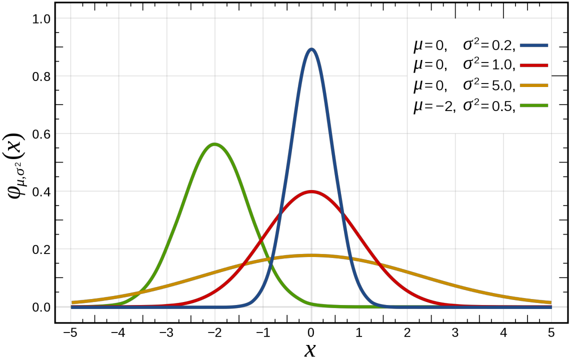
DJL and Metric JL

Key advantage: mapping is *oblivious* to data!

Normal Distribution

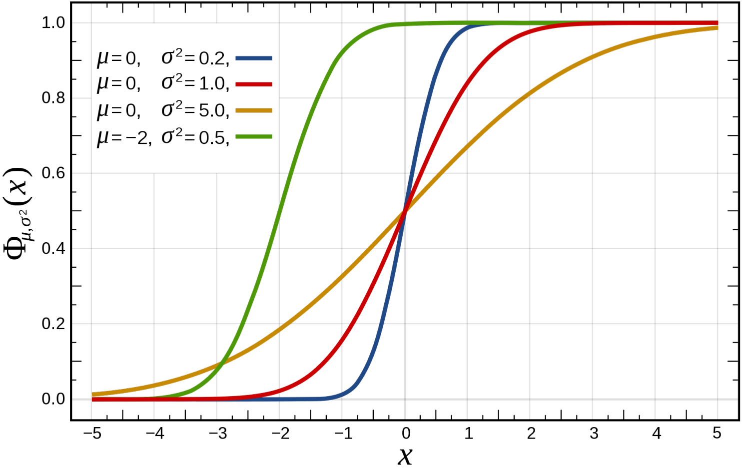
Density function: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Standard normal: $\mathcal{N}(0, 1)$ is when $\mu = 0, \sigma = 1$



Normal Distribution

Cumulative density function for standard normal:
$\boldsymbol{\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{t} e^{-t^2/2}}$ (no closed form)



Sum of independent Normally distributed variables

Lemma

Let X *and* Y *be independent random variables. Suppose* $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ *and* $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$*. Let* $Z = X + Y$*. Then* $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$*.*

Sum of independent Normally distributed variables

Lemma

Let X *and* Y *be independent random variables. Suppose* $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ *and* $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. *Let* $Z = X + Y$. *Then* $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Corollary

Let X *and* Y *be independent random variables. Suppose* $X \sim \mathcal{N}(0, 1)$ *and* $Y \sim \mathcal{N}(0, 1)$. *Let* $Z = aX + bY$. *Then* $Z \sim \mathcal{N}(0, a^2 + b^2)$.

Sum of independent Normally distributed variables

Lemma

Let X and Y be independent random variables. Suppose $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Let $Z = X + Y$. Then $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Corollary

Let X and Y be independent random variables. Suppose $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$. Let $Z = aX + bY$. Then $Z \sim \mathcal{N}(0, a^2 + b^2)$.

Normal distribution is a *stable* distributions: adding two independent random variables within the same class gives a distribution inside the class. Others exist and useful in F_p estimation for $p \in (0, 2)$.

Concentration of sum of squares of normally distributed variables

$\chi^2(k)$ distribution: distribution of sum of k independent standard normally distributed variables
$Y = \sum_{i=1}^{k} Z_i$ where each $Z_i \simeq \mathcal{N}(0, 1)$.

Concentration of sum of squares of normally distributed variables

$\chi^2(k)$ distribution: distribution of sum of k independent standard normally distributed variables

$Y = \sum_{i=1}^{k} Z_i^2$ where each $Z_i \simeq \mathcal{N}(0,1)$. $E[Z_i] = 0$

$\mathbf{E}\left[Z_i^2\right] = 1$ hence $\underline{\underline{\mathbf{E}[Y] =}}\ k$.

Concentration of sum of squares of normally distributed variables

$\chi^2(k)$ distribution: distribution of sum of k independent standard normally distributed variables

$Y = \sum_{i=1}^{k} Z_i$ where each $Z_i \simeq \mathcal{N}(0,1)$.

$\mathbf{E}\left[Z_i^2\right] = 1$ hence $\mathbf{E}[Y] = k$.

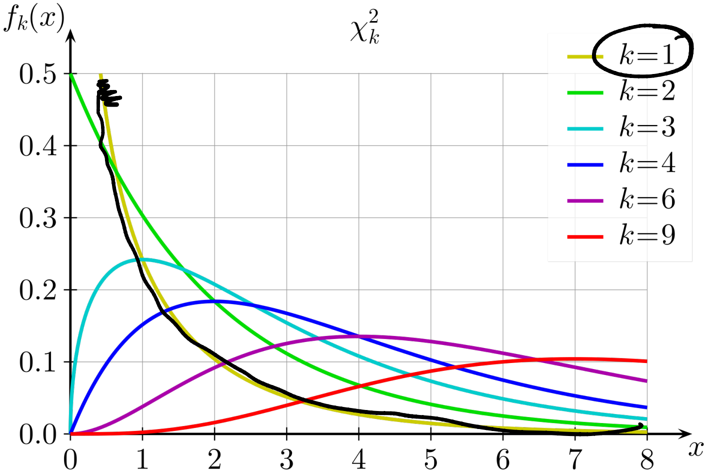
Lemma

Let $Z_1, Z_2, \ldots, Z_k$ be independent $\mathcal{N}(0,1)$ random variables and let $Y = \sum_i Z_i^2$. Then, for $\epsilon \in (0, 1/2)$, there is a constant c such that,

$$\Pr[(1-\epsilon)^2 k \le Y \le (1+\epsilon)^2 k] \ge 1 - 2e^{c\epsilon^2 k}.$$

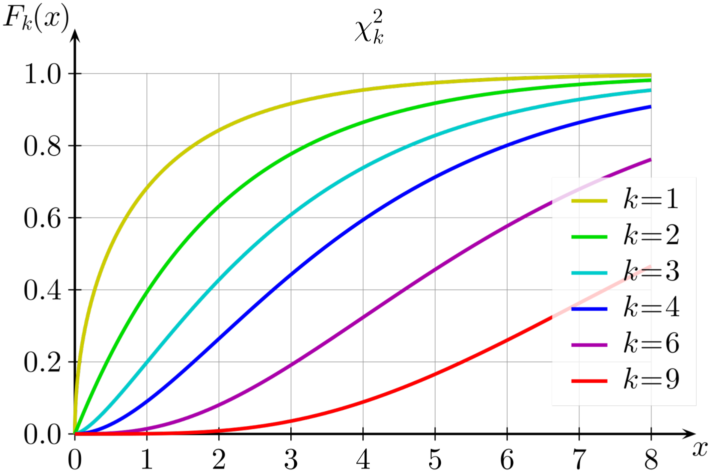
χ^2 distribution

Density function



χ^2 distribution

Cumulative density function



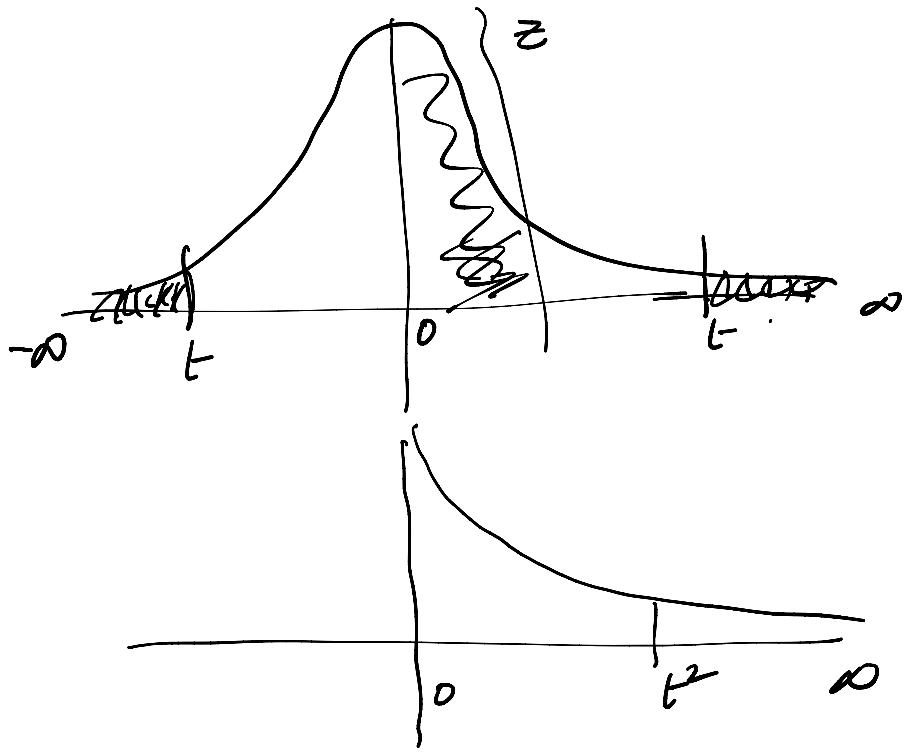
Concentration of sum of squares of normally distributed variables

$\chi^2(k)$ distribution: distribution of sum of k independent standard normally distributed variables

Lemma

Let $Z_1, Z_2, \ldots, Z_k$ be independent $\mathcal{N}(0,1)$ random variables and let $Y = \sum_i Z_i^2$. Then, for $\epsilon \in (0, 1/2)$, there is a constant c such that, $\Pr[(1-\epsilon)^2 k \leq Y \leq (1+\epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k}$.

Recall Chernoff-Hoeffding bound for *bounded* independent non-negative random variables. Z_i^2 is not bounded, however Chernoff-Hoeffding bounds extend to sums of random variables with exponentially decaying tails.



z
$-\infty$
t
0
t
∞
0
t^2
∞

Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$x \in \mathbb{R}^n$

$Z_i = \sum_{j=1}^{n} \Pi_{ij} x_i$

$$\Pi \begin{array}{c|c} & 1 \quad 2 \quad \cdots \quad n \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ k \end{matrix} & \begin{bmatrix} - & - & - & - & - \\ - & - & - & - \\ - \end{bmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{bmatrix}$$

- $Z_i \sim \mathcal{N}(0,1)$

$\Pi \in \mathbb{R}^{k \times n}$

$$(1-\varepsilon)\|\bar{x}\|_2 \leq \left\| \frac{1}{\sqrt{k}} \Pi \bar{x} \right\|_2 \leq (1+\varepsilon)\|\bar{x}\|_2$$

with prob $(1-\delta)$ where

$\Pi_{ij} \sim N(0,1)$.

$k \geq \frac{1}{\varepsilon^2} \ln \frac{1}{\delta}$

Proof of DJL Lemma

Without loss of generality assume $\boxed{\|x\|_2 = 1}$ (unit vector)

$Z_i = \sum_{j=1}^n \Pi_{ij} x_j$

$Z_i = \sum_{j=1}^{n} \Pi_{ij} x_j$

$\Sigma x_i^2 = 1$

- $Z_i \sim \mathcal{N}(0, 1)$
- Let $Y = \sum_{i=1}^k Z_i^2$. Y's distribution is $\chi^2_{(k)}$ since $Z_1, \ldots, Z_k$ are iid.

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{bmatrix}$$

Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$Z_i = \sum_{j=1}^{n} \Pi_{ij} x_i$ =

$\Pi \bar{x} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_k \end{bmatrix}$

- $\boxed{Z_i \sim \mathcal{N}(0, 1)}$
- Let $Y = \sum_{i=1}^{k} Z_i^2$. Y's distribution is χ^2 since $Z_1, \ldots, Z_k$ are iid.
- Hence $\Pr[(1-\epsilon)^2 k \leq Y \leq (1+\epsilon)^2 k] \geq 1 - 2e^{-c\epsilon^2 k}$

$1 - 2e^{-c \cancel{\epsilon^2} \frac{1}{\cancel{\epsilon^2}} \ln \frac{1}{\delta}}$

$\geq 1 - \delta$

Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$Z_i = \sum_{j=1}^{n} \Pi_{ij} x_i$

- $Z_i \sim \mathcal{N}(0, 1)$
- Let $Y = \sum_{i=1}^{k} Z_i^2$. Y's distribution is χ^2 since $Z_1, \ldots, Z_k$ are iid.
- Hence $\Pr[(1-\epsilon)^2 k \leq Y \leq (1+\epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k}$
- Since $k = \Omega(\frac{1}{\epsilon^2} \log(1/\delta))$ we have $\Pr[(1-\epsilon)^2 k \leq Y \leq (1+\epsilon)^2 k] \geq 1 - \delta$

Proof of DJL Lemma

Without loss of generality assume $\|x\|_2 = 1$ (unit vector)

$Z_i = \sum_{j=1}^{n} \Pi_{ij} x_i$

- $Z_i \sim \mathcal{N}(0, 1)$
- Let $Y = \sum_{i=1}^{k} Z_i^2$. Y's distribution is χ^2 since $Z_1, \ldots, Z_k$ are iid.
- Hence $\Pr[(1-\epsilon)^2 k \leq Y \leq (1+\epsilon)^2 k] \geq 1 - 2e^{c\epsilon^2 k}$
- Since $k = \Omega(\frac{1}{\epsilon^2}\log(1/\delta))$ we have $\Pr[(1-\epsilon)^2 k \leq Y \leq (1+\epsilon)^2 k] \geq 1 - \delta$
- Therefore $\|z\|_2 = \sqrt{Y/k}$ has the property that with probability $(1-\delta)$, $\|z\|_2 = (1 \pm \epsilon)\|x\|_2$.

JL lower bounds

Question: Are the bounds achieved by the lemmas tight or can we do better? How about non-linear maps?

Essentially optimal modulo constant factors for worst-case point sets.

n vectors in $\mathbb{R}^d$ $\rightarrow$ n vectors in $\mathbb{R}^k$ $k = \frac{4}{\varepsilon^2} \ln n$ s.t distances are preserved in ℓ_2 norm

Fast JL and Sparse JL

Projection matrix $\mathbf{\Pi}$ is dense and hence $\mathbf{\Pi}x$ takes $\Theta(kd)$ time.

Question: Can we find $\mathbf{\Pi}$ to improve time bound?

Two scenarios: x is dense and x is sparse

$\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n \in \mathbb{R}^d \longrightarrow v_1', v_2', \ldots, v_n' \in \mathbb{R}^k$ k.

Π

$\underline{\Pi}\, \bar{v}_i = v_i'$

$\Theta(kd)$ $\boxed{\frac{1}{\varepsilon^2} \ln d}$

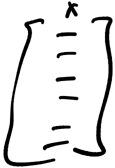
$\begin{bmatrix} - & - & - & - & - \\ - & - & - & - \\ - & - & - \end{bmatrix} \begin{bmatrix} v_i \\ \vdots \\ \vdots \end{bmatrix} = \underline{v_i'}$

Fast JL and Sparse JL

Projection matrix Π is dense and hence Πx takes $\Theta(kd)$ time.

Question: Can we find Π to improve time bound?

$\underline{kd}$



Two scenarios: x is dense and x is sparse

Known results:

- Choose Π_{ij} to be $\{-1, 0, 1\}$ with probability $1/6, 1/3, 1/6$. Also works. Roughly $1/3$ entries are 0
- Fast JL: Choose Π in a dependent way to ensure Πx can be computed in $O(\underline{d \log d + k}^2)$ time. For dense x.
- Sparse JL: Choose Π such that each column is s-sparse. The best known is $s = O(\frac{1}{\epsilon} \log(1/\delta))$. Helps in sparse x.

Part I

(Oblivious) Subspace Embeddings

Subspace Embedding

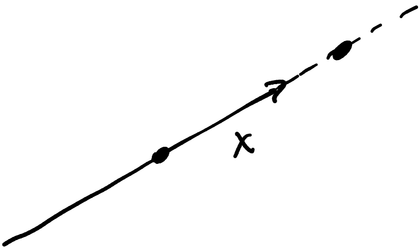
Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

$\bar{x}$ is a vector in R^n

$\Pi x \in R^k$

Then $\underline{\|\Pi \bar{x}\|_2} \simeq \|x\|_2$

$k << n.$

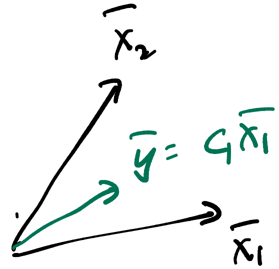


$y = c\bar{x} \quad \forall c$

$\|\Pi \bar{y}\|_2 = (1 \pm \epsilon)\|y\|_2$

Two vectors.

Fix $\bar{x}_1, \bar{x}_2$



$\bar{y} = c_1 \bar{x}_1 + c_2 \bar{x}_2$

want projection matrix Π

s.t. $\forall\ \bar{y}$ in span$(\bar{x}_1, \bar{x}_2)$

$\|\Pi \bar{y}\|_2 \approx (1 \pm \varepsilon)\|y\|$

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why?

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? Π maps E to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $\mathbf{0}$

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? Π maps E to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $\mathbf{0}$
- Possible if $k = d$. Why?

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? Π maps E to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $\mathbf{0}$
- Possible if $k = d$. Why? Pick Π to be an orthonormal basis for E.

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? Π maps E to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $\mathbf{0}$
- Possible if $k = d$. Why? Pick Π to be an orthonormal basis for E. **Disadvantage:** This requires knowing E and computing orthonormal basis which is slow.

Subspace Embedding

Question: Suppose we have linear subspace E of $\mathbb{R}^n$ of dimension d. Can we find a projection $\Pi : \mathbb{R}^n \to \mathbb{R}^k$ such that for *every* $x \in E$, $\|\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2$?

- Not possible if $k < d$. Why? Π maps E to a lower dimension. Implies some non-zero vector $x \in E$ mapped to $\mathbf{0}$
- Possible if $k = d$. Why? Pick Π to be an orthonormal basis for E. **Disadvantage:** This requires knowing E and computing orthonormal basis which is slow.

What we really want: *Oblivious* subspace embedding ala JL based on random projections

Oblivious Supspace Embedding

Theorem

Suppose E is a linear subspace of $\mathbb{R}^n$ of dimension d. Let Π be a DJL matrix $\Pi \in \mathbb{R}^{k\times n}$ with $k = O(\frac{d}{\epsilon^2}\log(1/\delta))$ rows. Then with probability $(1-\delta)$ for every $x \in E$,

$$\|\frac{1}{\sqrt{k}}\Pi x\|_2 = (1 \pm \epsilon)\|x\|_2.$$

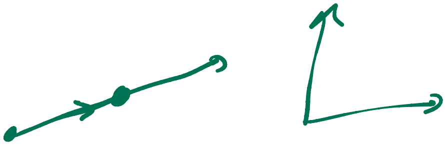
In other words JL Lemma extends from one dimension to arbitrary number of dimensions in a graceful way.

Proof Idea

How do we prove that Π works for *all* $x \in E$ which is an infinite set?

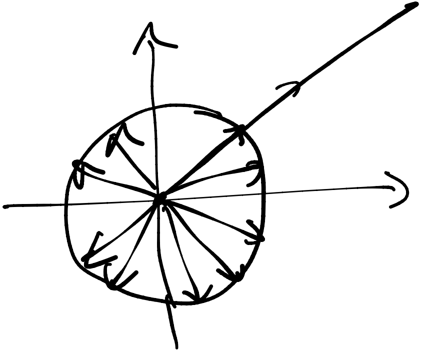
Several proofs but one useful argument that is often a starting hammer is the "net argument"

- Choose a large but finite set of vectors T carefully (the net)
- Prove that Π preserves lengths of vectors in T (via naive union bound)
- Argue that *any* vector $x \in E$ is sufficiently close to a vector in T and hence Π also preserves length of x



Net argument

Sufficient to focus on unit vectors in E. Why?



Net argument

Sufficient to focus on unit vectors in E. Why?

Also assume wlog and ease of notation that E is the subspace formed by the first d coordinates in standard basis.

E is linear subspace of d-dim in $\mathbb{R}^n$.

Net argument

Sufficient to focus on unit vectors in E. Why?

Also assume wlog and ease of notation that E is the subspace formed by the first d coordinates in standard basis.

Claim: There is a net T of size $e^{O(d)}$ such that preserving lengths of vectors in T suffices.

d^d

Net argument

Sufficient to focus on unit vectors in E. Why?

Also assume wlog and ease of notation that E is the subspace formed by the first d coordinates in standard basis.

Claim: There is a net T of size $e^{O(d)}$ such that preserving lengths of vectors in T suffices.

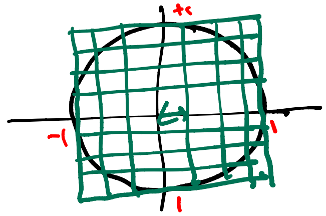
d^d

$d \log d$.

Assuming claim: use DJL with $k = O(\frac{d}{\epsilon^2}\log(1/\delta))$ and union bound to show that all vectors in T are preserved in length up to $(1 \pm \epsilon)$ factor.

any fixed vector is pres

w.p. $1 - \exp(-\frac{d}{\delta})$



$T =$ all "grid points" of the box $[-1,1]^d$

what is the spacing?

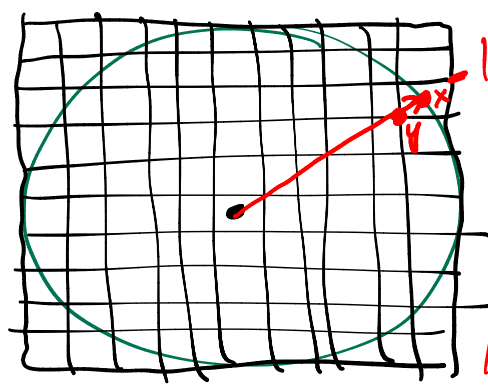
$\frac{\varepsilon}{d}$

$|T| = ?$ $\boxed{\left(\frac{2d}{\varepsilon}\right)^d}$

Claim: If all vectors in T are preserved to $(1\pm\varepsilon)$ then all unit vectors preserved to $(1\pm 2\varepsilon)$.

$$k = \frac{1}{\varepsilon^2} \ln\left(\left(\frac{2d}{\varepsilon}\right)^d \cdot \frac{1}{\delta}\right).$$

$\underline{\underline{d \ln d}}$



$\|\frac{1}{\sqrt{k}}\Pi\, \bar{x}\|_2 \approx (1\pm\varepsilon)\|x\|_2$

If x on the circle. We know

$\|\frac{1}{\sqrt{k}}\Pi\, \bar{y}\|_2 \approx (1\pm\varepsilon)\|y\|_2$

If $x = y + z$

$\underline{\underline{z}} = x - y.$

Net argument

Sufficient to focus on unit vectors in E.

Also assume wlog and ease of notation that E is the subspace formed by the first d coordinates in standard basis.

A weaker net:

- Consider the box $[-1,1]^d$ and make a grid with side length ϵ/d
- Number of grid vertices is $(2d/\epsilon)^d$
- Sufficient to take T to be the grid vertices
- Gives a weaker bound of $O(\frac{1}{\epsilon^2} d \log(d/\epsilon))$ dimensions
- A more careful net argument gives tight bound

Net argument: analysis

Fix any $x \in E$ such that $\|x\|_2 = 1$ (unit vector)
There is grid point y such that $\|y\|_2 \leq 1$ and x is close to y
Let $z = x - y$. We have $|z_i| \leq \epsilon/d$ for $1 \leq i \leq i \leq d$ and $z_i = 0$ for $i > d$

Net argument: analysis

Fix any $x \in E$ such that $\|x\|_2 = 1$ (unit vector)
There is grid point y such that $\|y\|_2 \leq 1$ and x is close to y
Let $z = x - y$. We have $|z_i| \leq \epsilon/d$ for $1 \leq i \leq i \leq d$ and $z_i = 0$ for $i > d$

$\|\Pi z\| = z_1 \bar{e}_1 + z_2 \bar{e}_2 + \cdots + z_d \cdot \bar{e}_d$

$$\begin{aligned}
\|\Pi x\| = \|\Pi y + \Pi z\| &\leq \|\Pi y\| + \|\Pi z\| \\
&\leq (1+\epsilon) + (1+\epsilon)\sum_{i=1}^{d} |z_i| \\
&\leq (1+\epsilon) + \epsilon(1+\epsilon) \leq 1 + 3\epsilon
\end{aligned}$$

$\|\Pi x\| \geq 1 - 3\epsilon$

$\|\Pi y\| \geq (1-\epsilon)\|y\| \geq (1-\epsilon)(1-\epsilon) \geq 1-2\epsilon$

Net argument: analysis

Fix any $x \in E$ such that $\|x\|_2 = 1$ (unit vector)
There is grid point y such that $\|y\|_2 \leq 1$ and x is close to y
Let $z = x - y$. We have $|z_i| \leq \epsilon/d$ for $1 \leq i \leq i \leq d$ and $z_i = 0$ for $i > d$

$$\begin{aligned}\|\Pi x\| = \|\Pi y + \Pi z\| &\leq \|\Pi y\| + \|\Pi z\| \\ &\leq (1+\epsilon) + (1+\epsilon)\sum_{i=1}^{d} |z_i| \\ &\leq (1+\epsilon) + \epsilon(1+\epsilon) \leq 1 + 3\epsilon\end{aligned}$$

Similarly $\|\Pi x\| \geq 1 - O(\epsilon)$.

Application of Subspace Embeddings

Faster algorithms for approximate

- matrix multiplication
- regression
- SVD

Basic idea: Want to perform operations on matrix $\boldsymbol{A}$ with $\boldsymbol{n}$ data columns (say in large dimension $\mathbb{R}^{\boldsymbol{h}}$) with small effective rank $\boldsymbol{d}$. Want to reduce to a matrix of size roughly $\mathbb{R}^{\boldsymbol{d}\times\boldsymbol{d}}$ by spending time proportional to nnz$(\boldsymbol{A})$.

Π where Π is a DJL

$\frac{1}{\epsilon^2}\ln\frac{1}{\delta}$

$\Pi\bar{x}$

Later in course.

A

$\Pi \in R^{k\times n}$

$A \in R^{n\times d}$

Π

$\underline{\underline{\Pi A}}$ very $\underline{\underline{\text{fast}}}$

in time $\underline{\underline{nnz(A)}}$