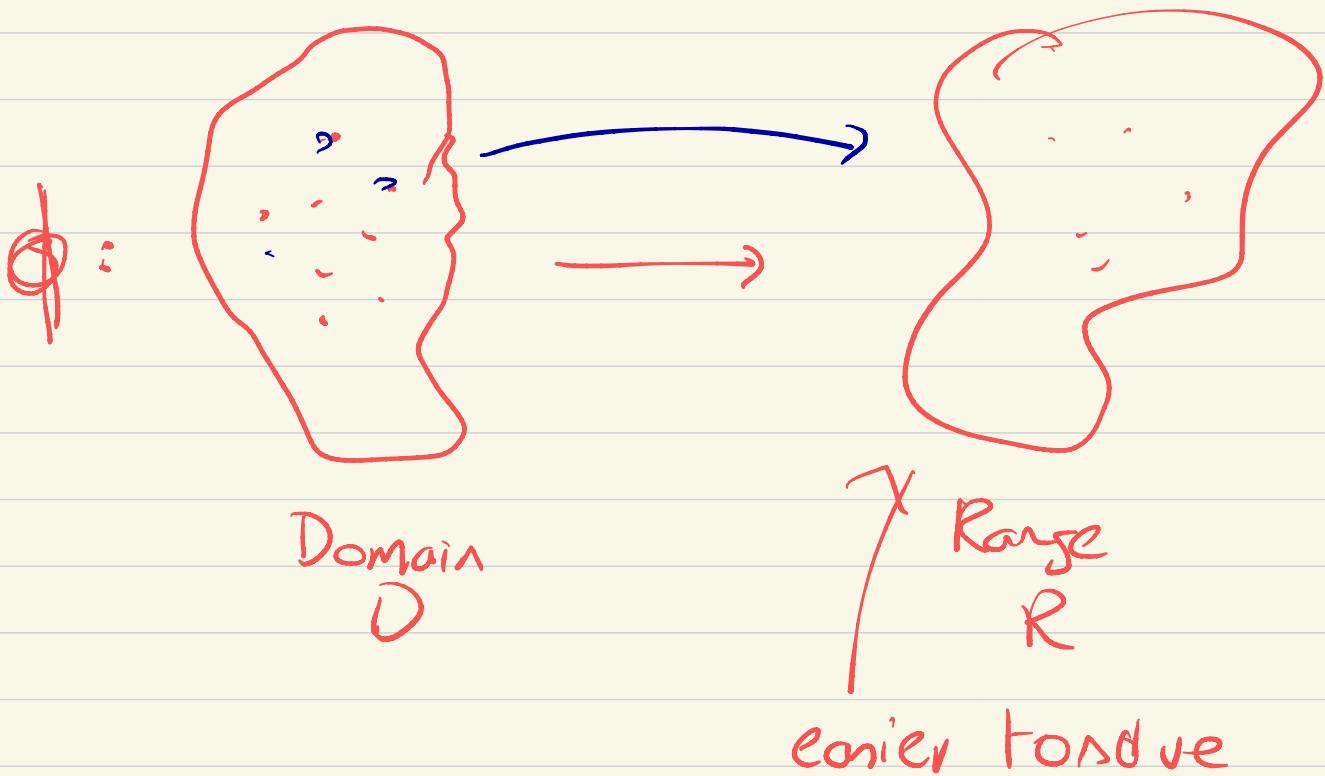


LECTURE 9

"Embedding"



ϕ embedding : distances / norms

Dimension Reduction

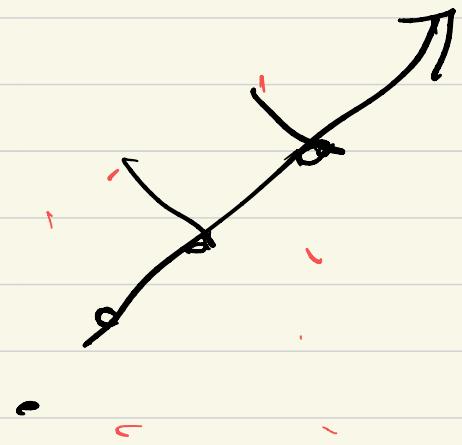
$$\phi: X_1 \dots X_n \in \mathbb{R}^d \rightarrow \mathbb{R}^k$$

for small k .

1-dimensional embedding

$$\phi: \mathbb{R}^d \rightarrow \mathbb{R}$$

T
randomized



Random Projection

- Pick a random direction $g \in \mathbb{R}^d$
 $g \sim \mathcal{N}(0, 1)^d$
- $g: \mathbb{R}^d \rightarrow \mathbb{R}$

$$g(x) = \langle x, g \rangle = \sum x_i g_i$$

Claim: Random 'g' gives an unbiased estimator for squared lengths.

$$\forall x \in \mathbb{R}^d \quad \underset{g}{\mathbb{E}}[(g(x))^2] = \underset{g}{\mathbb{E}}[\langle g, x \rangle^2] = \|x\|^2$$

$$\mathbb{E}_{\mathbf{g}}[\langle \mathbf{g}, \mathbf{x} \rangle^2] = \mathbb{E}_{\mathbf{g}} \left[\left(\sum_{i=1}^d g_i x_i \right)^2 \right]$$

$$= \sum_{i,j} x_i x_j \underbrace{\mathbb{E}_{\mathbf{g}}[g_i g_j]}_{\| \mathbf{x} \|^2}$$

$$(g_1 - g_n) \sim (\mathcal{N}(0, 1))^d \quad \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$$

$$= \sum x_i^2 = \underline{\underline{\|\mathbf{x}\|^2}}$$

$$\boxed{\underbrace{g(\mathbf{u}) = \langle \mathbf{g}, \mathbf{u} \rangle}_{\text{linear}} \rightarrow \text{linear}}$$

$$\mathbb{E}_{\mathbf{g}} \|g(\mathbf{x}) - g(\mathbf{y})\|^2 = \mathbb{E}_{\mathbf{g}} \langle \mathbf{g}, \mathbf{x} - \mathbf{y} \rangle^2 = \underline{\underline{\|\mathbf{x} - \mathbf{y}\|^2}}$$

$$g(\mathbf{x}) = \sum_{i=1}^d g_i x_i$$

$\mathbf{x} \in \mathbb{R}^d$
 $\underline{g \sim N(0, 1)^d}$

Properties:

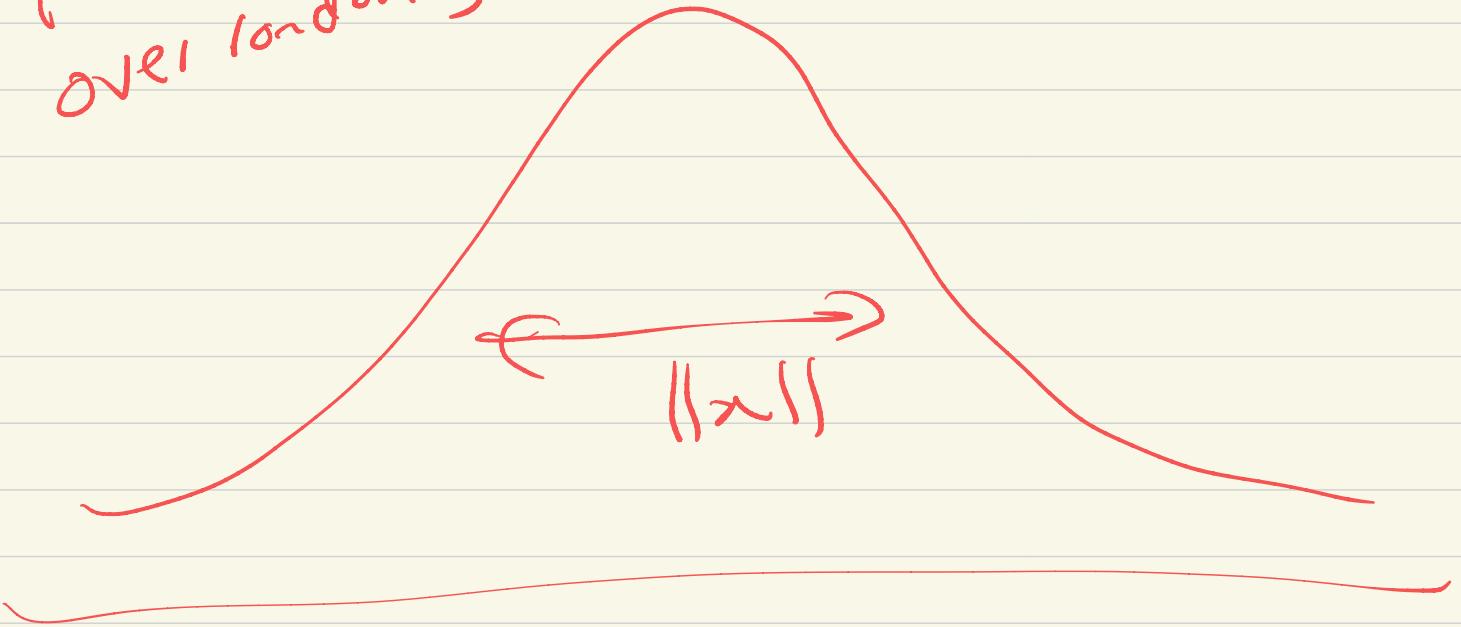
- linear combinations of Gaussians are Gaussians
- $X \sim N(\mu_1, \sigma_1^2)$
- $Y \sim N(\mu_2, \sigma_2^2)$

If $x, y: x+y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$$E[g(\mathbf{x})] = E[\sum g_i x_i] = 0$$

$$\begin{aligned} \text{Var}[g(\mathbf{x})] &= \text{Var}\left[\sum g_i x_i\right] = \sum x_i^2 \\ &= \|\mathbf{x}\|^2 \end{aligned}$$

Fix α
over random S



Streaming Alg:

- Streaming Alg to estimate l_2 moment -

Estimator \hat{X} for some X

Define

$$h: \mathbb{R}^d \rightarrow \mathbb{R}^k$$

- Pick $g_1, \dots, g_k \in N(0, 1)^d$

$$- h(x) = \frac{1}{\sqrt{k}} (\langle g_1, x \rangle, \langle g_2, x \rangle, \dots, \langle g_k, x \rangle)$$

$$\|h(x)\|^2 = \frac{1}{k} \sum_{i=1}^k \langle g_i, x \rangle^2$$

average of k estimates

$$- E[\|h(x)\|^2] = \|x\|^2$$

$$- \boxed{\text{Var}[\|h(x)\|^2] = \frac{1}{k} \text{Var}[\langle g_i, x \rangle^2]}$$

DJL

[Distributional Johnson-Lindenstrauss]

Fix $\varepsilon \in (0, \frac{1}{2})$ and $x \in \mathbb{R}^d$

$$\Pr \left[\left\| G(x) \right\|^2 = \underbrace{(1 \pm \varepsilon) \|x\|^2}_{\in (1-\varepsilon)\|x\|^2, (1+\varepsilon)\|x\|^2} \right] \geq 1 - \delta$$

if
$$K \geq \frac{10 \log(\gamma_f)}{\varepsilon^2}$$

Proof: Fix x . $\|x\|=1$

$$\|G(x)\|^2 = \sum_{i=1}^K \langle g_i, x \rangle^2$$

↑ standard square of a Gaussian

Lemma:

$$Y_1, \dots, Y_K \sim \mathcal{N}(0, 1)$$

$$Z = \sum_{i=1}^K Y_i^2 \quad E[Z] = K$$

$$\Rightarrow \Pr [Z \in [(-\varepsilon, +\varepsilon)]] \geq 1 - \delta$$

Proof:

Pick some $t \in \mathbb{R}$

$$\Pr[\underline{\text{Z}} > ((t\epsilon)K)] \leq \frac{E[e^{tZ}]}{e^{t(t-\epsilon)K}},$$

Z^2

Moment inequality $\theta > 0$

$$\Pr[A > \theta] \leq \frac{E[A]}{\theta}$$

$$E(A) = \sum_i \Pr(A=i) i > \Pr(A > \theta) \theta$$

$$E[e^{tZ}] = E[e^{t(\sum Y_i^2)}]$$

$$= E\left[\prod_{i=1}^K e^{tY_i^2}\right]$$

$$= \prod_{i=1}^K E[e^{tY_i^2}] = (E[e^{tY^2}])^K$$

$$E\left[e^{tY^2}\right] = \frac{1}{\sqrt{2\pi}} \int e^{-y^2/2} \cdot e^{ty^2} dy$$

$Y \sim N(0, 1)$

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

$$E\left[e^{tz}\right] = \left(\frac{1}{\sqrt{1-2t}}\right)^K \quad \longrightarrow (1)$$

$$P_1\left[Z \geq (1+\varepsilon)K \right] \leq \left(\frac{1}{\sqrt{1-2t}}\right)^{(1+\varepsilon)t} K$$

Optimal value

$$t = \frac{\varepsilon}{2}$$

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

$$\Pr[Z \in K(1-\varepsilon, 1+\varepsilon)] \geq 1 - 2e^{-\varepsilon^2 K/8}$$

$$\Rightarrow K \approx \frac{10 \log(1/\delta)}{\varepsilon^2}$$

$$\geq 1 - \delta$$

JL dimension reduction
 $\bar{X} = X_1, \dots, X_n \in \mathbb{R}^d$ data points



$g(x_1) \dots g(x_n)$ all pairwise distances
 $\underbrace{\quad\quad\quad}_{\text{all pairwise distances}} =$

$$n^2 \text{ distances. } \|g(x_i) - g(x_j)\|$$

$$= \|g - (x_i - x_j)\|$$

each distance $1 - \frac{\delta}{n^2}$

\Rightarrow all distances are preserved up to $1 - \delta$.

Fix

$$\Rightarrow K = O\left(\log n + \log \frac{1}{\delta}\right)$$

ε^2

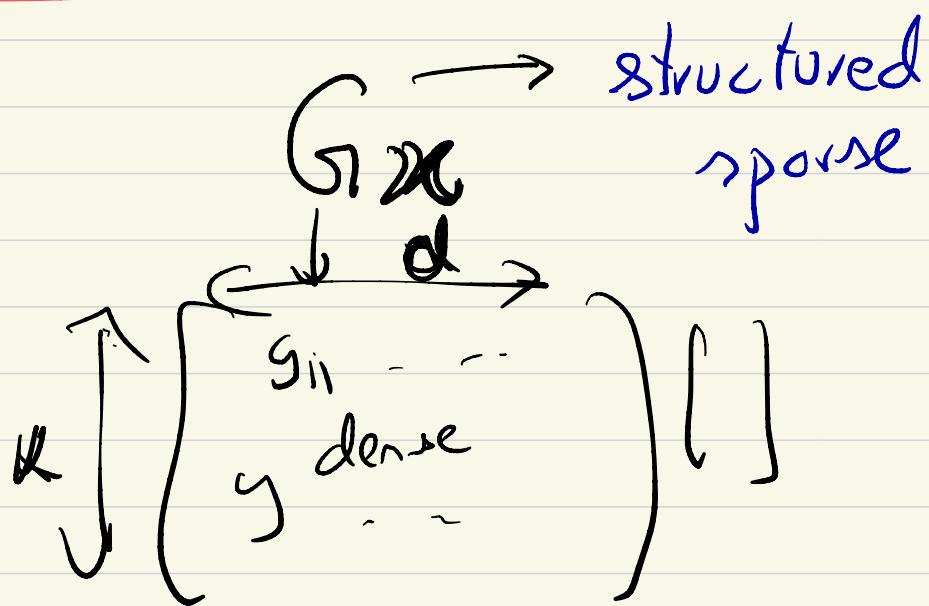
OPTIMAC

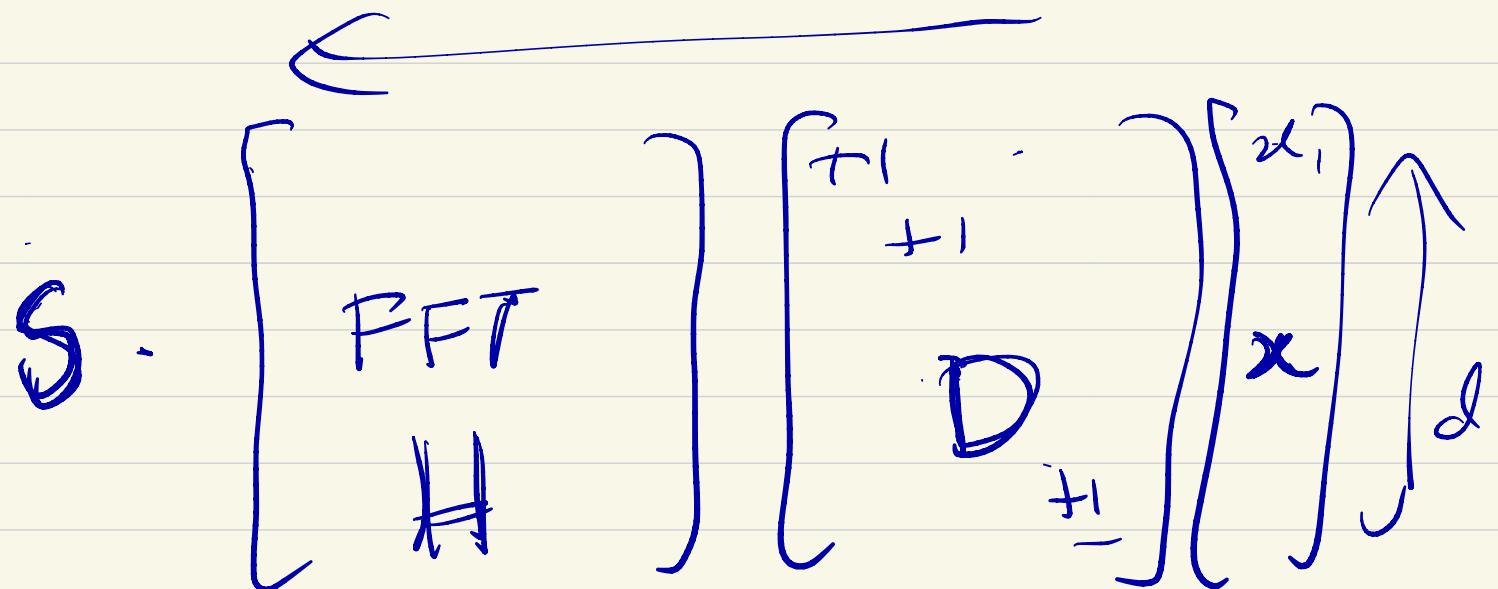
w.p 1- δ all pairs x_i, x_j

$$\|g(x_i) - g(x_j)\|_2 \approx (1 + \varepsilon) \|x_i - x_j\|_2$$

- Data independent

Fast JL

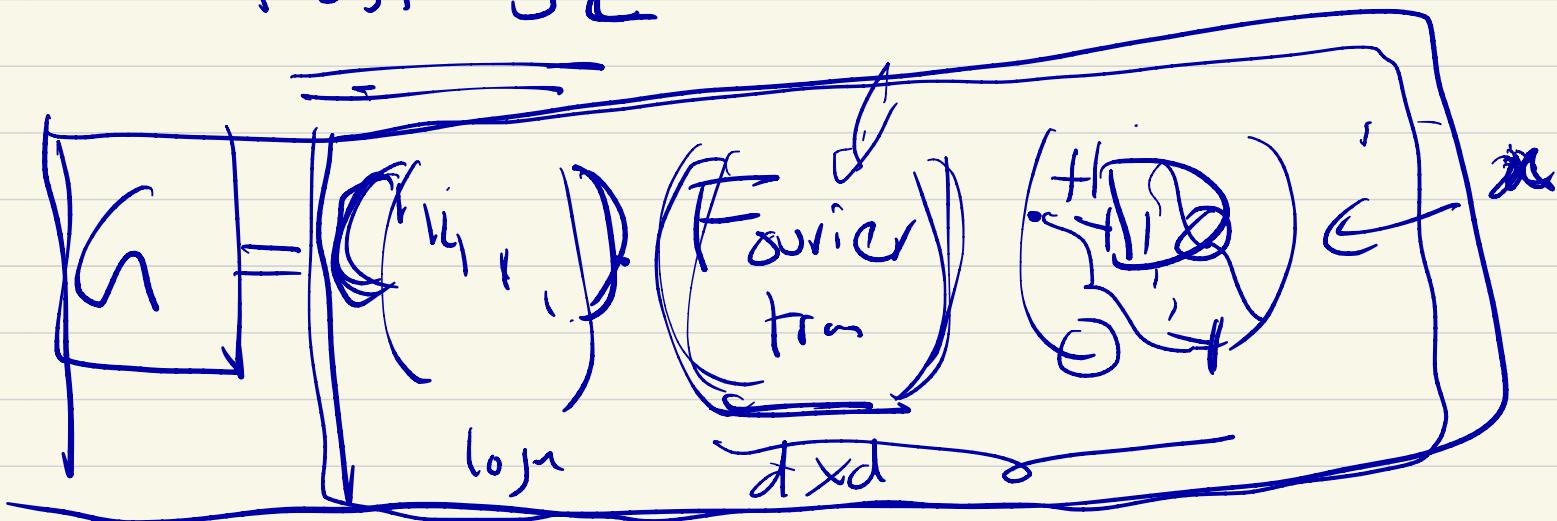




flip
 signs of
 a random's

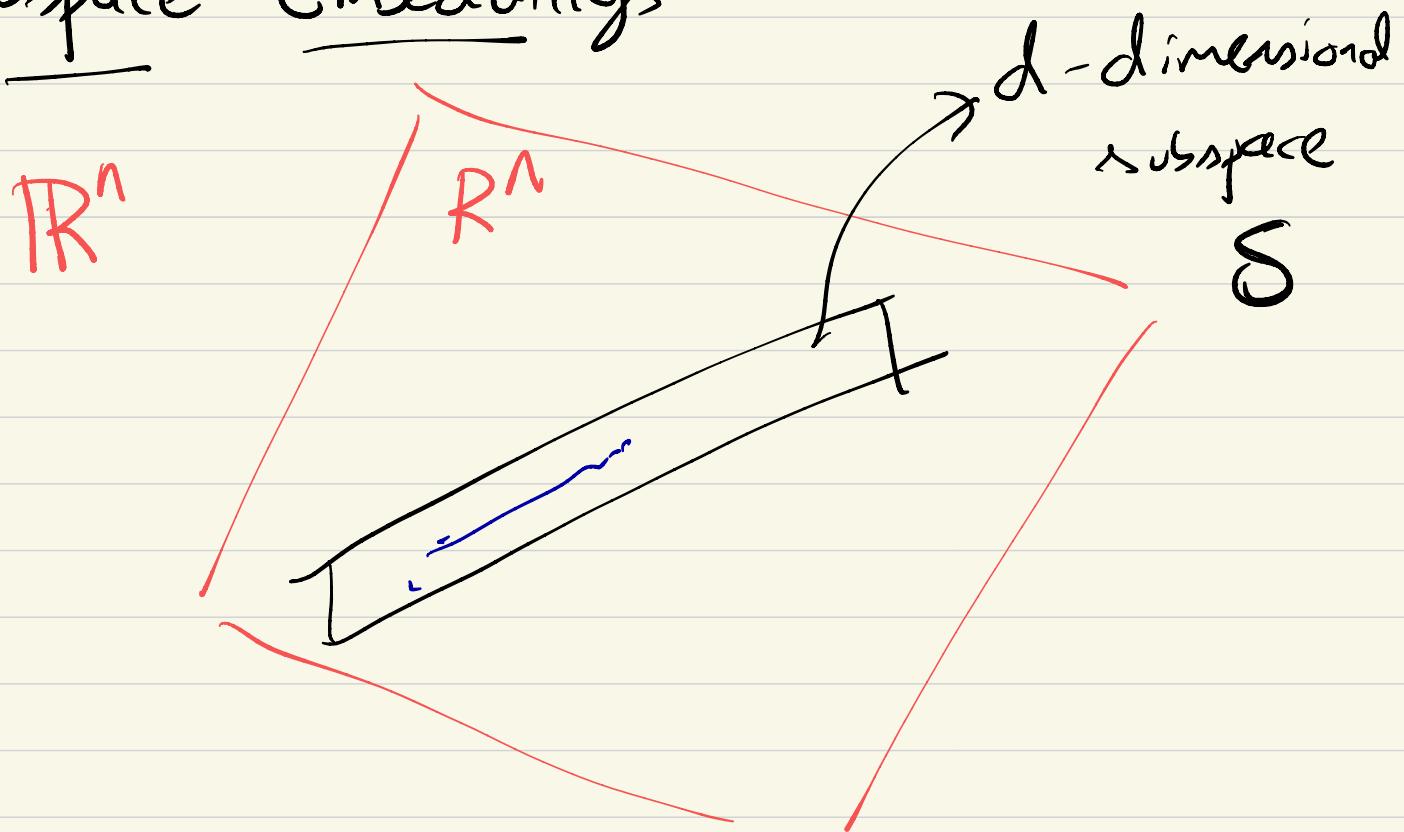
Fast JL

(Gx)



$$\begin{pmatrix} 1 & & & & & \\ & 1 & 0 & 0 & 0 & 0 \\ 0 & & 1 & & & \end{pmatrix}$$

Oblivious Subspace Embeddings



preserve all lengths within S

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

$$\begin{bmatrix} & \\ A & \\ & \end{bmatrix}$$

↑
n
↓
100

$$S = \text{colspan}(A)$$

Construct a map

$$g: \mathbb{R}^n \rightarrow \mathbb{R}^K$$

st $x \in S$

$$\|gx\| = (1+\varepsilon)\|x\|$$

Linear Regression (ℓ^2)

$$\min_x \|Ax - b\|_2^2$$

Diagram illustrating the linear regression problem:

The left side shows the expression $\|Ax - b\|_2^2$. A red bracket labeled d encloses the vector b . A red grid labeled A encloses the matrix A . A red bracket labeled n encloses the vector x .

The right side shows the expression $\|Ax - b\|_2^2$. A red bracket labeled d encloses the vector b . A red bracket labeled n encloses the vector x .

Apply G : and solve

$$\min_x \|Gx - b\| \text{ to get a}$$

$$\hat{x} = \arg \min_x \|Gx - b\|$$

$$\boxed{x^* = \arg \min_x \|Ax - b\|}$$

$$\|A\hat{x} - b\| \leq (1+\varepsilon) \|Ax^* - b\| \quad -(1)$$

$$\underbrace{\varepsilon}_{\text{S} = \text{Span of columns of } A, b} \leq (1+\varepsilon) \|Ax^* - b\| \quad -(2)$$

$$\leq \frac{(1+\varepsilon)}{(1-\varepsilon)} \|Ax^* - b\|$$

↓

true minimum

Thm: Pick $G = (g_1, \dots, g_K)$

$$\uparrow \\ N(0, 1)^n$$

and $\left\{ K \geq 0 \left(\frac{d + \log(1/\delta)}{\varepsilon^2} \right) \right\}$

then H subspace S

w.p. $1-\delta$, $\|Gx\| \approx (1 \pm \varepsilon) \|x\|$

for $x \in \underline{S}$.

Proof: $K \approx O\left(\frac{d + \log \frac{1}{\delta}}{\epsilon^2}\right)$

$\forall x \in S$

$$B_x = \{y \mid \|y\| \notin (1 \pm \epsilon) \|x\|\}$$

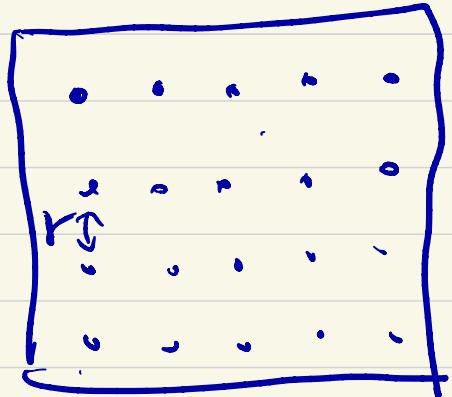
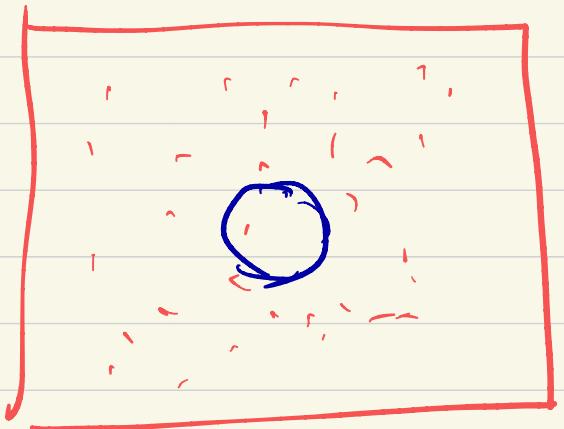
ϵ -Net:

Def: An ϵ -net for
a set $B \subseteq \mathbb{R}^d$ is

a set of points N s.t

$\forall x \in B \quad \exists y \in N$

$$\|y - x\| \leq \epsilon$$



Unit Ball in subspace $S \rightarrow$ ϵ -net for Ball

Strategy:

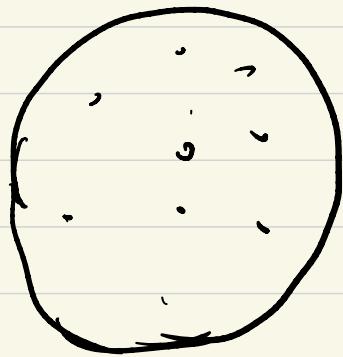
a) With probability $1-\delta$,

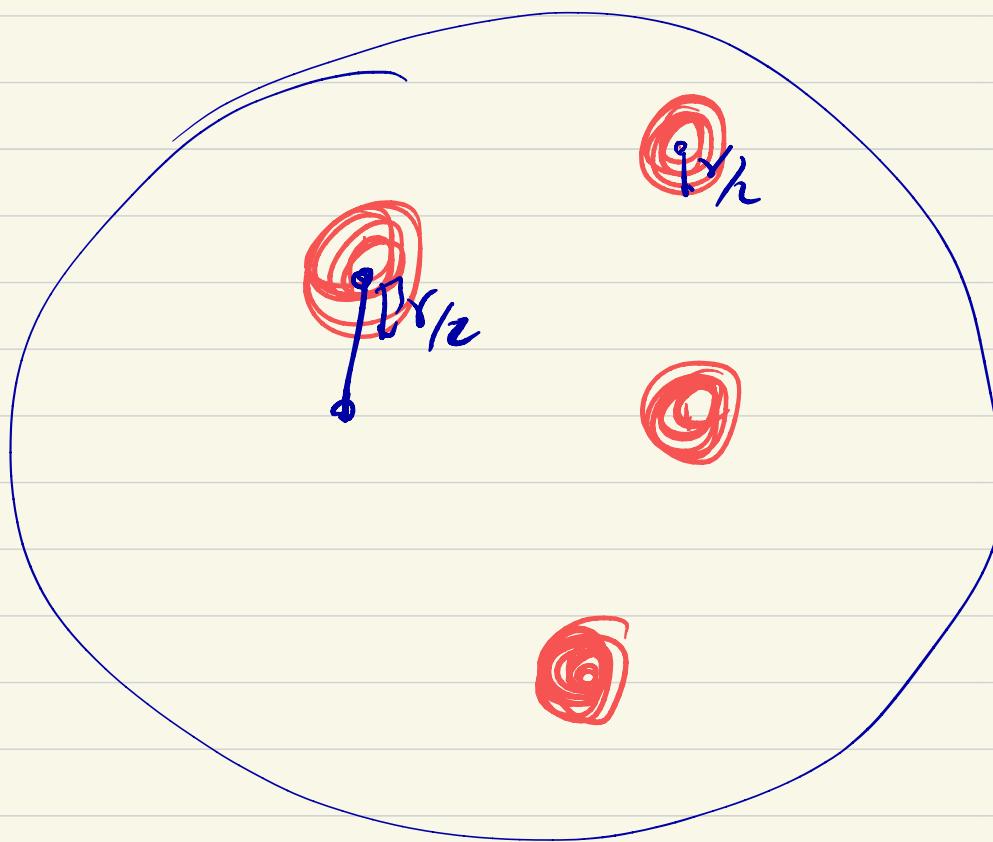
$$\forall y \in N, \|Gy\| \approx \|y\| (1 \pm \varepsilon)$$

$$|\mathcal{N}| \leq \left(\frac{1}{\delta}\right)^d \quad \leftarrow K > \frac{\log |\mathcal{N}| + \log \frac{1}{\delta}}{\varepsilon^2} = \frac{(d + \log \frac{1}{\delta})}{\varepsilon^2}$$

b) If $\forall y \in N \|G(y)\| \approx \|y\|$

then $\forall x \in \text{Ball}_{\text{unit}}$ $\|Gx\| \approx \|x\|$





$$\frac{\text{Vol}(\text{Ball}(0,1))}{\text{Vol}(\text{Ball}(0,r_s/2))}$$