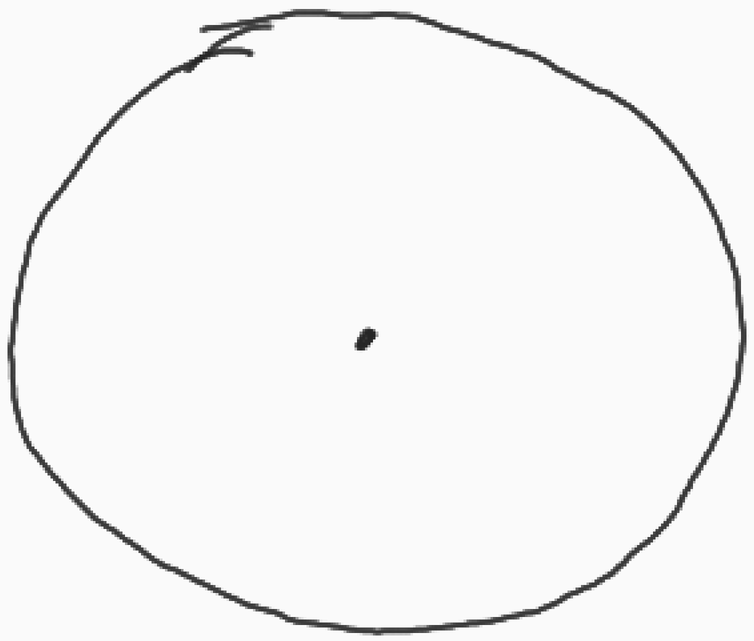
Lecture 12 : Dimension Reduction

- JL theorem
  - proof
- Subgaussian RVs
- geometry of high dimensions



Thm: Given $\underbrace{x_1, x_2 \ldots x_n}_{n \text{ pts}} \in \mathbb{R}^D$, there exists linear map $A : \mathbb{R}^D \to \mathbb{R}^k$

$$\boxed{k = \Theta\left(\frac{\lg n}{\varepsilon^2}\right)}$$

$(\varepsilon < 1/2 \text{ say})$

st $$1-\varepsilon \le \frac{\|A(x_i) - A(x_j)\|_2^2}{\|x_i - x_j\|_2^2} \le 1+\varepsilon \qquad \forall i,j$$

Thm (one vector) $\exists$ random choice of linear A st for any $x \in \mathbb{R}^D$, x unit vector

$$\|A(x)\|^2 \in [1 \pm \varepsilon] \qquad \text{wp } 1 - \frac{1}{n^2}$$

$$\frac{x_i - x_j}{\|x_i - x_j\|}$$

$$\|x\|^2 = \sum_i x_i^2$$

$$A(x) = \frac{1}{\sqrt{k}} \; k \overset{D}{\boxed{\quad \square \quad M \quad}} \begin{pmatrix} x \in \mathbb{R}^D \end{pmatrix} = \frac{1}{\sqrt{k}} M x$$

$\square \sim N(0,1)$

Pf. ① $E[\|Ax\|^2] = 1 = \|x\|^2$

② Tail Bd.

$$\frac{1}{\sqrt{k}}\begin{pmatrix} G \quad G \quad G \quad G \\ M \\ \end{pmatrix}\begin{pmatrix} x_1 \\ x_1 \\ X \\ x_D \end{pmatrix} \qquad \frac{1}{\sqrt{k}}\begin{pmatrix} \circ \\ \circ \\ \circ \\ \circ \end{pmatrix}$$

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

$X \sim N(\mu, \sigma^2)$

① $E[\|Ax\|^2] = E \sum_{i=1}^{k} (Ax)_i^2$

$= E\left[\frac{1}{k}\sum_{i=1}^{k} G^2\right] = 1.$

$\sum x_i G_i$

$\sim N(0, \sum_i x_i^2)$

$= N(0,1)$

$cX \sim N(c\mu, c^2\sigma^2)$

$\underbrace{X_1 + X_2}_{\text{ind}} \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

② $Ax \sim \frac{\text{sum of } k \text{ indep (Gaussians)}^2}{k} = \sum_{i=1}^{k} G_i^2 / k$

$EX^2 = Var(X) + (EX)^2$

$= \sigma^2 + 0 = \sigma^2$

$Pr\left[\frac{\sum G_i^2}{k} \geq 1+\varepsilon\right] \leq$ ______

$E[e^{t \sum G_i^2}]$ — MGF of G^2, X^2

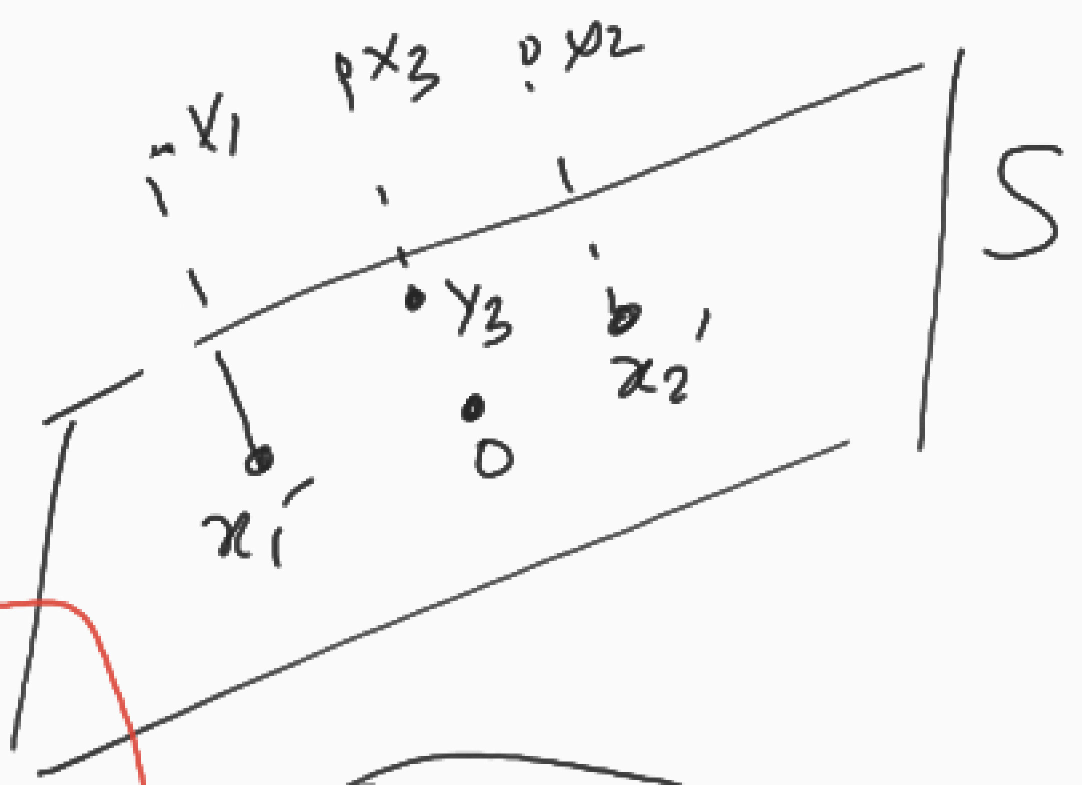
$Pr[\sum G_i^2 \geq k(1+\varepsilon)] = Pr[e^{t\sum G_i^2} \geq e^{tk(1+\varepsilon)}] \leq \frac{\prod E[e^{t G_i^2}]}{e^{tk(1+\varepsilon)}} = \frac{\left(\frac{1}{\sqrt{1-2t}}\right)^k}{e^{tk(1+\varepsilon)}} \leq \boxed{e^{-\varepsilon^2 k/8}}$

$(t > 0)$

Algebra omitted

if $k = \frac{16 \ln n}{\varepsilon^2} \Rightarrow e^{-\varepsilon^2 k/8} = \frac{1}{n^2}$

$\left\{\begin{array}{l} \text{Pick a } k\text{-dim subspace } S \text{ of } \mathbb{R}^D \text{ u.a.r.} \\ \text{Project all pts } x_i \text{ onto } S \\ \text{Scale up by } \sqrt{\frac{D}{k}} \end{array}\right\}$



$\Rightarrow$ Pick v_1 u.a.r. from surface of unit ball

v_2 u.a.r. $\perp v_1$

v_3 . ——— $\perp \{v_1, v_2\}$

v_{D-k} . . . . .

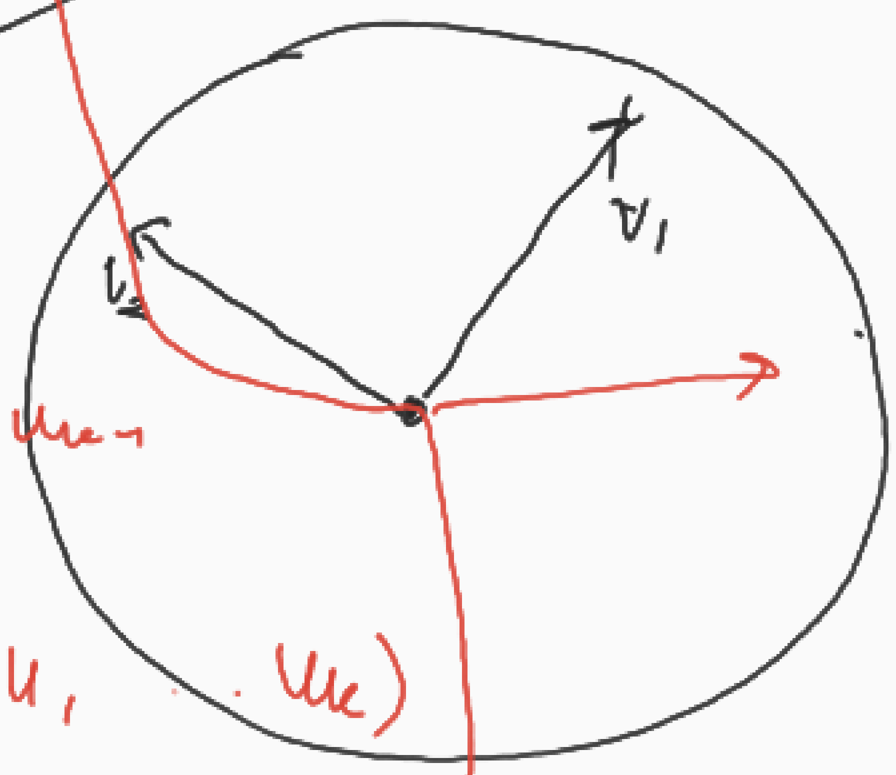
$\{x \perp v_1, v_2 \ldots v_{D-k}\}$ subspace

u_1

$u_2 \perp$

$u_k \perp u_1 \ldots u_{k-1}$

span$(u_1, \ldots u_k)$



Pick v uar from surface of sphere in $\mathbb{R}^D$

$$v' = \frac{(G, G, \dots, G)}{\sqrt{D}} \qquad V = \frac{v'}{\|v'\|}$$

indep $N(0,1)$ r.v.s.

indep from $N(0,1)^D / \sqrt{D}$

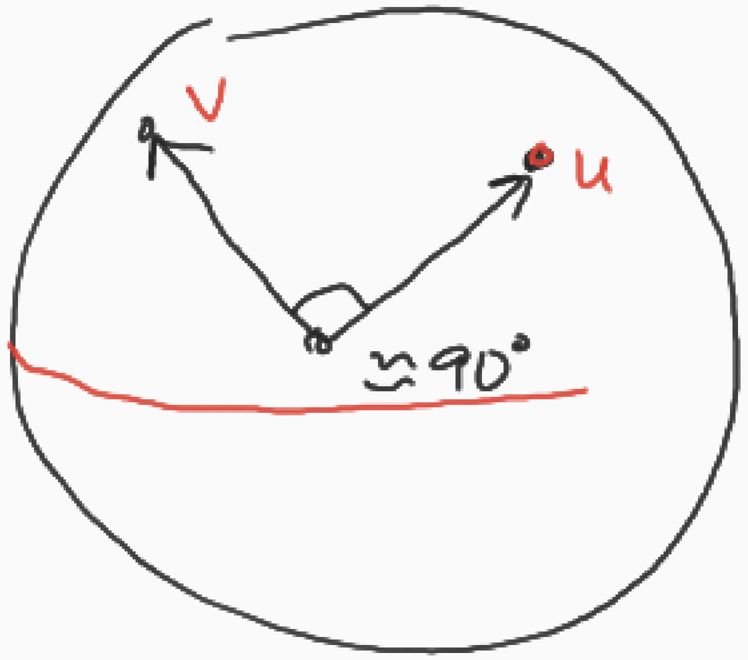
$v_1, v_2, \dots, v_k$ ~~indep from surface of sphere~~

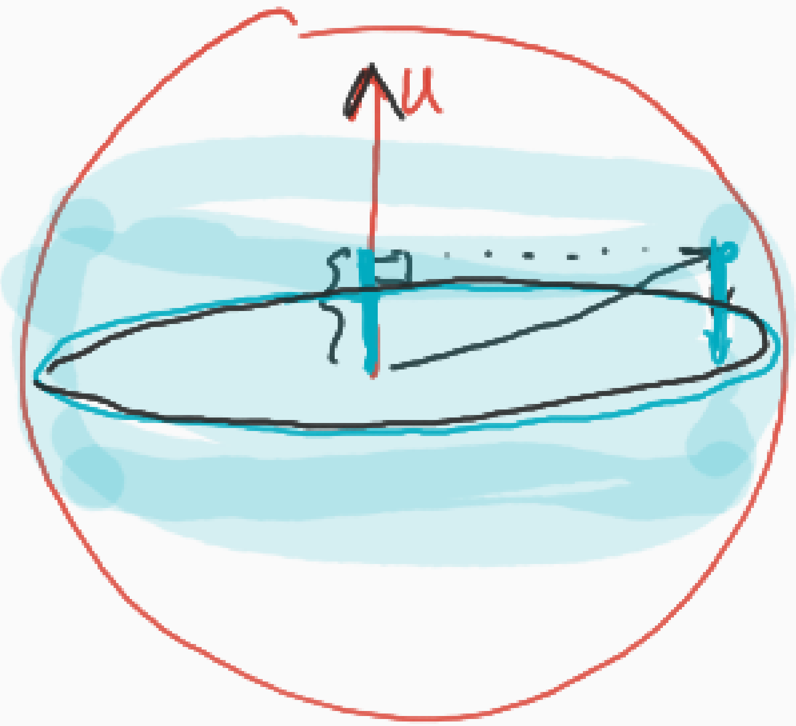
$\simeq$ almost ortho~~gonal~~ to each other.

orthonormal

$$v' \leftarrow \frac{(G, G \dots G)}{\sqrt{D}}$$ has squared length $(1 \pm \varepsilon)$ whp.

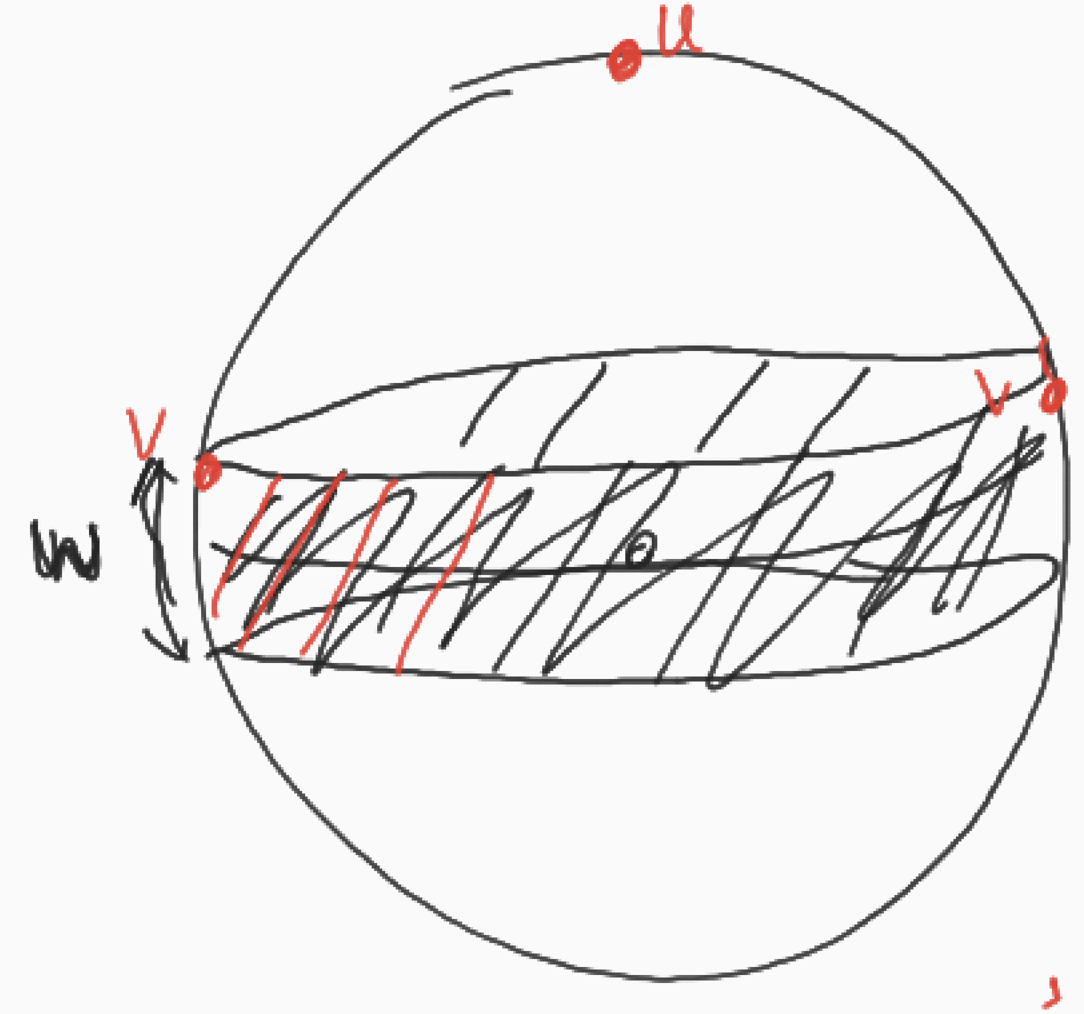
$$\|v'\|^2 = \frac{\sum G^2}{D} \simeq (1 \pm \varepsilon)$$





Fact: unit sphere in $\mathbb{R}^D$

want band to contain 50% of the surface



$\Rightarrow w = 1/2, \quad 1/\log D, \quad \boxed{1/\sqrt{D}} \quad 1/D$

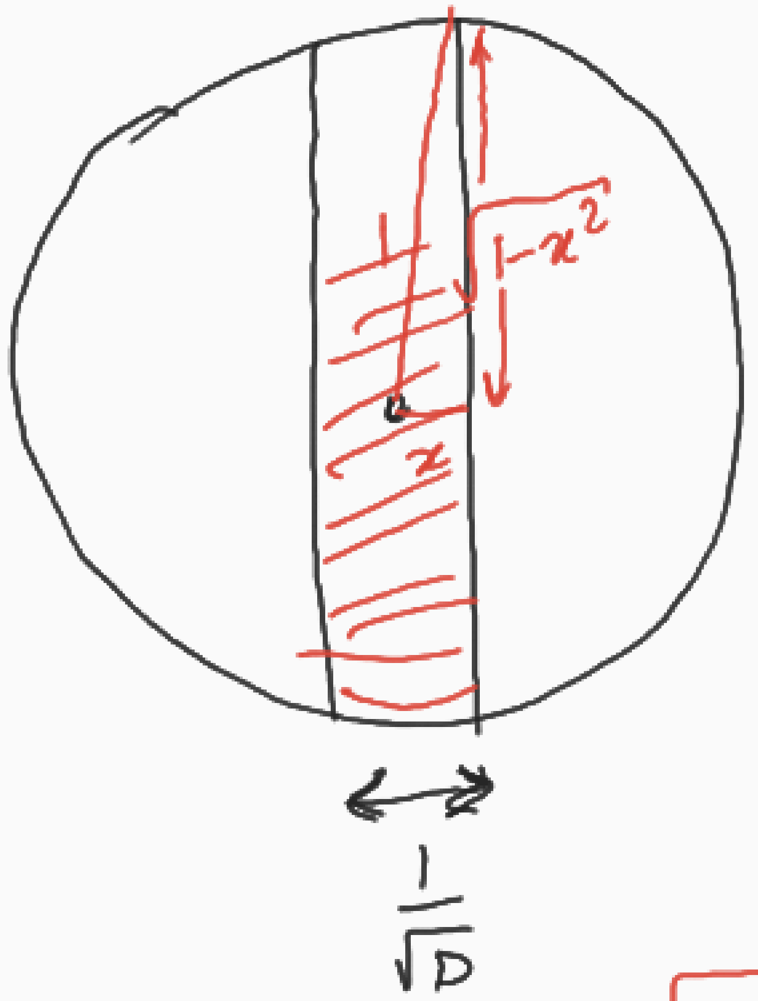
Moral: Pick k random vector from $\dfrac{G, G \ldots G}{\sqrt{D}}$

$$Ax = \frac{1}{\sqrt{K}} Mx \quad \text{vs} \quad A(x) = \left(\text{proj of } x \text{ onto random } k \text{ subspace}\right)\sqrt{\frac{D}{K}}$$

length of circle @ x $= \underline{\left(\sqrt{(1-x^2)}\right)^{d-1} \cdot C_d}$

$$\simeq e^{-x^2/2\,(d-1)}\, C_d$$

$$\underbrace{\int_{x=0}^{W} e^{-x^2/2\,(d-1)}}_{\text{at } W = \sqrt{\frac{\boxed{\text{constant}}}{d-1}}} C_d \quad \text{vs.} \quad 50\%\ C_d$$



$$1+z \simeq e^z$$

$$Ax = \frac{1}{\sqrt{k}} \begin{bmatrix} & & M & \end{bmatrix} \begin{pmatrix} 1 \\ x \\ \\ \end{pmatrix}$$

$k \cdot D$ gaussians

$\geq D$.

$M_{ij} \sim N(0,1)$ ← invariance principle

$\boxed{\sim \{\pm 1\}}$ ← Yes

$\boxed{\sim \text{sparse?} \leftarrow \cancel{\text{No}}\ \text{Yes.}}$

$\{+1, -1\}$ Rademacher

indep.

$\|Ax\|^2 \in [1 \pm \varepsilon]$ whp.

① $E[\|Ax\|^2] = 1$ if x unit vector ✓

② $\Pr[\|Ax\|^2 \geq 1+\varepsilon] \leq e^{-\varepsilon^2 k/8}$

$$\Pr[X \geq \mu + \lambda] = \Pr[e^{t(X-\mu)} \geq e^{t\lambda}] \leq \frac{E[e^{t(X-\mu)}]}{e^{t\lambda}} = e^{-\underbrace{(t\lambda - \psi(t))}}$$

def $\psi(t) := \log E[e^{t(X-\mu)}]$ (log MGF)

$\rightarrow$ $\psi^*(\lambda) = \sup_t (t\lambda - \psi(t))$ Legendre dual of ψ (-Fenchel)

$$\boxed{\Pr[X \geq \mu + \lambda] \leq e^{-\psi^*(\lambda)}}$$ Generic Chernoff

$X \sim N(0, \sigma^2)$ $\quad \psi(t) = \sigma^2 t^2/2 \quad \psi^*(\lambda) = \frac{\lambda^2}{2\sigma^2}$ $\Rightarrow \Pr(G \geq \lambda) \leq e^{-\lambda^2/2\sigma^2}$

$X \sim \{-1, 1\}$ $\quad \psi(t) = \log\left(\frac{e^t + e^{-t}}{2}\right)$

$= \log\left(1 + \frac{t^2}{2!} + \frac{t^4}{4!} \cdots\right) \leq \log\left(e^{t^2/2}\right) = t^2/2$

$\Pr(X \geq \lambda) \leq e^{-\lambda^2/2}$

Def: X is σ-subgaussian if $\psi_X(t) \leq \underbrace{\psi_{Gaus(\sigma^2)}(t)}_{\frac{\sigma^2 t^2}{2}}$

$$\Rightarrow \Pr(X \geq \lambda) \leq e^{-\lambda^2/2\sigma^2}$$

. Rademacher are 1-subgaussian

$\Rightarrow$ Sq. of Rademacher have no heavier tails than sq. of Gaussians.