LECTURE 19

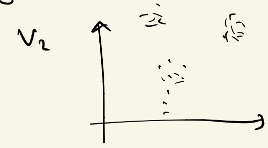
- Last class: How to visualize a high-dim. distribution? (e.g. human genome)

  $X \in \mathbb{R}^d$ r. vector, $\mathbb{E}X = 0$, $\mathrm{Cov}(X) = \mathbb{E}XX^T =: \Sigma$.

  e.g. X = genome of a random person in the world

  Eigenvectors v_i of Σ = "principal components" of X

- PCA: reduce dimension $\mathbb{R}^d \to \mathbb{R}^2$ by projecting X onto span$\{v_1, v_2\}$



- PCA assumes that we can compute the population cov. matrix

$$\Sigma = \mathrm{Cov}(X) = \mathbb{E}XX^T$$

average over population

But we don't have data of all population. We have:

- Finite sample $X_1, \dots, X_n \in \mathbb{R}^d$ iid copies of X. $\Rightarrow$ We approximate Σ by

$$\boxed{\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T.}$$ "Sample covariance matrix"

and hope that PCA(sample) $\approx$ PCA(population), i.e.

$$\lambda_i(\Sigma_n) \approx \lambda_i(\Sigma) \quad \text{and} \quad v_i(\Sigma_n) \approx v_i(\Sigma). \qquad (*)$$

- How large is n? $n = O(\log d)$? $n = O(d)$? $\underbrace{n = O(e^d)}$?

  Curse of H.D?

COVARIANCE ESTIMATION PROBLEM

- OUR GOAL $\boxed{n = O(d) \text{ suffices for } (*);}$ smaller for ~low rank.

- PLAN
  1. Approximate $\Sigma_n \approx \Sigma$ in operator norm
  2. Use perturbation theory to conclude $(*)$

- By **HW** (operator norm for symmetric matrices),


S^{d-1}

$$\|\Sigma_n - \Sigma\| = \max_{v \in S^{d-1}} \Big| \underbrace{v^T(\Sigma_n - \Sigma)v}_{= \, v^T\Sigma_n v - v^T \Sigma v} \Big| \qquad \text{where } S^{d-1} = \text{unit sphere in } \mathbb{R}^d$$

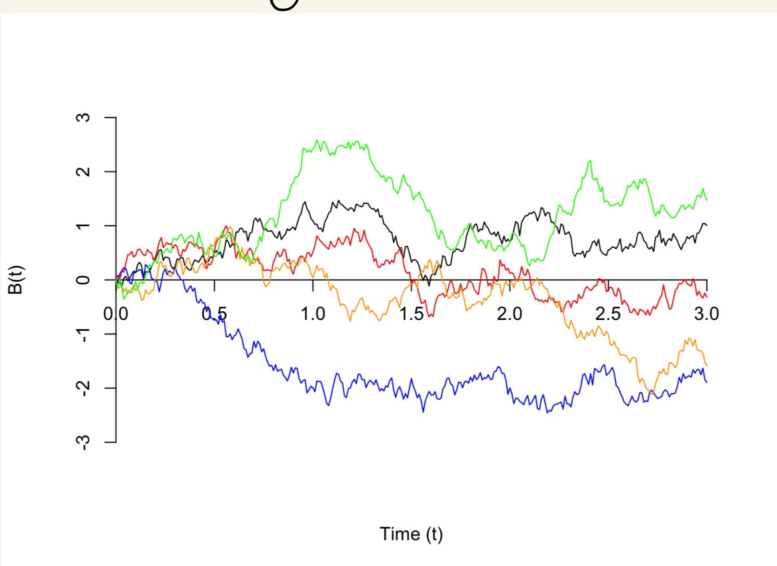
- $$v^T\Sigma v = v^T \mathbb{E} XX^T v = \mathbb{E}\, \underbrace{v^T X}_{= \langle X, v\rangle} \underbrace{X^T v}_{= \langle X, v\rangle} = \mathbb{E}\langle X, v\rangle^2$$

- $$v^T \Sigma_n v = \frac{1}{n}\sum_1^n \langle X_i, v\rangle^2$$

$$\Rightarrow \quad \|\Sigma_n - \Sigma\| = \max_{v \in S^{d-1}} \underbrace{\left| \frac{1}{n}\sum_1^n \langle X_i, v\rangle^2 - \mathbb{E}\langle X, v\rangle^2 \right|}_{=: \, Z(v) \text{ random variable.}}$$

- $(Z(v))_{v \in S^{d-1}}$ <u>Random process</u> indexed by $v \in S^{d-1}$.

- Compare to the <u>Brownian motion</u> (a.k.a. Wiener process) $(B(t))_{t \in [0,\infty)}$ indexed by time.



- $\mathbb{E}\max_{t \le T} |B(t)| \simeq \sqrt{T}$. $\qquad \mathbb{E}\max_{v \in S^{d-1}} |Z(v)| \le \, ?$

- Difficulty: a continuum of points in S^{d-1}. $\Rightarrow$ Discretize:

THE ε-NET METHOD

Prop (Finding a net)

$\forall \varepsilon > 0$, the unit sphere S^{d-1} has an ε-net $x_1, \dots, x_N$ with

$$\left(\frac{1}{\varepsilon}\right)^n \leq N \leq \left(\frac{2}{\varepsilon}+1\right)^d$$

↑ lecture 3

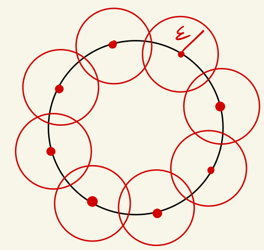
$\iff$ i.e. $\forall$ pt $\in S^{d-1}$ is within dist. ε from some x_i:

$\forall x \in S^{d-1} \ \exists i : \|x - x_i\|_2 \leq \varepsilon$

$\iff$ ε-balls centered at x_i cover S^{d-1}

- **Algorithm**:
  - Choose $\forall x_1$
  - Choose $\forall x_2$ at dist $> \varepsilon$ from x_1
  - Choose $\forall x_3$ at dist $> \varepsilon$ from $\{x_1, x_2\}$.
  - Choose $\forall x_4$ at dist $> \varepsilon$ from $\{x_1, x_2, x_3\}$
  - ....

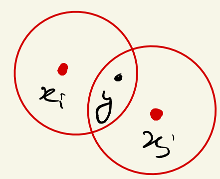
  STOP whenever impossible



$\Rightarrow$ cover

- **Claim**: The $(\varepsilon/2)$-balls centered at x_i are disjoint

If not, $\exists\, i \neq j, \exists\, y$: $\begin{cases} \|x_i - y\|_2 \leq \varepsilon/2 \\ \|x_j - y\|_2 \leq \varepsilon/2 \end{cases}$



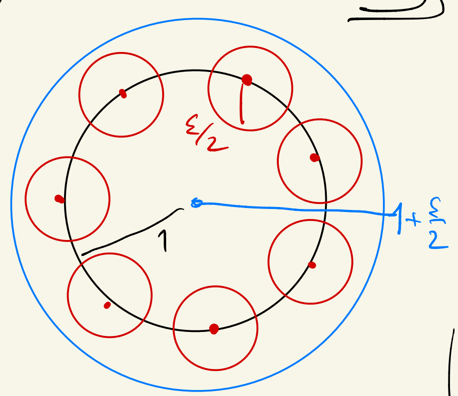
$\xRightarrow{\Delta} \|x_i - x_j\|_2 \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$

But all x_i are ε-separated by construction ↯

- All these balls lie in the $(1+\varepsilon/2)$-ball centered at 0 $\Rightarrow$

$$\mathrm{Vol}(B(1+\varepsilon/2)) \geq N \cdot \mathrm{Vol}(B(\varepsilon/2))$$

$$\Rightarrow \quad N \leq \frac{\mathrm{Vol}(B(1+\varepsilon/2))}{\mathrm{Vol}(B(\varepsilon/2))} = \left(\frac{1+\varepsilon/2}{\varepsilon/2}\right)^d = \left(\frac{2}{\varepsilon}+1\right)^d.$$



Remark Covering ≈ packing.

Prop (Computing the operator norm on a net)

Let A be an $m\times n$ matrix, $\mathcal{N}\subset S^{n-1}$ an ε-net. Then
$$\|A\| \le \frac{1}{1-\varepsilon}\max_{x\in\mathcal{N}}\|Ax\|$$

By def of operator norm, $\exists u\in S^{n-1}$:
$$\|Au\|_2 = \|A\|. \qquad (*)$$
By def of ε-net, $\exists x\in\mathcal{N}$:
$$\|x-u\|_2\le\varepsilon.$$
$$\Rightarrow \|Ax-Au\|_2 = \|A(x-u)\|_2 \le \|A\|\cdot\|x-u\|_2 \quad \text{(def of operator norm)}$$
$$\le \|A\|\cdot\varepsilon. \qquad (**)$$
$$\Rightarrow \|Ax\|_2 = \|Au-(Ax-Au)\|_2 \ge \|Au\|_2 - \|Ax-Au\|_2 \quad (\Delta\text{ ineq.})$$
$$\ge \|A\|-\|A\|\cdot\varepsilon \quad \text{(by } (*) \text{ and } (**))$$
$$= (1-\varepsilon)\|A\|.$$