Johnson Lindenstrauss

ECE 4200

PLAN:-

1. Recap PCA.
2. Johnosn Lindenstrauss
3. Clustering.

Recap of PCA.

Goal

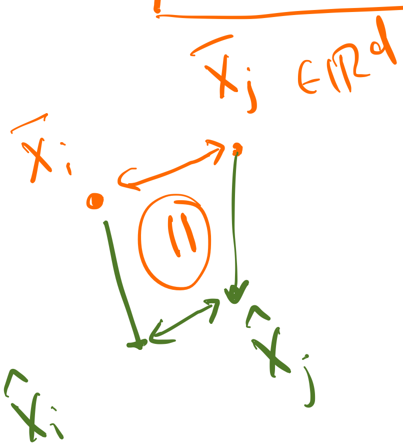
Perform a Linear projection to **preserve interpoint distances.**

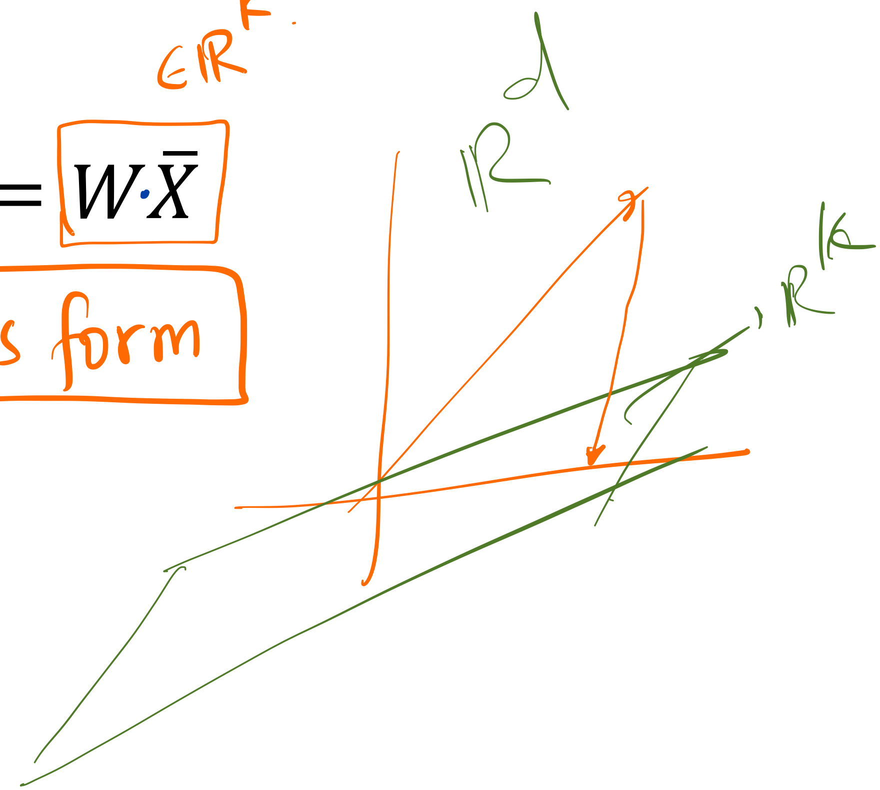
A **linear projection** from $\mathrm{R}^d \rightarrow \mathrm{R}^k$ is specified by a $\boxed{k \times d}$ matrix W.

$\in \mathbb{R}^k$

A point $\bar{X} \in \mathrm{R}^d$ is mapped to $\hat{\bar{X}} = \boxed{W \cdot \bar{X}}$

EX :- $\boxed{\text{Write PCA in this form}}$





Why **preserve interpoint distances?**

If all pairwise distances are preserved, the overall geometric structure is preserved

Faster computations in lower dimensions

$$S = \{(\bar{x}_i, y_i)\}, i = 1 \ldots n\} \longrightarrow \hat{S} = \{(\hat{x}_i, y_i), i = 1, \ldots n\}.$$

ε-distance preservation

Suppose $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n$ are n points in d (large) dimensions. A matrix $W: R^d \rightarrow R^k$ is ε-**distance** preserving if:

$$(1-\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2 \leq \left\|\hat{\bar{X}}_i - \hat{\bar{X}}_j\right\|_2^2 \leq (1+\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2$$

Interpoint distances are preserved up to a multiplicative factor ε

Q1: how small can k be made?

Q2: how to find such a W?

JL theorem

(ind of $\underline{\underline{d}}$)

Q1: Let $0.5 > \varepsilon > 0$ and $k > \frac{30 \log n}{\varepsilon^2}$.

There exists an ε-**distance** preserving map $W: R^d \to R^k$.

Q2: **Random** projection works

A random linear transform

Let W be a $k \times d$ matrix such that

$$W_{ij} \sim N\left(0, \frac{1}{k}\right).$$

$i = 1, \ldots k$
$j = 1, \ldots d$

Suppose we map each $\bar{X}_i$ to $W\bar{X}_i \in R^k$.

We will show that there is a ***W*** *that satisfies the theorem.*

Comparison to PCA

PCA:

- low-dimensional data
- Inherently in a subspace
- Time: $d^2 n + \boxed{d^3}$ → eigen-decomposition.
- Needs more memory (d^2)

Random projection:

- Faster (nkd) $\quad k \approx \log n$
- Data can come in and out with small memory
- Resulting dimension independent of d

$d \cdot n \log n$ → $\boxed{n.d}$

's' 'n' points in $\mathbb{R}^d$.

Pick 'w' at $\boxed{\text{random}}$ (Gaussian)

- interpoint ...

Step 1: Mean preservation

Let W be the $k \times d$ matrix such with each

$\mathbb{E}[W_{ij}] = 0$

$\mathbb{E}[W_{ij}^2] = \frac{1}{k}$ $\Leftarrow$ $W_{ij} \sim N\left(0, \frac{1}{k}\right)$.

(Dasgupta, Gupta, Simple proofs, Johnson ...)

Claim: Given a fixed $\bar{X}$, and W generated as above, then

$$\mathbb{E}[\|W\bar{X}\|_2^2] = \|\bar{X}\|_2^2 \quad \text{— Good.}$$

$\hat{X}$

$$\bar{W} = \begin{pmatrix} W_{11} & \cdots & W_{1d} \\ W_{k1} & \cdots & W_{kd} \end{pmatrix} \qquad \bar{X} = (\bar{X}^1, \ldots, \bar{X}^d) \qquad W \cdot \bar{X} = \begin{pmatrix} W_{11}\bar{X}^1 + W_{12}\bar{X}^2 + \cdots + W_{1d}\bar{X}^d \\ \vdots \\ W_{k1}\bar{X}^1 + W_{k2}\bar{X}^2 + \cdots + W_{kd}\bar{X}^d \end{pmatrix} \begin{matrix} \to V_1 \\ \\ \to V_k \end{matrix}$$

$\in k \times 1$

$$\mathbb{E}\left[\|W\bar{x}\|_2^2\right] = \mathbb{E}\left[\sum_{j=1}^{K}\left(w_{j1}\bar{x}^1 + w_{j2}\bar{x}^2 + \dots + w_{jd}\bar{x}^d\right)^2\right]$$

$$\mathbb{E}\left[w_{j1}^2(\bar{x}^1)^2 + w_{j2}^2(\bar{x}^2)^2 + \dots + w_{jd}^2(\bar{x}^d)^2 + \overset{0}{\cancel{w_{j1}w_{j2}\bar{x}^1\bar{x}^2}} + \dots\right]$$

$$= \frac{1}{K}\cdot\left((\bar{x}^1)^2 + (\bar{x}^2)^2 + \dots + (\bar{x}^d)^2\right) = \boxed{\frac{1}{K}}\cdot\|\bar{x}\|_2^2$$

$$\mathbb{E}\left[\|W\bar{x}\|_2^2\right] = \sum_{j=1}^{K}\frac{1}{K}\|\bar{x}\|_2^2 = \|\bar{x}\|_2^2.$$

Step 1: Mean preservation

Claim: Given a fixed $\bar{X}$, and W generated as above, then

$$\mathbb{E}[\|W\bar{X}\|_2^2] = \|\bar{X}\|_2^2$$

$\bar{x}_i - \bar{x}_j$

Using this, for any $\boxed{i, j}$

$$\mathbb{E}\left[\left\|\hat{\bar{X}}_i - \hat{\bar{X}}_j\right\|_2^2\right] = \mathbb{E}\left[\|W(\bar{X}_i - \bar{X}_j)\|_2^2\right] = \left\|\bar{X}_i - \bar{X}_j\right\|_2^2$$

Step2: Concentration

Claim: Suppose $k > \frac{30 \log n}{\varepsilon^2}$, when W is chosen at random

$$\Pr\left((1-\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2 \le \left\|\hat{\bar{X}}_i - \hat{\bar{X}}_j\right\|_2^2 \le (1+\varepsilon)\left\|\bar{X}_i - \bar{X}_j\right\|_2^2 \right) > 1 - \frac{1}{2n^2}$$

Once we have this we can finish the proof by the union bound.

Step2: $\boxed{\text{Concentration}}$ r.v.

For i, j, let

$$A_{ij}^{c} := \left\{(1-\varepsilon)\boxed{\|\bar{X}_i - \bar{X}_j\|_2^2} \le \left\|\hat{\bar{X}}_i - \hat{\bar{X}}_j\right\|_2^2 \le (1+\varepsilon)\boxed{\|\bar{X}_i - \bar{X}_j\|_2^2}\right\}.$$

$\bar{X}$

$\boxed{A_{ij}}$: event that for the distance between $\bar{X}_i$, and $\bar{X}_j$ is not preserved.

Then, we will show that

$$\boxed{\Pr(A_{ij}) < \frac{1}{2n^2}.}$$

UNION Bound

For any events $E_1, \ldots$

$$\Pr(E_1 \cup E_2 \cup \cdots) \leq \Pr(E_1) + \Pr(E_2) + \cdots$$

Applying this,

$$\Pr(\boxed{A_{12} \cup A_{13} \cup \cdots}) \leq \Pr(A_{12}) + \Pr(A_{13}) + \cdots$$
$$< \binom{n}{2}\frac{1}{2n^2} < \frac{1}{4}.$$

This shows that

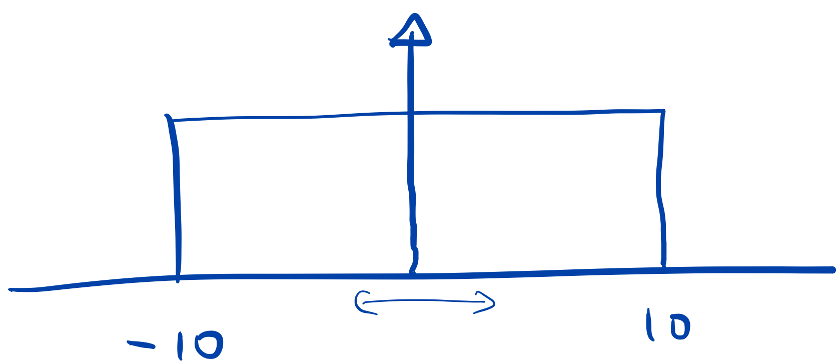
$$\boxed{\Pr\left((A_{12} \cup A_{13} \cup \cdots)^c\right) > \cancel{0}\,\frac{3}{4}}$$

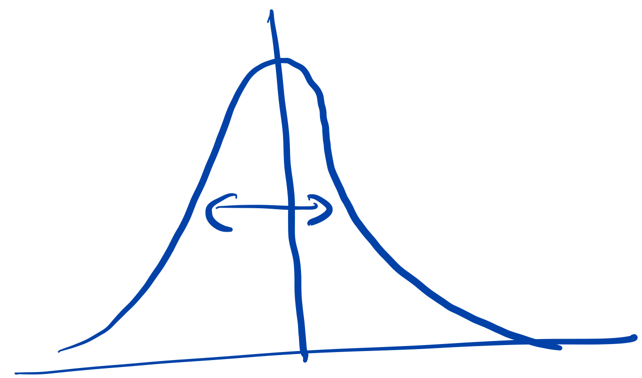
Showing that there is some W for which this event happens! (what is this event?)

How to prove concentration?

A general statement reads as:

A random variable X is pretty close to its mean, with **high probability.**





Markov Inequality

For a non-negative random variable X, and $\boxed{\alpha > 1:}$

$$\Pr(X > \underbrace{\alpha\mathbb{E}[X]}) < \frac{1}{\alpha}.$$

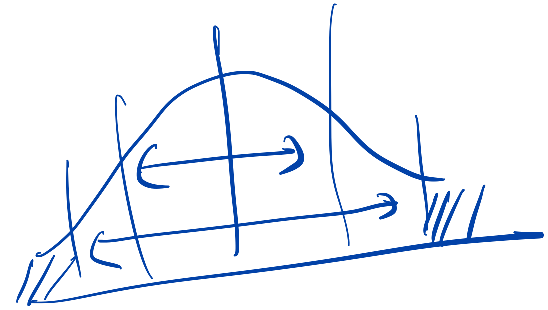
Eg, $\mathbb{E}[X] = 10$

$$\Pr(X > 2\underbrace{\mathbb{E}[X]}) < \frac{1}{2}.$$

$$\Pr(X > 20) < \frac{1}{2}.$$

PROOF: $\boxed{\text{HZ}}$

Chebychev's Inequality



VARIANCE implies CONCENTRATION.

For a random variable X with variance σ^2, and $\alpha > 1$:

$$\boxed{\Pr(|X - \mathbb{E}[X]| > \alpha \cdot \sigma) < \frac{1}{\alpha^2}.} \rightarrow e^{-\alpha}$$

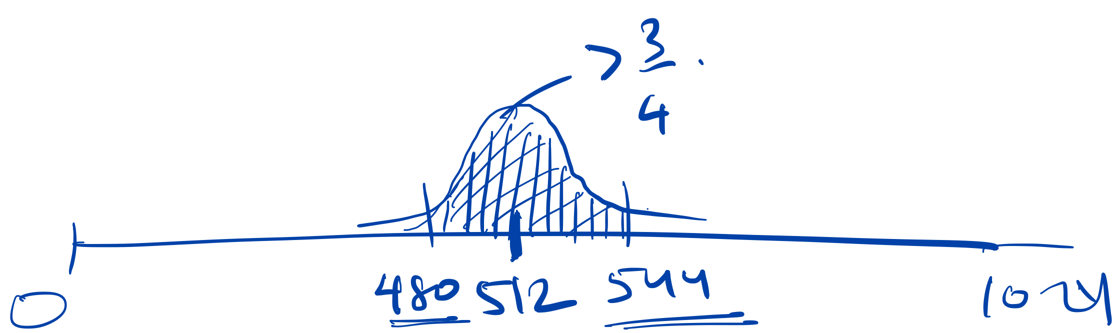
Eg, X: # Heads in 1024 coin tosses. $\boxed{\mathrm{Var}(X) = 256.}$ $\frac{1}{4} \times 1024$.

$$\Pr(|X - \mathbb{E}[X]| > 32) < \frac{1}{4}.$$

$\boxed{\sigma = 16}$

$\boxed{\alpha = 2}$

PROOF: HZ



To prove JL

Markov and Chebychev will be insufficient.

$\boxed{\textbf{Chernoff Bounds!}}$

$\| \hat{\bar{x}}_i - \hat{\bar{x}}_j \|_2^2 \rightarrow$ close to its $\| \bar{x}_i - x_j \|^2$

$Pr\left((1-\varepsilon) \|\bar{x}\|^2 < \| W \bar{x} \|_2^2 < (1+\varepsilon) \|\bar{x}\|^2 \right) \longrightarrow$ high.

$Pr\left(\| W \bar{x} \|_2^2 > (1+\varepsilon) \|\bar{x}\|_2^2 \right) \longrightarrow$ small.

$\Rightarrow \quad Pr\left(\left\| W \cdot \frac{\bar{x}}{\|\bar{x}\|_2} \right\|^2 > 1+\varepsilon \right)$

Chebychev's Inequality

Exponential bounds!

$$\frac{\overline{w} \cdot \boxed{\overline{x}}}{\|x\|_2} = \boxed{\overline{v} = (\overline{v}_1, \ldots, \overline{v}_k)} \rightarrow \boxed{\begin{aligned} E[\|\overline{v}\|^2] &= \sum E[(v_i)^2] \\ &= 1 \end{aligned}}$$

Claim :- $\overline{v}_i \sim N\left(0, \frac{1}{k}\right)$, ind.

$$\Pr(\underbrace{\|\overline{v}\|_2^2}_{} > 1+\varepsilon) \rightarrow \text{small}. \qquad v_1^2 + v_2^2 + \cdots + v_k^2$$

If 'k' is large, $\boxed{\text{small}}$

4/22/2020.

1. Recap JL, clustering,

Goal:- Linear projection to preserve inter-point distances.

Answer:- A random projection. $\mathbb{R}^d \to \mathbb{R}^k$

$\underbrace{\bar{x}_1, \dots, \bar{x}_n}$ $\in \mathbb{R}^d$ (d-dimensional data).

RP $\boxed{W \sim k \times d}$ matrix, $W_{ij} \sim N(0, \frac{1}{k})$

$\boxed{\hat{x}_i = W \cdot \bar{x}_i}$ is the projection. $\bar{x}_i \to \hat{x}_i$

If $k > \frac{30 \log n}{\varepsilon^2}$, then works!!!

For $\boxed{\text{each } i, j}$, $\|\hat{x}_i - \hat{x}_j\|_2^2 \in (1 \pm \varepsilon) \|\bar{x}_i - \bar{x}_j\|_2^2$ (goal).

---

1. $\underset{W}{\mathbb{E}} [\underbrace{\|W \cdot \bar{x}\|_2^2}] = \underbrace{\|\bar{x}\|_2^2} \longrightarrow$ projection preserves norm^2 in expection.

$\mathbb{E}[\|\hat{x}\|^2] =$

2. concentration. Take $\bar{x} = (1, 0, 0, \dots 0) \in \mathbb{R}^d$

then $W \cdot \bar{x} = [W_{11}\ W_{21} \dots W_{k1}]^T \rightarrow$ $\boxed{\mathbb{E}[W_{ij}^2] = \frac{1}{k}}$

$\mathbb{E}[\underbrace{\|W \cdot \bar{x}\|_2^2}] = \mathbb{E}(W_{11}^2 + \dots + W_{k1}^2) = 1 = \|\bar{x}\|^2.$

- we want $\underbrace{\|W \cdot \bar{x}\|_2^2} \in (1 \pm \varepsilon)$ with prob $\boxed{1 - \frac{1}{2n^2}}$

$\Leftrightarrow$ $\boxed{W_{11}^2 + \dots + W_{k1}^2 \in (1 \pm \varepsilon)}$

$N(0, \frac{1}{k})$

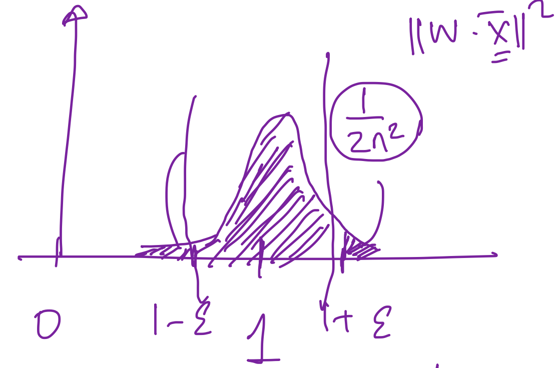
$$\begin{bmatrix} W_{11} & W_{12} & \dots & W_{1d} \\ \vdots & & & \\ W_{k1} & W_{k2} & \dots & W_{kd} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} W_{11} \\ \vdots \\ W_{k1} \end{bmatrix}$$

X

$X \sim N(0,\sigma^2)$

$a \cdot X \sim N(0, a^2\sigma^2)$

Let $v_i = W_{i1} \cdot \boxed{\sqrt{K}} \Rightarrow \boxed{v_i \sim N(0,1)}$



$$\Rightarrow \boxed{\frac{v_1^2 + \cdots + v_K^2}{K} \in 1 \pm \varepsilon}$$

guarantee this we need large k.

Sample mean of 'k' $N(0,1)^2$ random variables

once k- large enough good concentration

how large ?? $\rightarrow$ $\boxed{k \geq \frac{30 \log n}{\varepsilon^2}}$ $\Rightarrow$ $1 - \frac{1}{2n^2}$

$1 - \frac{1}{n^{20}}$

$*$ $v_1, \ldots, v_K \sim N(0,1)$,

$v_1^2, v_2^2, \ldots, v_K^2$

$$\underbrace{\frac{v_1^2 + v_2^2 + \cdots + v_K^2}{K}}$$

go over slides

- Dim reduction for supervised learning
- Denoising for ↓ & compression.
- Feature engineering.
- Some times leads to better generalization error.

$\bar{X}_1, \ldots, \bar{X}_n \longrightarrow \hat{X}_1, \ldots \hat{X}_n$

$\in \mathbb{R}^d \qquad \in \mathbb{R}^k$

$$\boxed{k = \frac{\log n}{\varepsilon^2}}$$

$\bar{W} \longrightarrow$ with probability $\boxed{1 - \frac{1}{n^{10}},}$ it succeeds.

$\boxed{\frac{1}{n^{10}}}$, $\frac{1}{50^{10}} \longrightarrow$

CLUSTERING