Machine Learning for Data Science (CS4786) Lecture 5

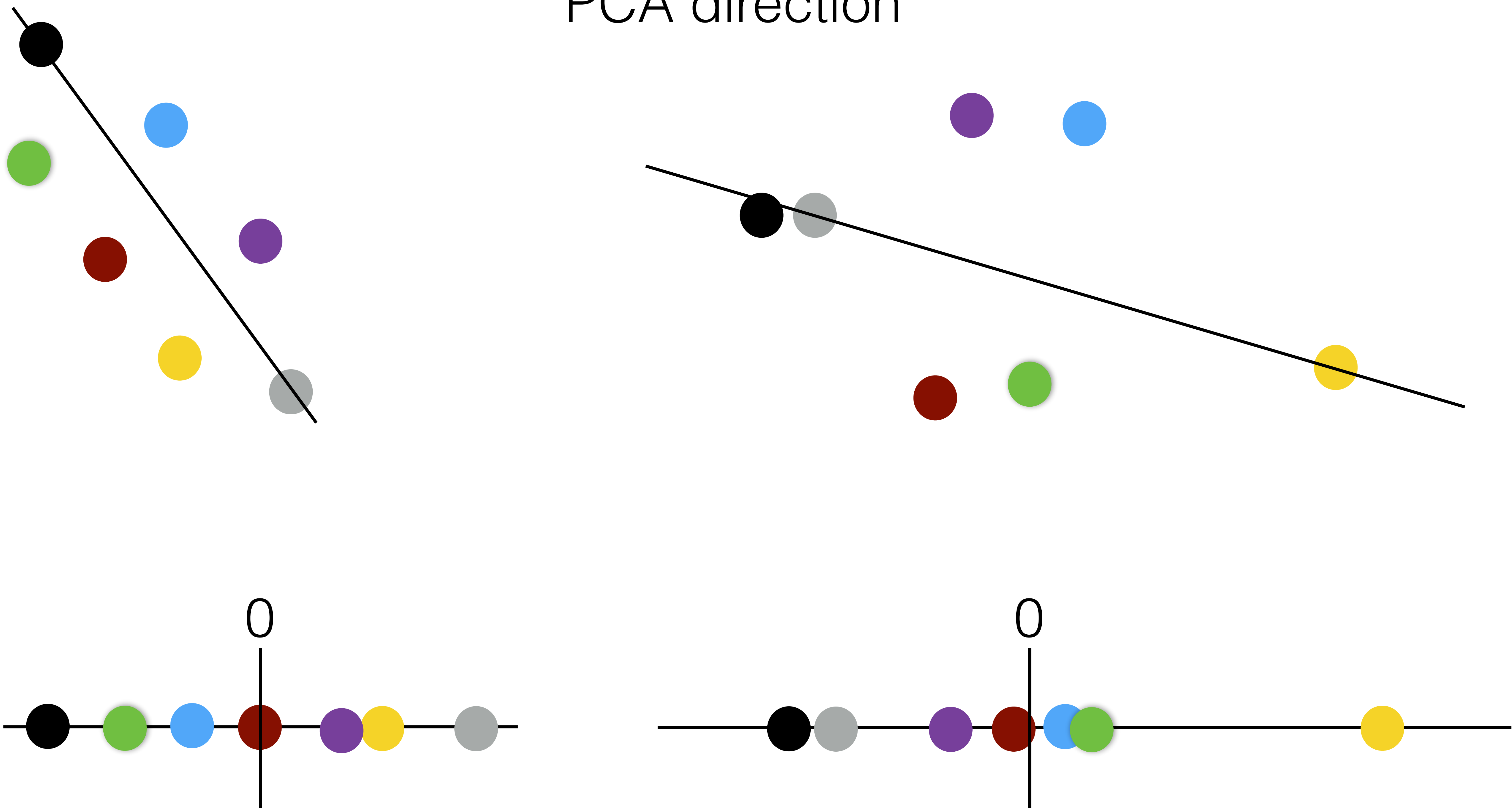
Random Projections

Course Webpage :

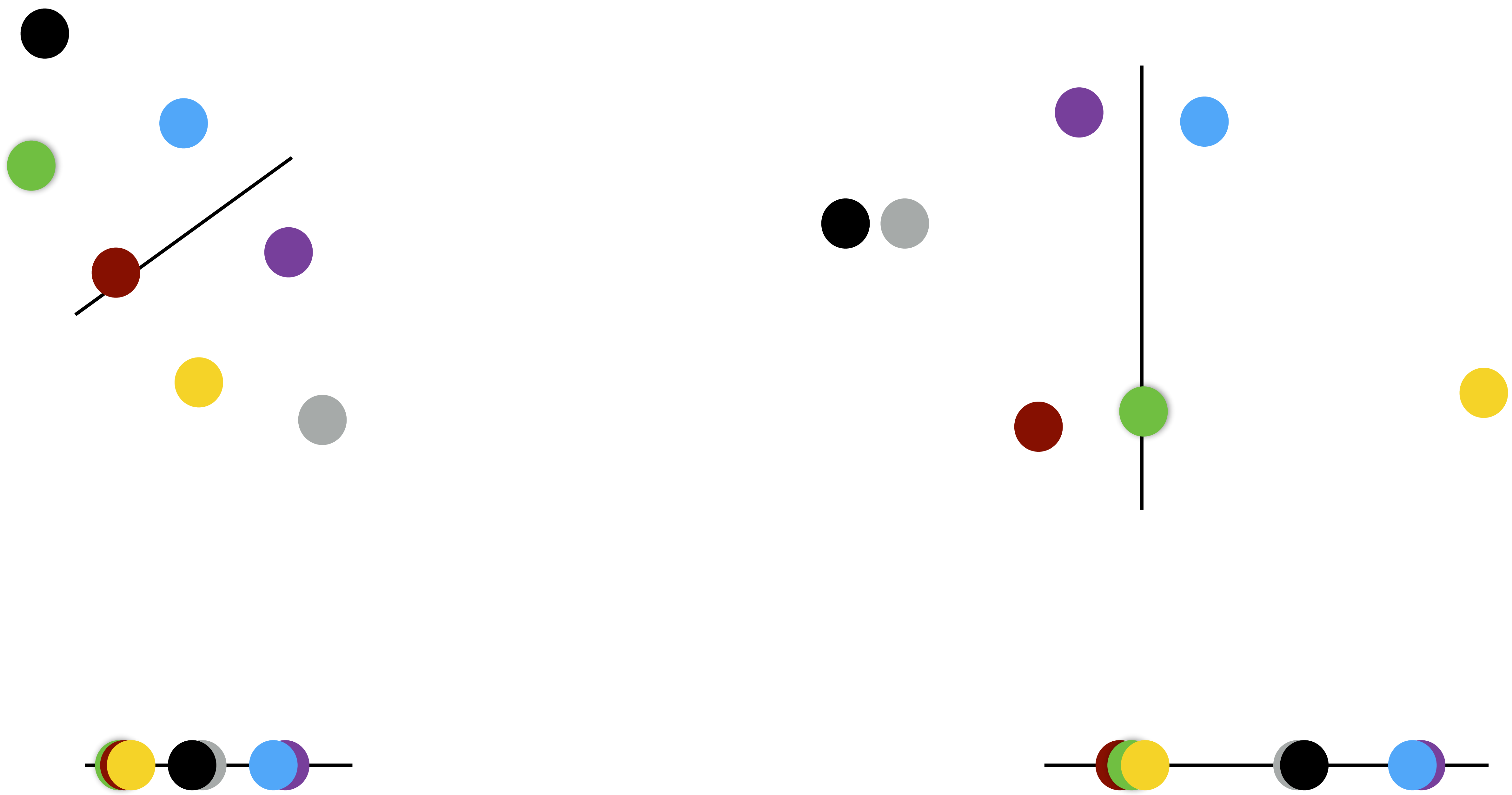
http://www.cs.cornell.edu/Courses/cs4786/2016fa/

Which Direction to Pick?

PCA direction



Which Direction to Pick?



Direction has large covariance

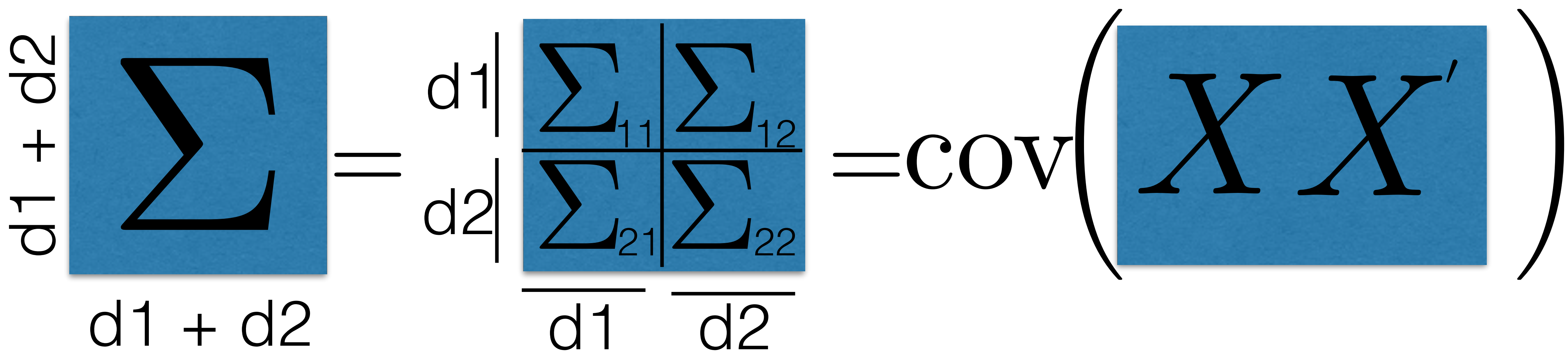
Maximizing Correlation Coefficient

- Say $\mathbf{w}_1$ and $\mathbf{v}_1$ are the directions we choose to project in views 1 and 2 respectively we want these directions to maximize,

$$\frac{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)\cdot\left(\mathbf{y}'_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t[1]\right)^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{y}'_t[1]-\frac{1}{n}\sum_{t=1}^{n}\mathbf{y}'_t[1]\right)}}$$

- Hence we want to solve for projection vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ that

$$\begin{aligned}&\text{maximize } \mathbf{w}_1^\top \Sigma_{1,2} \mathbf{v}_1 \\ &\text{subject to } \mathbf{w}_1^\top \Sigma_{1,1} \mathbf{w}_1 = \mathbf{v}_1^\top \Sigma_{2,2} \mathbf{v}_1 = 1\end{aligned}$$



Solution

$$\boxed{W_1} = \text{eigs}\left(\boxed{\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}, K\right)$$

$$\boxed{W_2} = \text{eigs}\left(\boxed{\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}, K\right)$$

CCA DEMO

i can't believe how awful is this movie i was expecting it to be really good especially with the actors that were in the cast this is depressing i'm so bummed that they ruined such a good plot

bummed to see such a bad game what an awful performance by everyone on the team as if everyone played to loose need to improve hitters more but fielders were also worse today one of the worst performance in the history of baseball

oh man this war movie was just too depressing for me some scenes were simply awful even though the plot closely follows the novel which i've read i was bummed at the end and had to secretly go cry

i will tell you what is wrong with it it is dead that's what is wrong with
about enough of this that team is definitely deceased tired and sha
after a long game you say look matey not a single soul in that lineup
a single ball even if i put 4000-volts through them they are bleeding
they are not pitching they passed on period plus pretty sure they m
awful elderberries after that game you should be depressed like me

this was so hilarious that's the best movie i've seen in a while i didn't know this actor before but he is so funny i was laughing from start to finish

it was hilarious to see playing these kids against experts throughout the game they were just running here and there and trying to get to the ball which they couldn't even once this was funny for viewers but organizers should ensure that inexperienced teams don't play against the experienced ones to keep the game interesting

dude that movie was so funny right i was laughing in like fits during some of the scenes i know the plot is supposed to be thought-provoking but i found it hilarious i really should stop laughing all the time but who cares right

now what seems to be the problem he says after leaning on the coach's limb body after a fast pitch struck him during the game his face was icy serious not laughing at all unlike everyone else jen said 'it is the coach he is not moving at all is he dead he said slowly course not we answered laughing again thank god

well that was a funny movie i enjoyed the plot with all those twists you never knew what was going to happen especially in this last scene i wasn't expecting this outcome at all haha

was it a game at all i felt as if everyone was just trying to stay warm by making as little move as possible laziness of fielders was making it appear as if they were running in 0.5x speed mode haha strikers made good use of pitch they got and it was an easy win

lol i can't even sit properly now i have a tummy ache because of all the rofling that actor's head looked like a volcano haha i swear it looked like it was about to erupt and his brains would spill out haha

fans at the game are encouraged to get out of their seats stretch a bit and sing take me out to the ball game that is the closest baseball gets to a halftime haha

really love that movie we saw yesterday i was really excited since i knew it was going to be released this week and i haven't been disappointed at all i especially enjoyed the acting of the actors they were so good

what an awesome game it was dwight evans set the path to unprecedented victory when he made his very first strike on the pitch he alone made the whole game enjoyable excited for the next match

omg i totally loved yesterday's movie we were all so excited to finally catch the third movie after months of scouring the fan pages for the plot there are mixed opinions on the acting but i think the actors did a brilliant job overall

80 years old and was still playing the game stuff like this keeps you excited motivated you know yes he did break his back walking to the pitch to take the strike but you know everyone has to expire and go to their maker at some point he was lucky to do it while doing something he loved i am sure he enjoyed every second of it we should learn to enjoy this

Recap

Dimensionality Reduction

Given feature vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^K$ where $K << d$

Principal Component Analysis:

- **Find directions that maximize variance (spread)**
- **Find directions that minimize reconstruction error**

Principal Component Analysis

1. $$\Sigma = \mathrm{cov}\left(X \right)$$

2. $$W = \mathrm{eigs}\left(\Sigma, K \right)$$

3. $$Y = X - \mu \times W$$

Reconstruction

4. $$\widehat{X} = Y \times W^\top + \mu$$

Two View Dimensionality Reduction

- Data can be split into pairs $(\mathbf{x}_1, \mathbf{x}'_1), \ldots, (\mathbf{x}_n, \mathbf{x}'_n)$ where $\mathbf{x}_t$'s are d_1 dimensional and $\mathbf{x}'_t$'s are d_2 dimensional

- Goal: Compress $\mathbf{x}_1, \ldots, \mathbf{x}_n$ into K dimensional vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ (or $\mathbf{x}'_1, \ldots, \mathbf{x}'_n$ into $\mathbf{y}'_1, \ldots, \mathbf{y}'_n$ or both)
  - Retain information redundant between the two views

Canonical Correlation Analysis:

- **Find directions that maximize correlations between the projections in the two views**

CCA Algorithm

1. $$X = \Big({}_{n}\, \underset{d_1}{X_1}, \underset{d_2}{X_2} \Big)$$

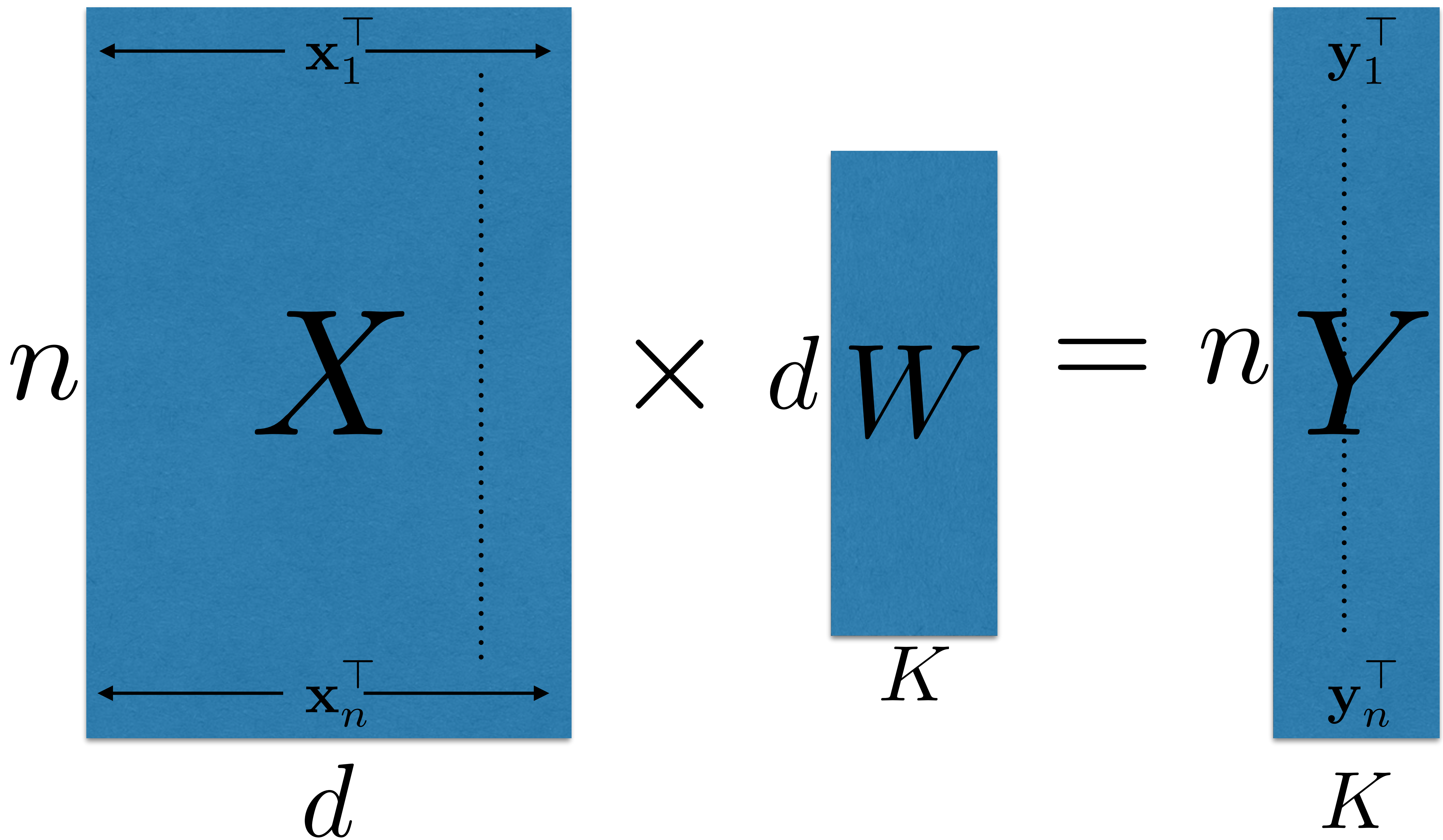
2. $$\Sigma = \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix} = \mathrm{cov}\Big(X \Big)$$

3. $$W_1 = \mathrm{eigs}\Big(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, K \Big)$$

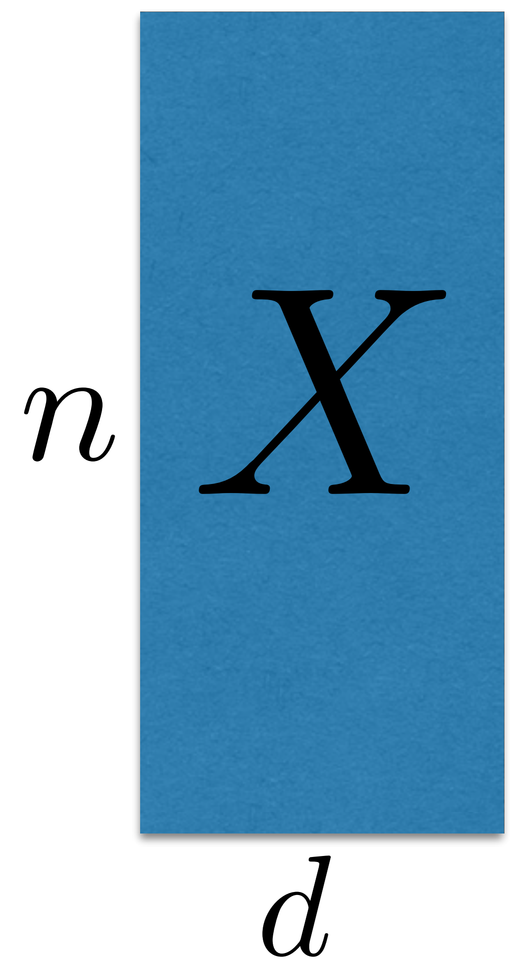
CCA Algorithm

4. $$Y_1 = X_1 - \mu_1 \times W_1$$

Back to Single View: Recap



The Tall, the Fat and the Ugly

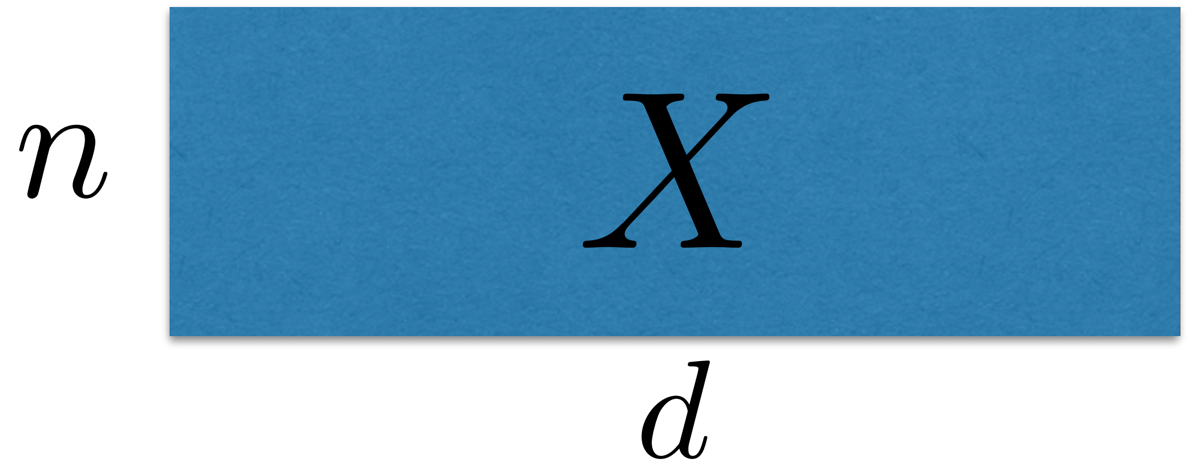


The Tall, the Fat and the Ugly

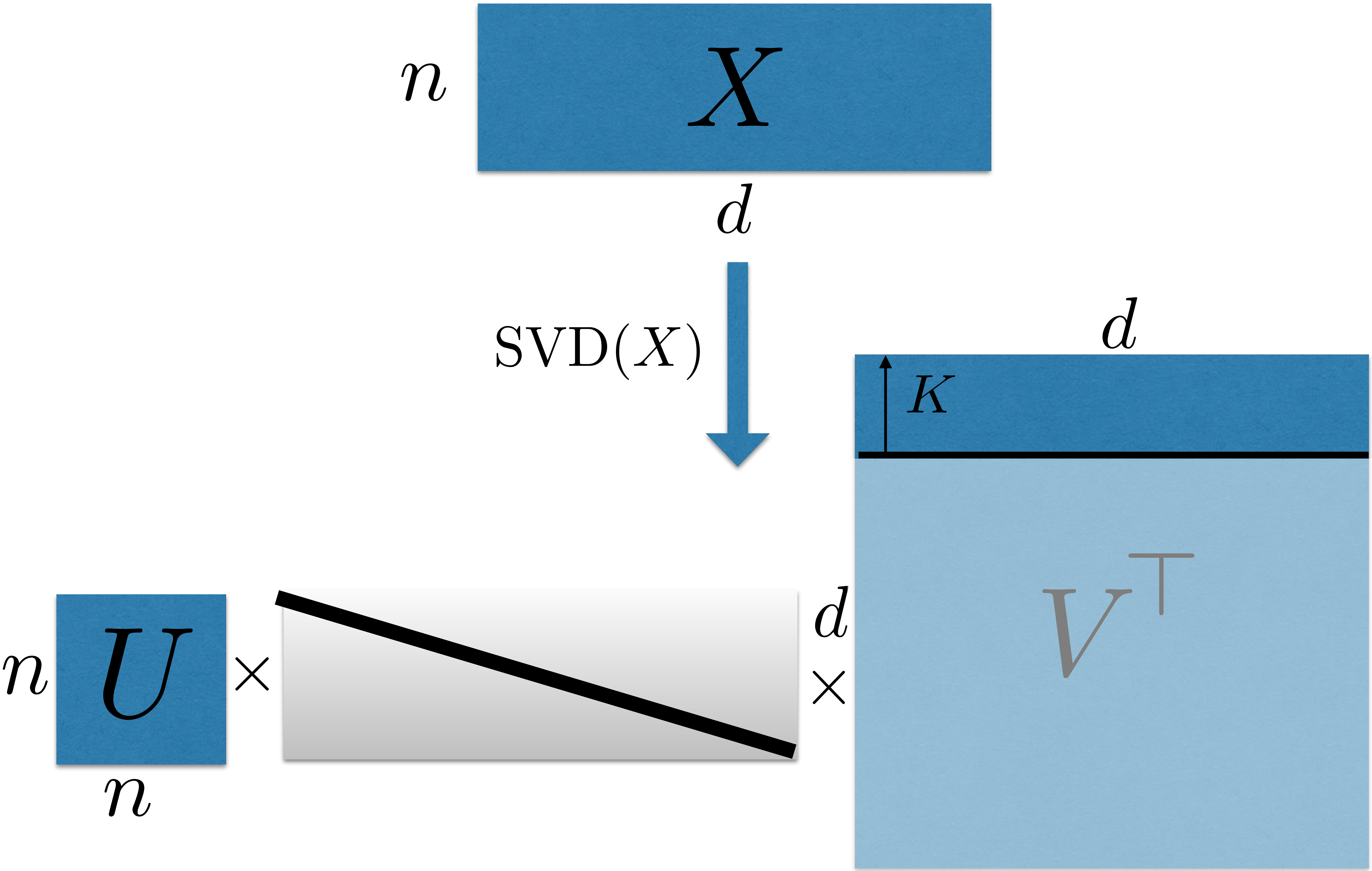
$$ {}_{d}\underset{n}{X^{\top}} \times {}_{n}\underset{d}{X} \Big/ n = {}_{d}\overset{d}{\Sigma} $$

$$ {}_{d}\underset{K}{W} = \mathrm{Eigs}\left(\Sigma, K \right) $$

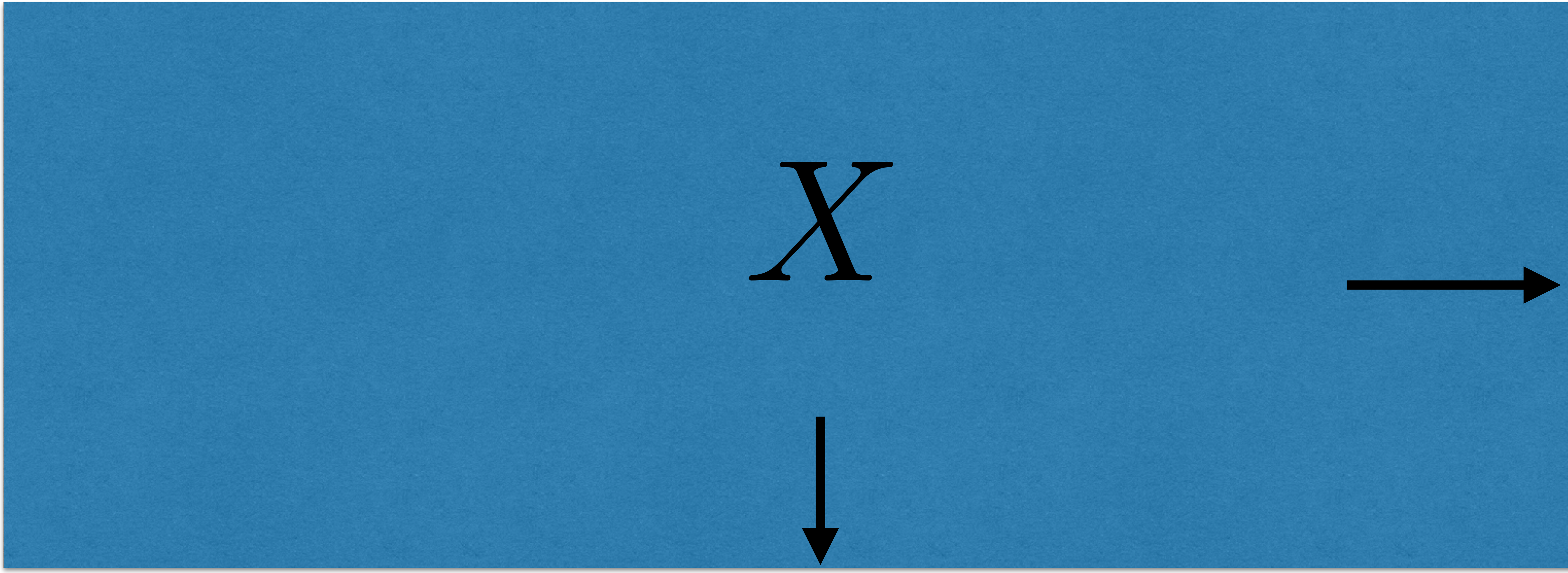
The Tall, **the Fat** and the Ugly



The Tall, **the Fat** and the Ugly



THE TALL, THE FAT AND **the Ugly**



- d and n so large we can't even store in memory
- Only have time to be linear in $\text{size}(X) = n \times d$

I there any hope?

Pick a Random W

$$Y = X \times \begin{array}{cc} \begin{bmatrix} +1 & \dots & -1 \\ -1 & \dots & +1 \\ +1 & \dots & -1 \\ & \vdots & \\ +1 & \dots & -1 \end{bmatrix} & d \\ K & \end{array} \Bigg/ \sqrt{K}$$

Random Projection

- What does "it works" even mean?

Distances between all pairs of data-points in low dim. projection is roughly the same as their distances in the high dim. space.

That is, when K is "large enough", with "high probability", for all pairs of data points $i, j \in \{1, \ldots, n\}$,

$$(1-\epsilon)\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2 \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|_2 \leq (1+\epsilon)\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2$$

Why should Random Projections even work?!

Say $K = 1$. Consider any vector $\tilde{\mathbf{x}} \in \mathbb{R}^d$ and let $\tilde{\mathbf{y}} = \tilde{\mathbf{x}}\, W$. Note that

$$\tilde{\mathbf{y}}^2 = \left(\sum_{i=1}^{d} W[i,1] \cdot \tilde{\mathbf{x}}[i]\right)^2$$

$$= \sum_{i=1}^{d} (W[i,1] \cdot \tilde{\mathbf{x}}[i])^2 + 2 \sum_{i'>i} (W[i,1] \cdot \tilde{\mathbf{x}}[i]) \left(W[i',1] \cdot \tilde{\mathbf{x}}[i']\right)$$

$$= \sum_{i=1}^{d} W^2[i,1]\tilde{\mathbf{x}}^2[i] + \sum_{i'>i} \left(W[i,1] \cdot W[i',1]\right) \cdot \left(\tilde{\mathbf{x}}[i] \cdot \tilde{\mathbf{x}}[i']\right)$$

However $W^2[i,1] = 1/K = 1$ when $K = 1$

$$= \sum_{i=1}^{d} \tilde{\mathbf{x}}^2[i] + \sum_{i'>i} \left(W[i,1] \cdot W[i',1]\right) \cdot \left(\tilde{\mathbf{x}}[i] \cdot \tilde{\mathbf{x}}[i']\right)$$

WHY SHOULD RANDOM PROJECTIONS EVEN WORK?!

Hence,

$$\mathbb{E}\left[\tilde{\mathbf{y}}^2\right] = \sum_{i=1}^{d} \tilde{\mathbf{x}}^2[i] + \sum_{i'>i} \mathbb{E}\left[W[i,1] \cdot W[i',1]\right] \cdot \left(\tilde{\mathbf{x}}[i] \cdot \tilde{\mathbf{x}}[i']\right)$$

However $W[i,1]$ and $W[i',1]$ are independent and so

$$\mathbb{E}\left[W[i,1] \cdot W[i',1]\right] = \mathbb{E}[W[i,1]] \cdot \mathbb{E}\left[W[i',1]\right] = 0$$

Using this we conclude that

$$\mathbb{E}\left[\tilde{\mathbf{y}}^2\right] = \sum_{i=1}^{d} \tilde{\mathbf{x}}^2[i] = \|\tilde{\mathbf{x}}\|^2$$

Why should Random Projections even work?!

Hence,

$$\mathbb{E}\left[|\tilde{\mathbf{y}}|^2\right] = \|\tilde{\mathbf{x}}\|_2^2$$

If we let $\tilde{\mathbf{x}} = \mathbf{x}_s - \mathbf{x}_t$ then

$$\tilde{\mathbf{y}} = \tilde{\mathbf{x}}W = \mathbf{x}_s W - \mathbf{x}_t W = \mathbf{y}_s - \mathbf{y}_t$$

Hence for any $s, t \in \{1, \ldots, n\}$,

$$\mathbb{E}\left[|\mathbf{y}_s - \mathbf{y}_t|^2\right] = \|\mathbf{x}_s - \mathbf{x}_t\|_2^2$$

Lets try this in Matlab ...

Why should Random Projections even work?!

- Setting K large is like getting K samples.
- Specifically since we take W to be random signs normalized by $\sqrt{K}$, for each $j \in [K]$, for any $\tilde{\mathbf{x}}$ if $\tilde{\mathbf{y}} = \tilde{\mathbf{x}}\, W$, then

$$\mathbb{E}\left[\tilde{\mathbf{y}}^2[j]\right] = \|\tilde{\mathbf{x}}\|_2^2 / K$$

Hence we can conclude that

$$\mathbb{E}\left[\sum_{j=1}^{K} \tilde{\mathbf{y}}^2[j]\right] = \sum_{j=1}^{K} \mathbb{E}\left[\tilde{\mathbf{y}}^2[j]\right] = \sum_{j=1}^{K} \frac{\|\tilde{\mathbf{x}}\|_2^2}{K} = \|\tilde{\mathbf{x}}\|_2^2$$

This is like taking an average of K independent measurements whose expectations are $\|\tilde{\mathbf{x}}\|_2^2$

Why should Random Projections even work?!

For large K, not only true in expectation but also with high probability

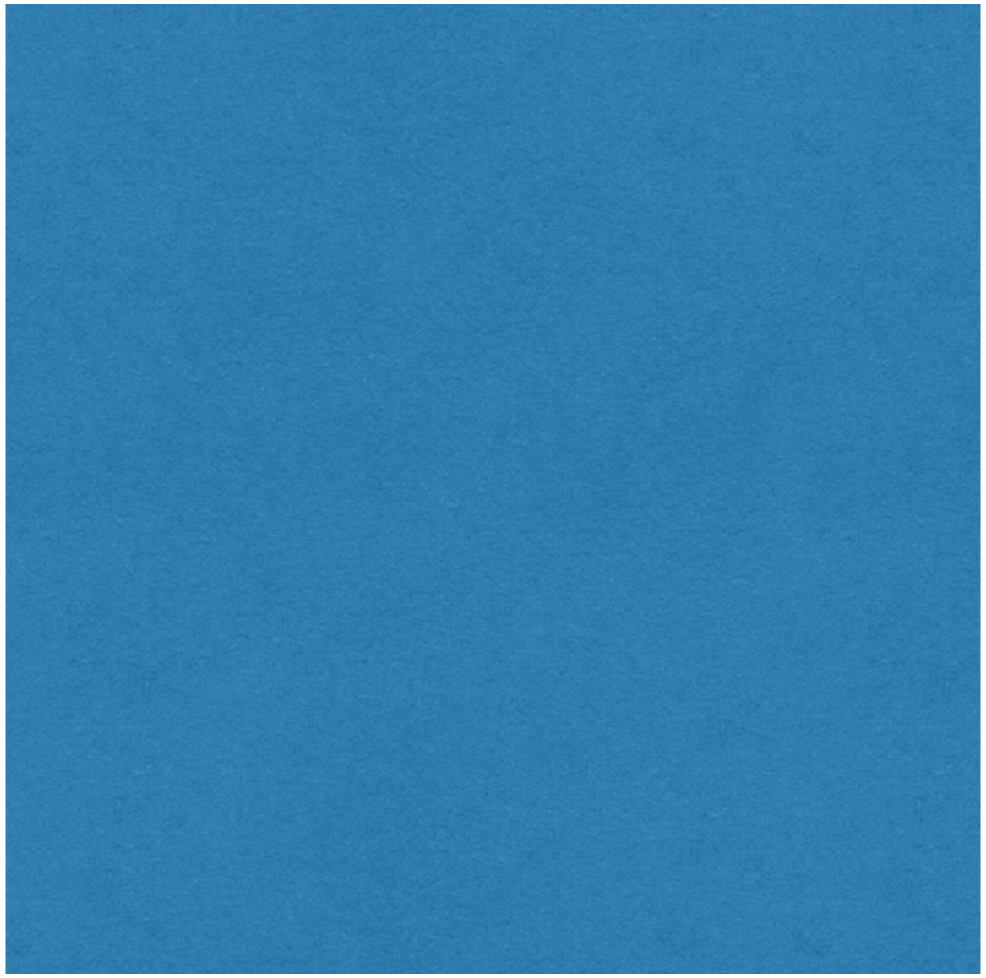
For any $\epsilon > 0$, if $K \approx \log(n/\delta)/\epsilon^2$, with probability $1-\delta$ over draw of W, for all pairs of data points $i, j \in \{1, \ldots, n\}$,

$$(1-\epsilon)\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^2 \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|_2 \leq (1+\epsilon)\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^2$$

Lets try on Matlab . . .

This is called the Johnson-Lindenstrauss lemma or JL lemma for short.

Why is this so Ridiculously Magical?

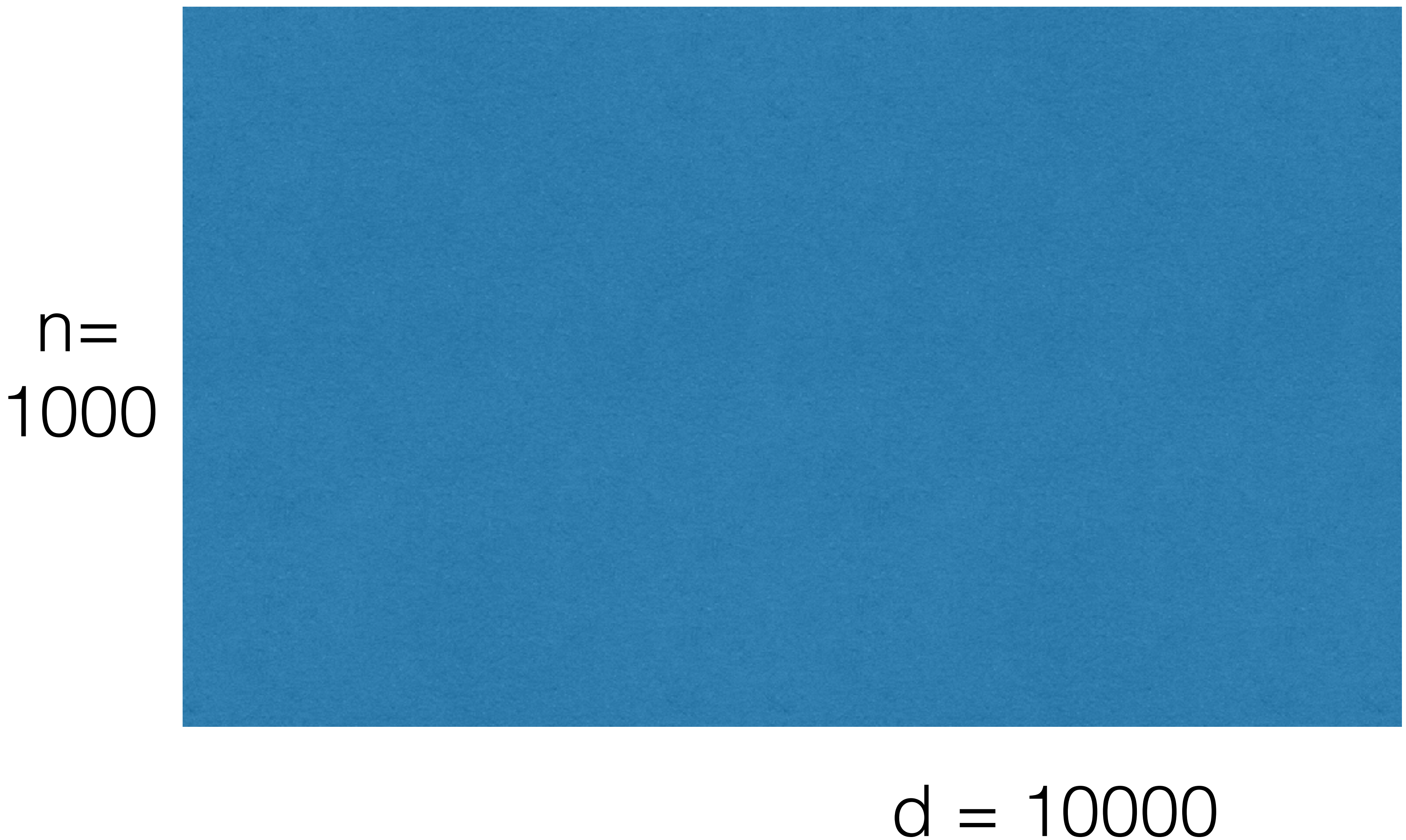


n=
1000

d = 1000

If we take $K = 69.1/\epsilon^2$, with probability 0.99 distances are preserved to accuracy ϵ

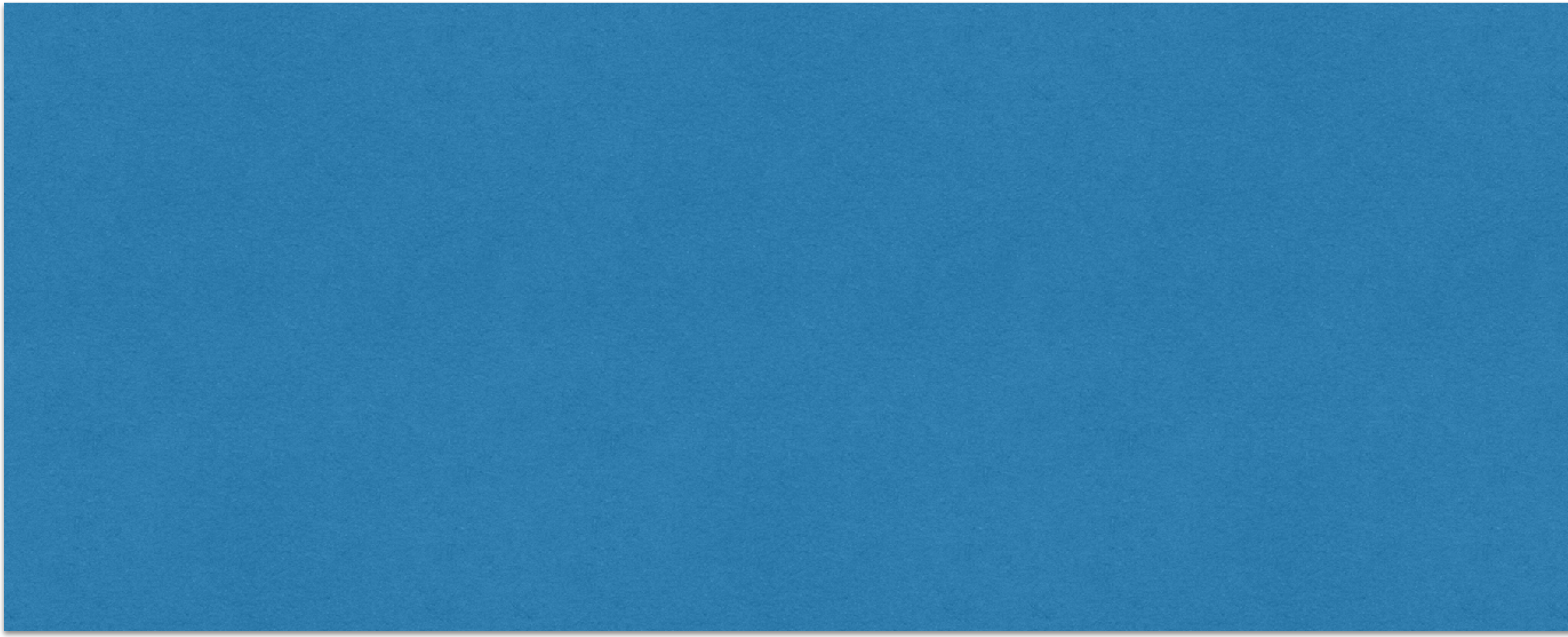
Why is this so Ridiculously Magical?



If we take $K = 69.1/\epsilon^2$, with probability 0.99 distances are preserved to accuracy ϵ

Why is this so Ridiculously Magical?

n=
1000



d = 1000000

If we take $K = 69.1/\epsilon^2$, with probability 0.99 distances are preserved to accuracy ϵ