Machine Learning for Data Science (CS4786) Lecture 6

Random Projections

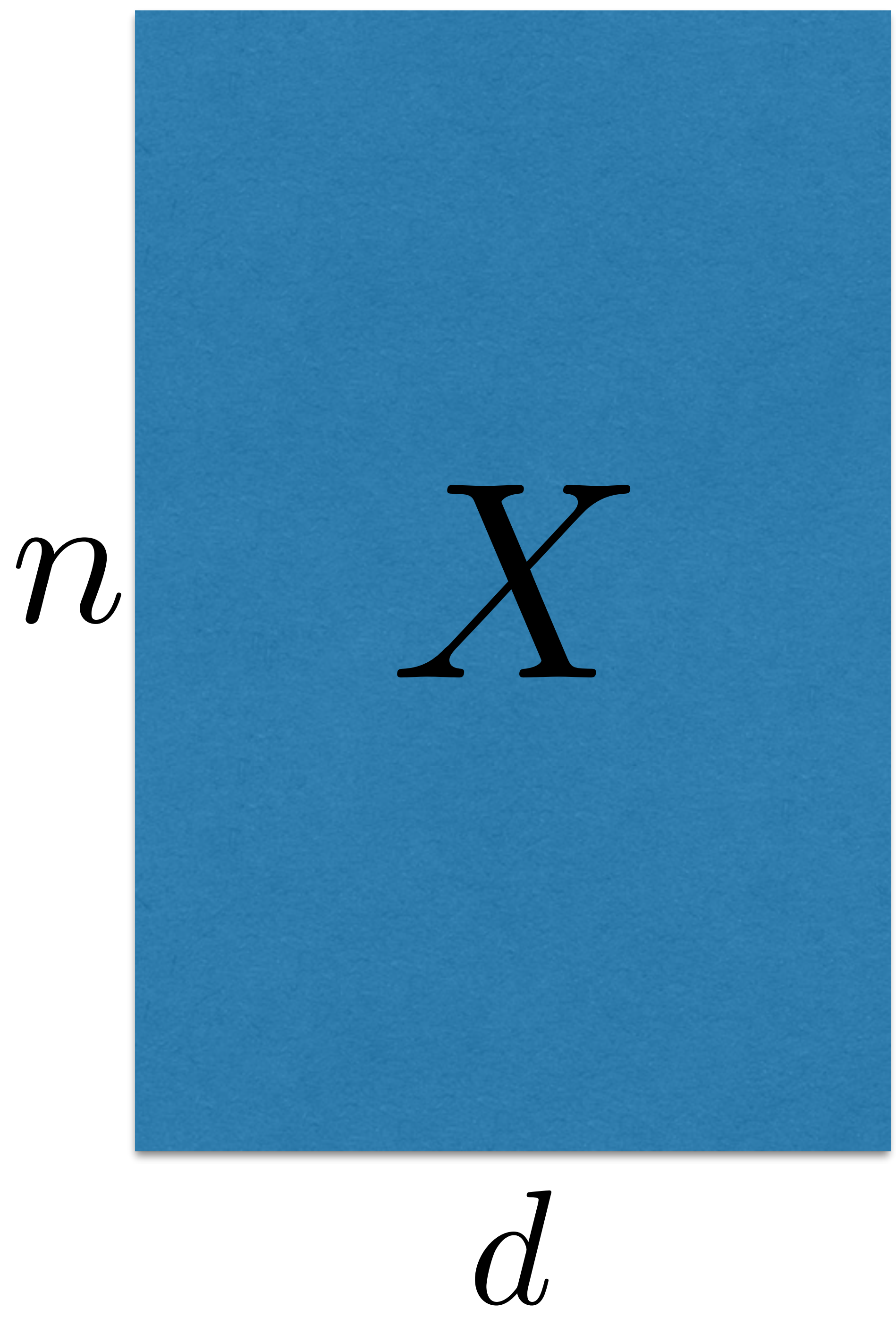
Feb 09, 2015

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2015sp/

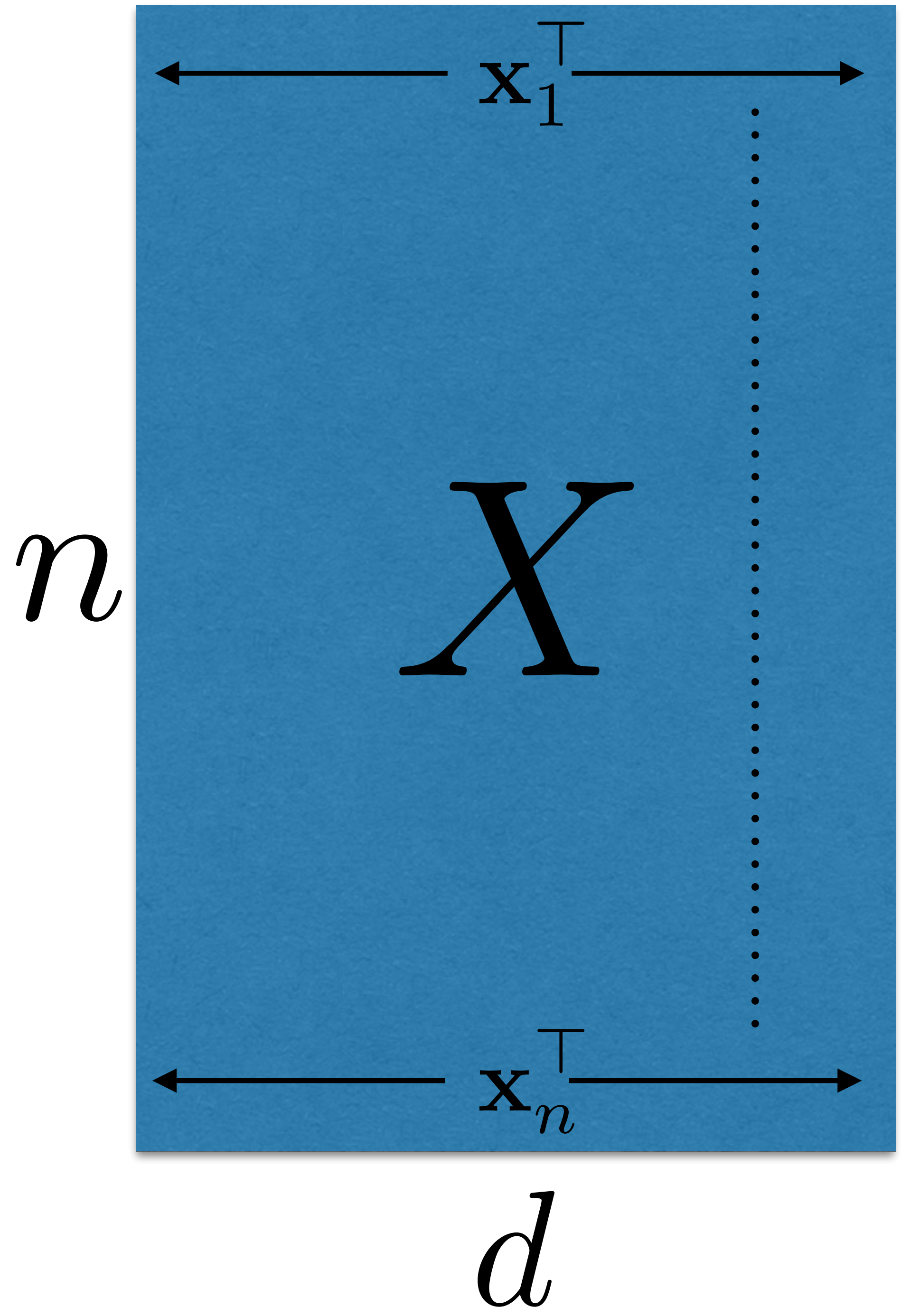
CCA Demo Redo?

BACK TO SINGLE VIEW: RECAP

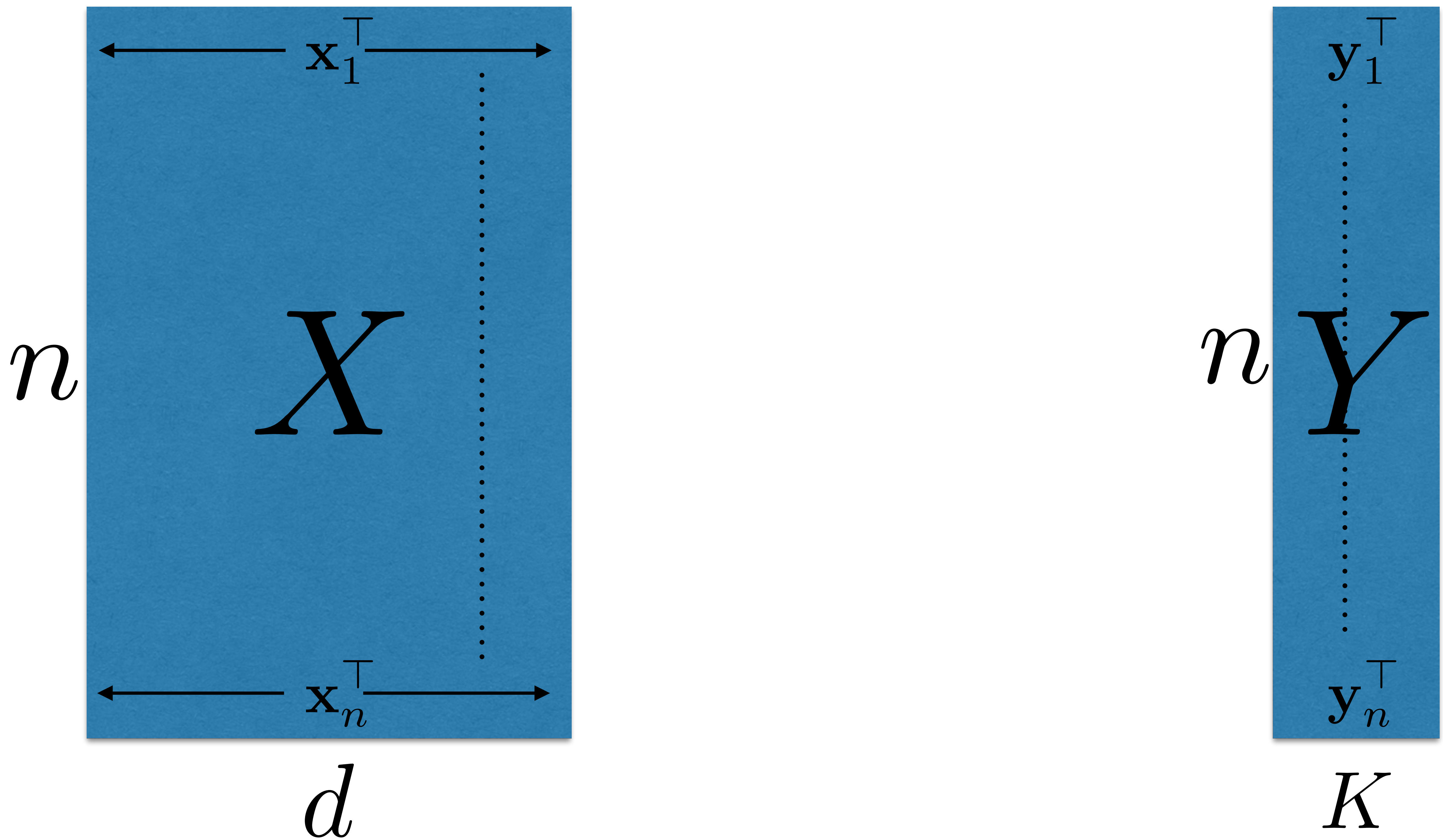
Back to Single View: Recap



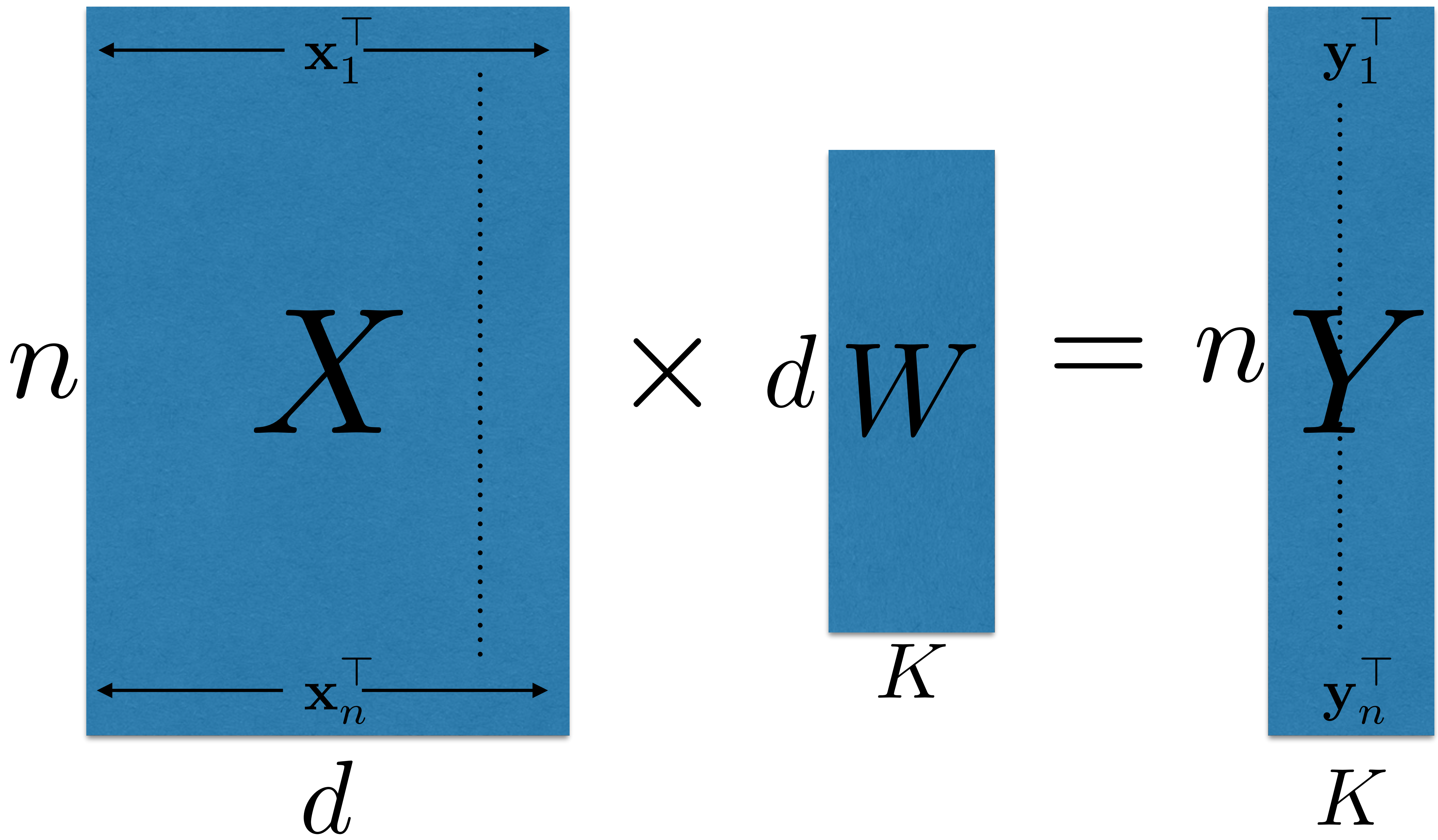
Back to Single View: Recap



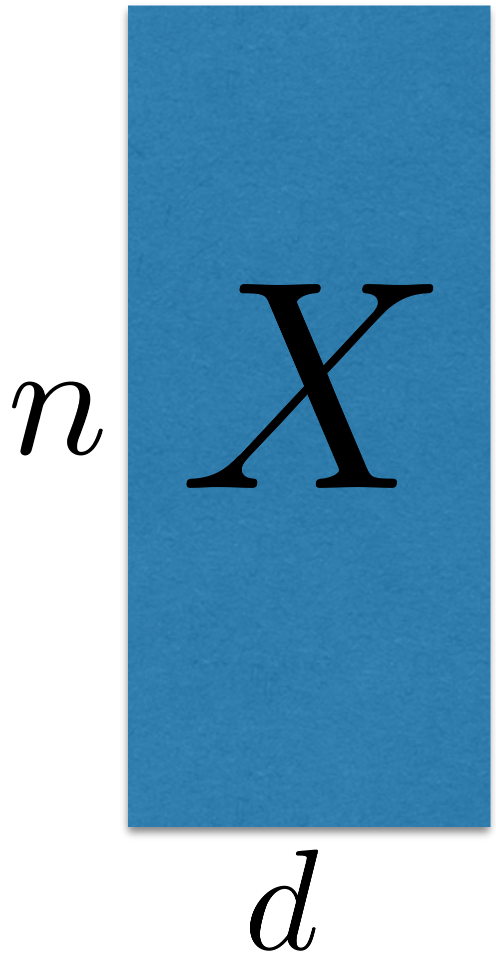
Back to Single View: Recap



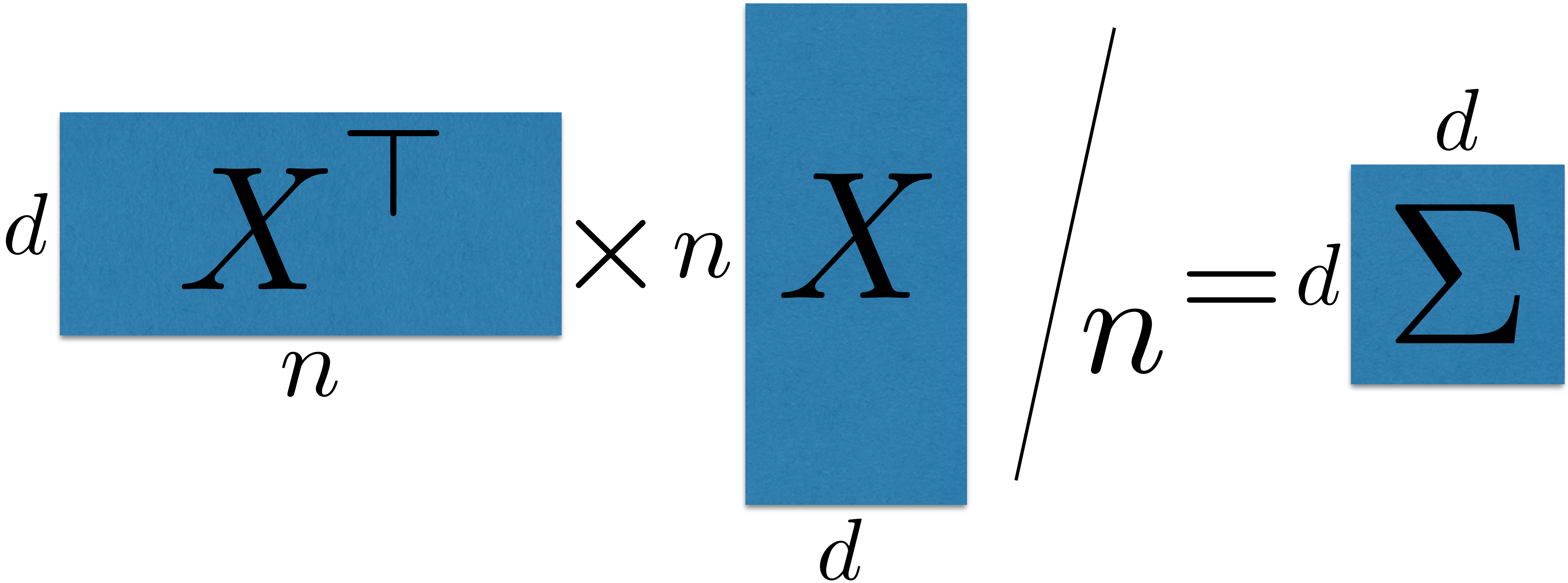
Back to Single View: Recap



The Tall, the Fat and the Ugly



The Tall, the Fat and the Ugly

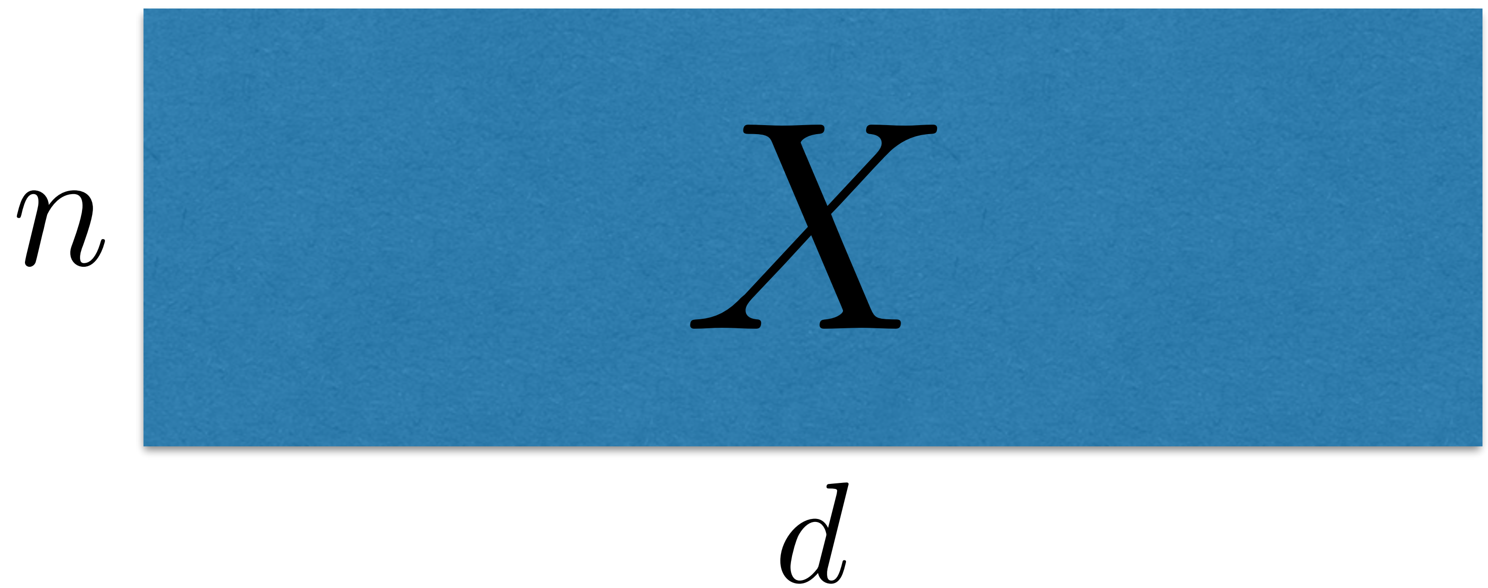


The Tall, the Fat and the Ugly

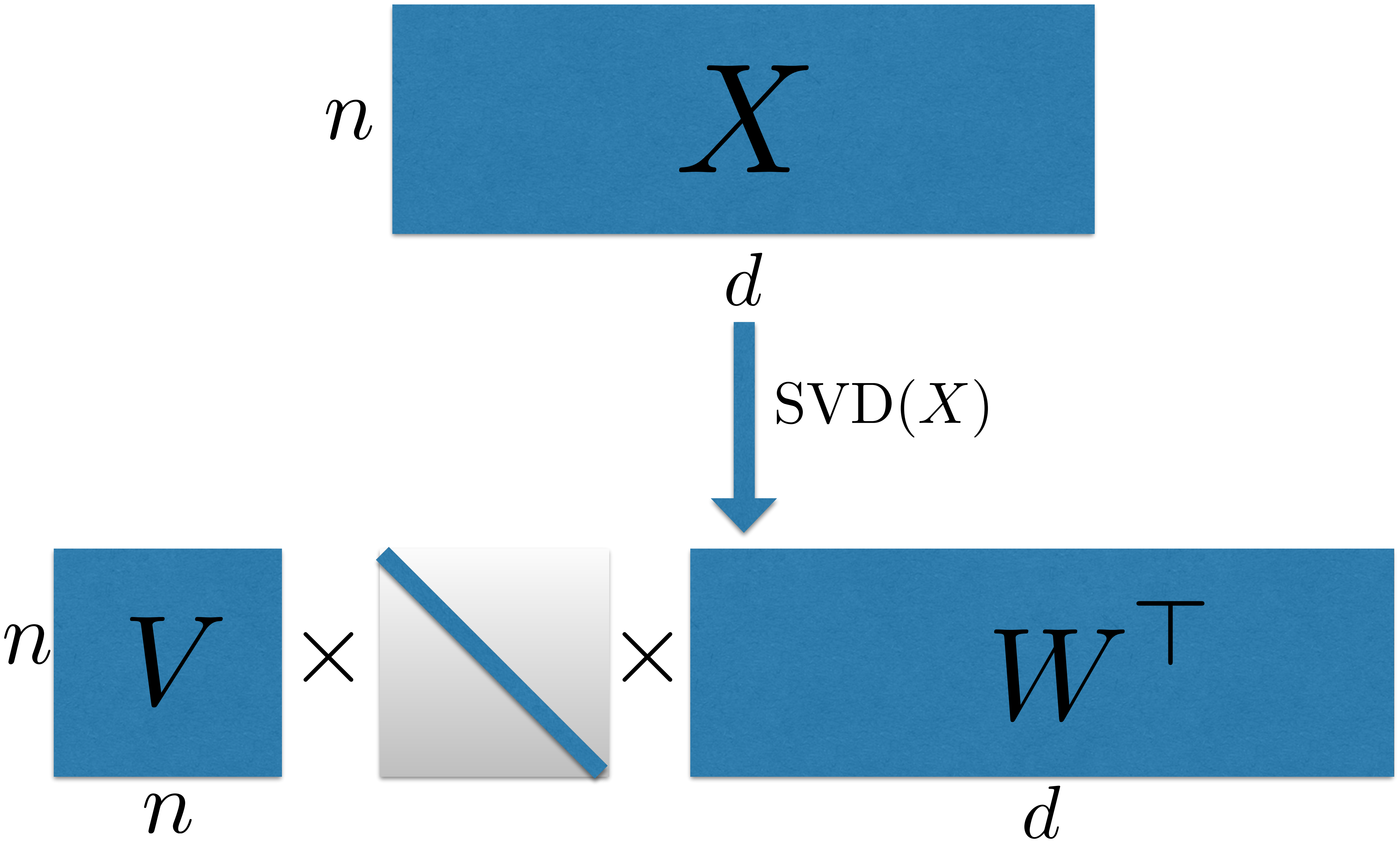
$$\underset{d \times n}{X^\top} \times \underset{n \times d}{X} \Big/ n = \underset{d \times d}{\Sigma}$$

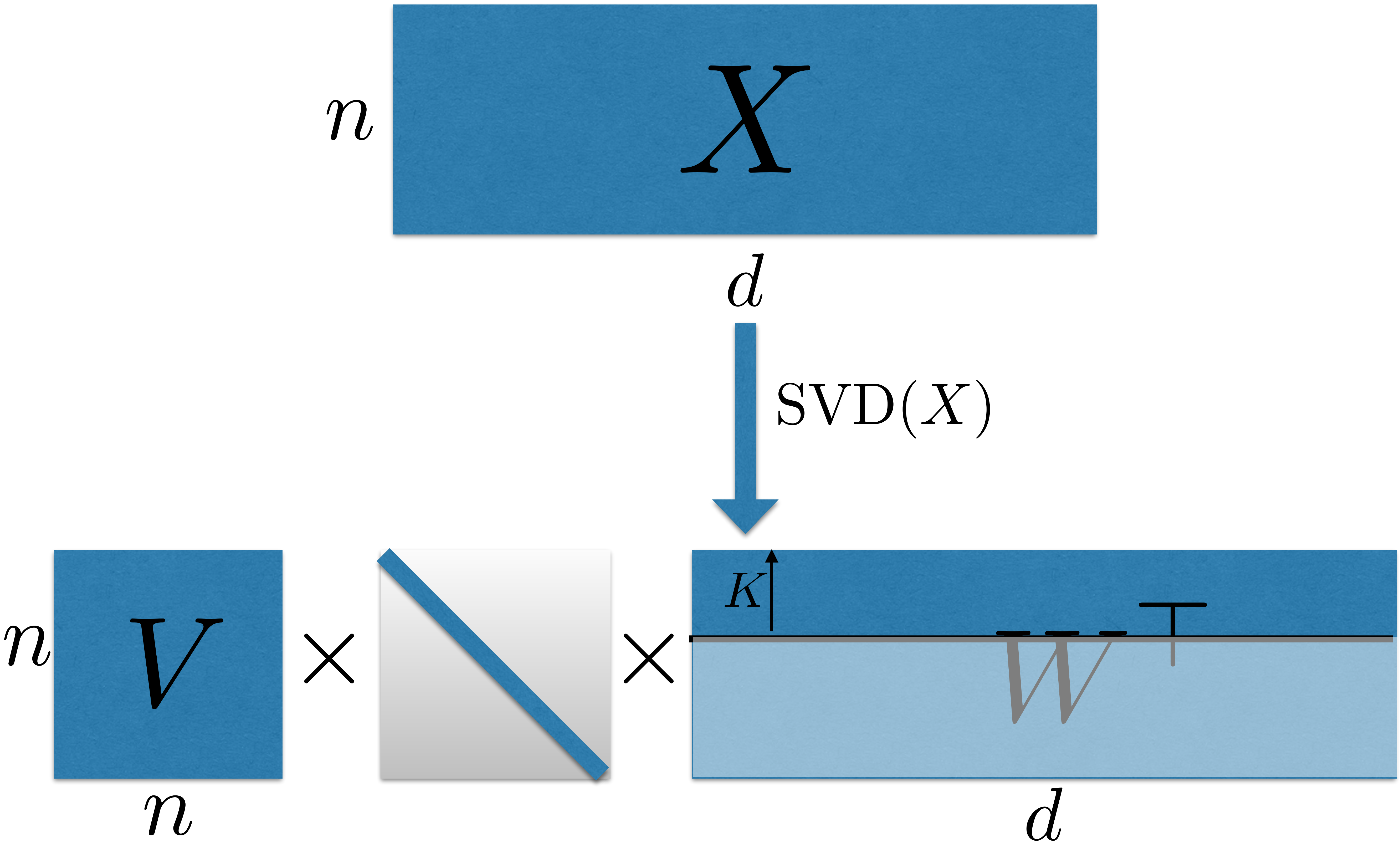
$$\underset{d \times K}{W} = \mathrm{Eigs}(\Sigma, K)$$

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



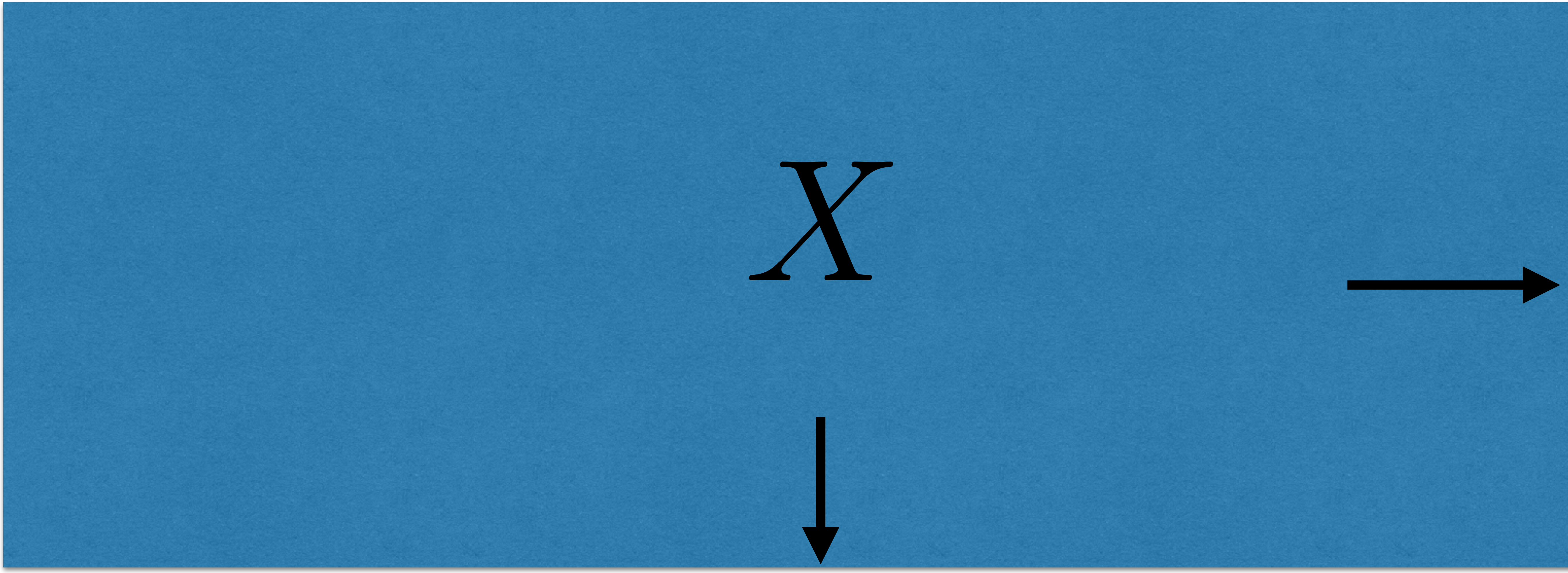
THE TALL, **the Fat** AND THE UGLY





n
X
d
SVD(X)
n
V
n
×
×
K
W⊤
d

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

Pick a Random W

$$Y = X \times \left.\left[\begin{array}{ccc} +1 & \dots & -1 \\ -1 & \dots & +1 \\ +1 & \dots & -1 \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ +1 & \dots & -1 \end{array}\right]_{K}^{d} \middle/ \sqrt{K} \right.$$

Why should Random Projections even work?!

Random Projection

- What does “it works” even mean?

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.

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?!

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

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

Why should Random Projections even work?!

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

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

WHY SHOULD RANDOM PROJECTIONS EVEN WORK?!

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

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

$$= \sum_{i,i'} \left(W[i,j] \cdot W[i',j]\right) \cdot \left(\tilde{\mathbf{x}}[i] \cdot \tilde{\mathbf{x}}[i']\right)$$

Why should Random Projections even work?!

Hence,

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

WHY SHOULD RANDOM PROJECTIONS EVEN WORK?!

Hence,

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

if $i \neq i'$, $W[i,j]$ and $W[i',j]$ are independent and so

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

Why should Random Projections even work?!

Hence,

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

if $i \neq i'$, $W[i,j]$ and $W[i',j]$ are independent and so

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

$$= \sum_{i=1}^{d} \tilde{\mathbf{x}}[i]^2/\sqrt{K}^2 = \|\tilde{\mathbf{x}}\|_2^2/K$$

Why should Random Projections even work?!

Hence,

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

Why should Random Projections even work?!

Hence,

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

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

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

WHY SHOULD RANDOM PROJECTIONS EVEN WORK?!

Hence,

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

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

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

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

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

Why should Random Projections even work?!

Hence,

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

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

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

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

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

Lets try this in Matlab …

Why should Random Projections even work?!

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

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 $i, j \in \{1, \ldots, n\}$,

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

Why should Random Projections even work?!

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 $i, j \in \{1, \ldots, n\}$,

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

Lets try on Matlab . . .

Why should Random Projections even work?!

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

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

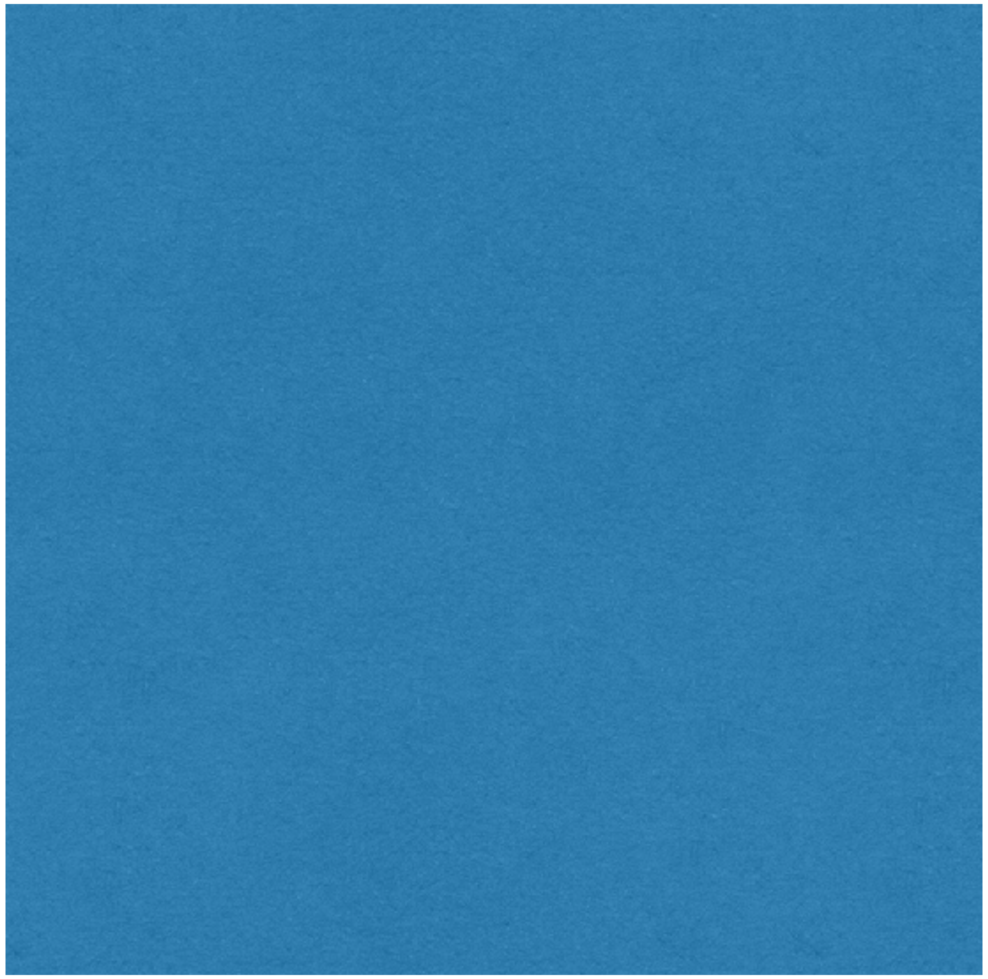
$$(1 - \epsilon)\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2 \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|_2 \leq (1 + \epsilon)\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2$$

Lets try on Matlab …

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

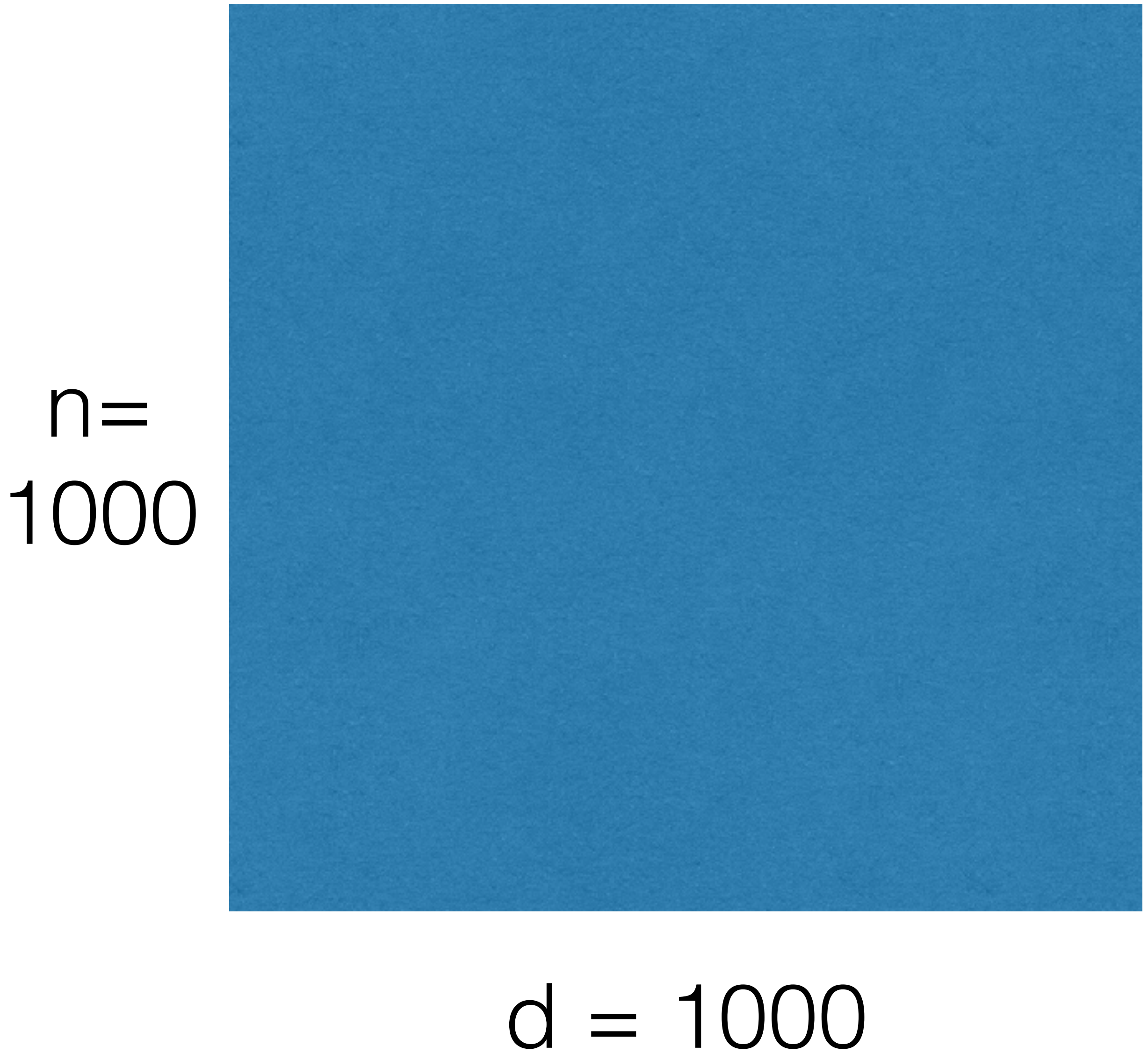
Why is this so Ridiculously Magical?

n=
1000



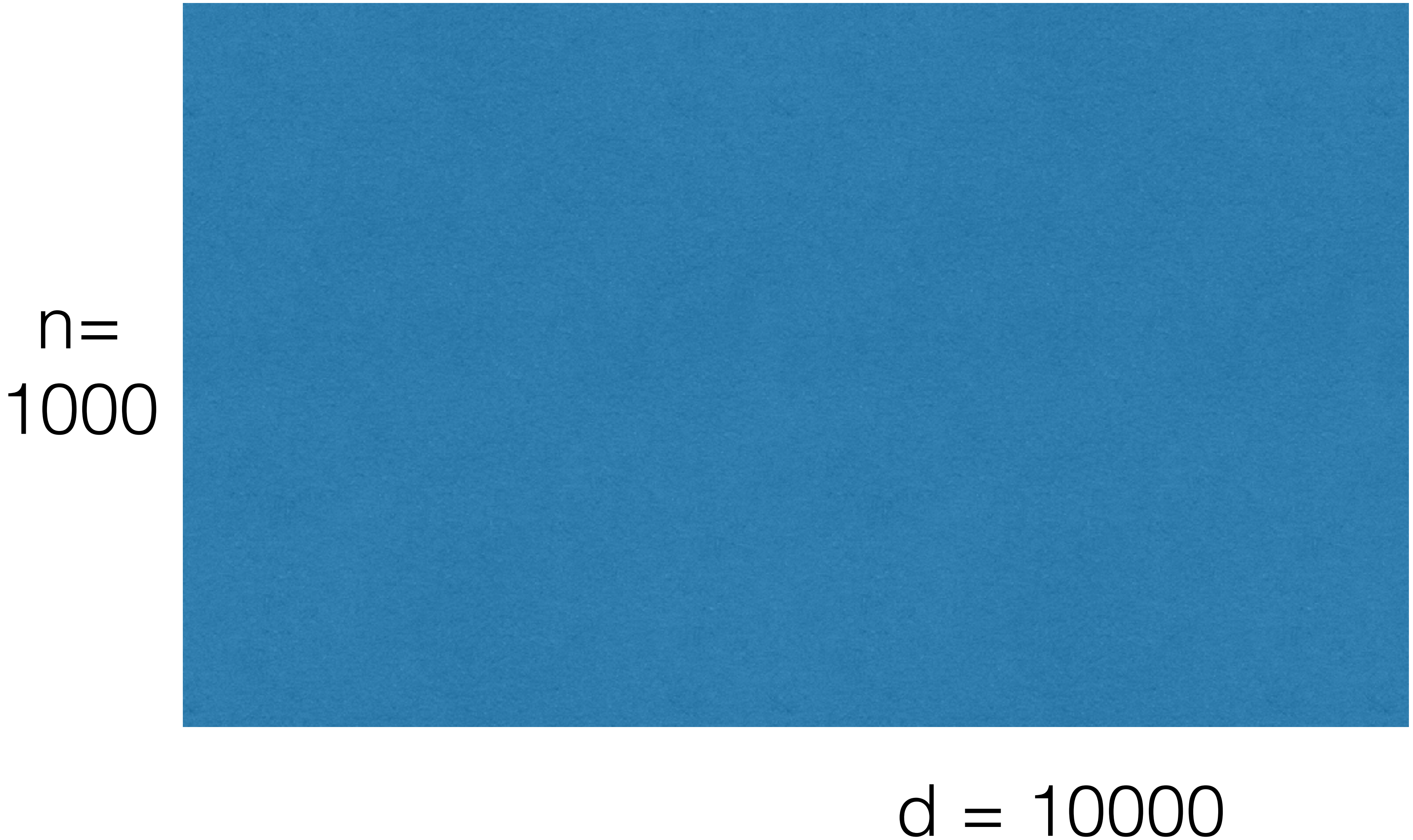
d = 1000

Why is this so Ridiculously Magical?



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

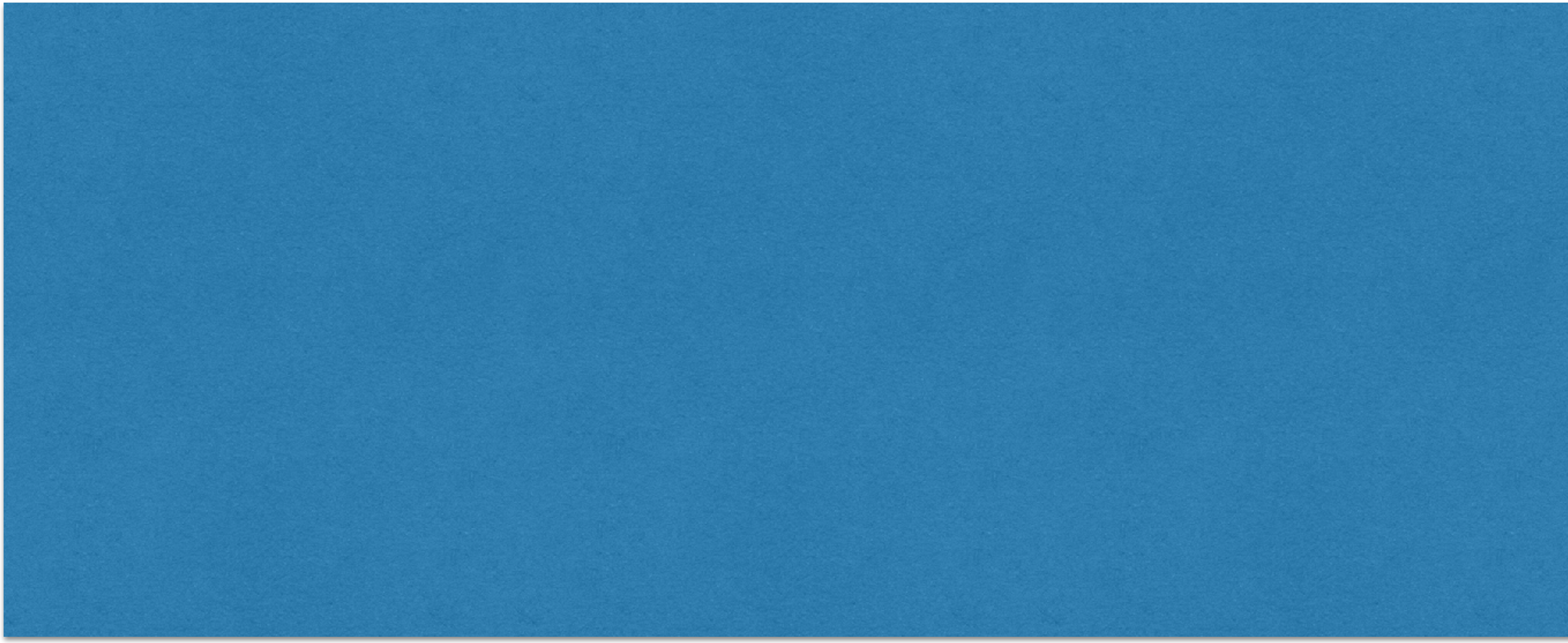
Why is this so Ridiculously Magical?



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