Business Data Science.

Oct. 27. Lecture no 18.

today: Unsupervised Learning.

There are no Labeled examples (no Y col).

→ Sometimes we choose of the columns as a target Y. and see how well the other features can predict Y.

→ Typical unsupervised Tasks:

- → Finding hidden structure in data
- → Clustering data. (K Means, TSNE).
- → Dimensionality Reduction. (PCA)

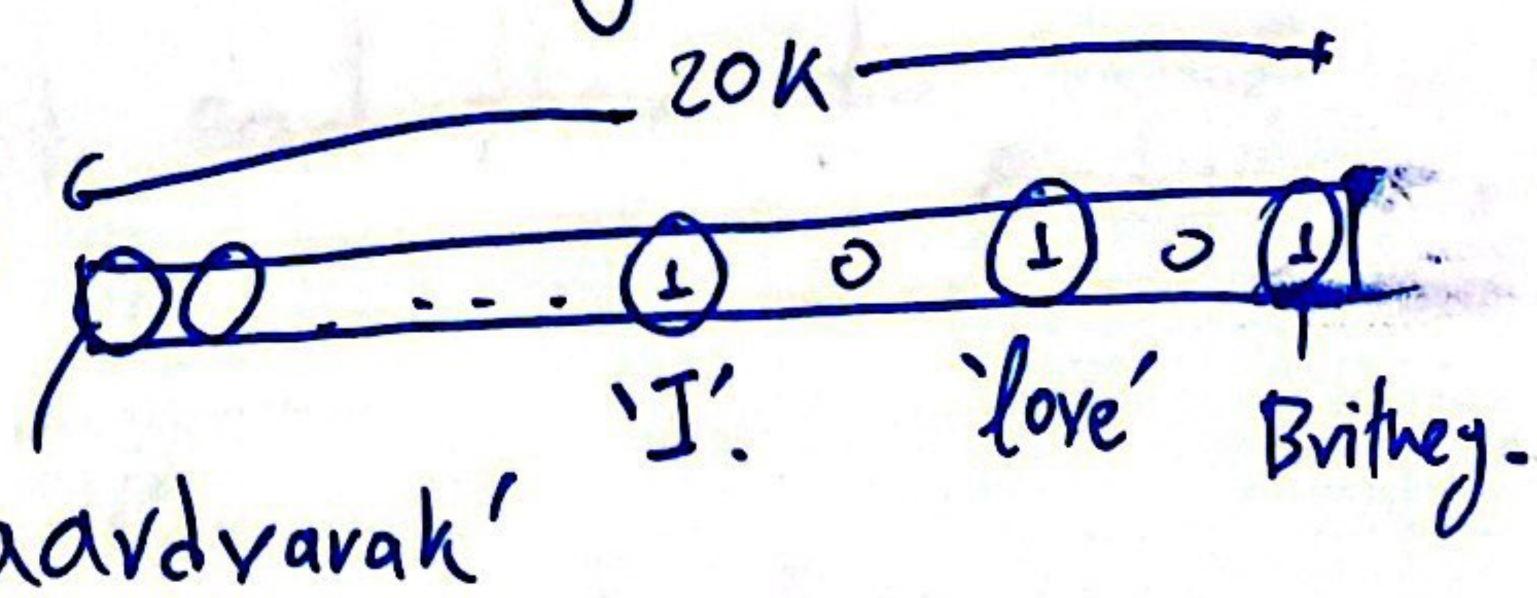
Both useful for visualization, understanding Roles of features.

→ Generative Modeling.

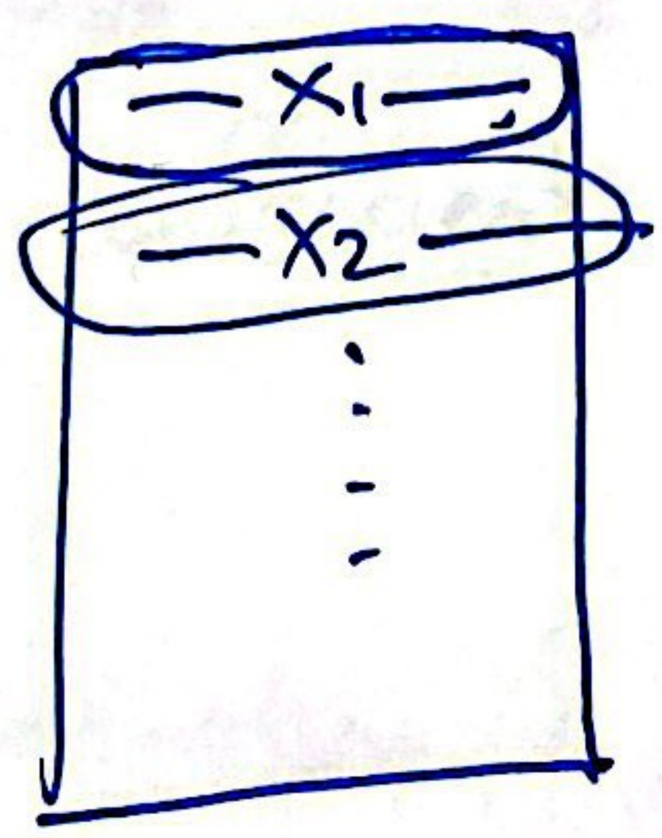
Example text data: Twitter. Bag of Words'

data → Vectors

"I love Britney." ⟷



X data Matrix

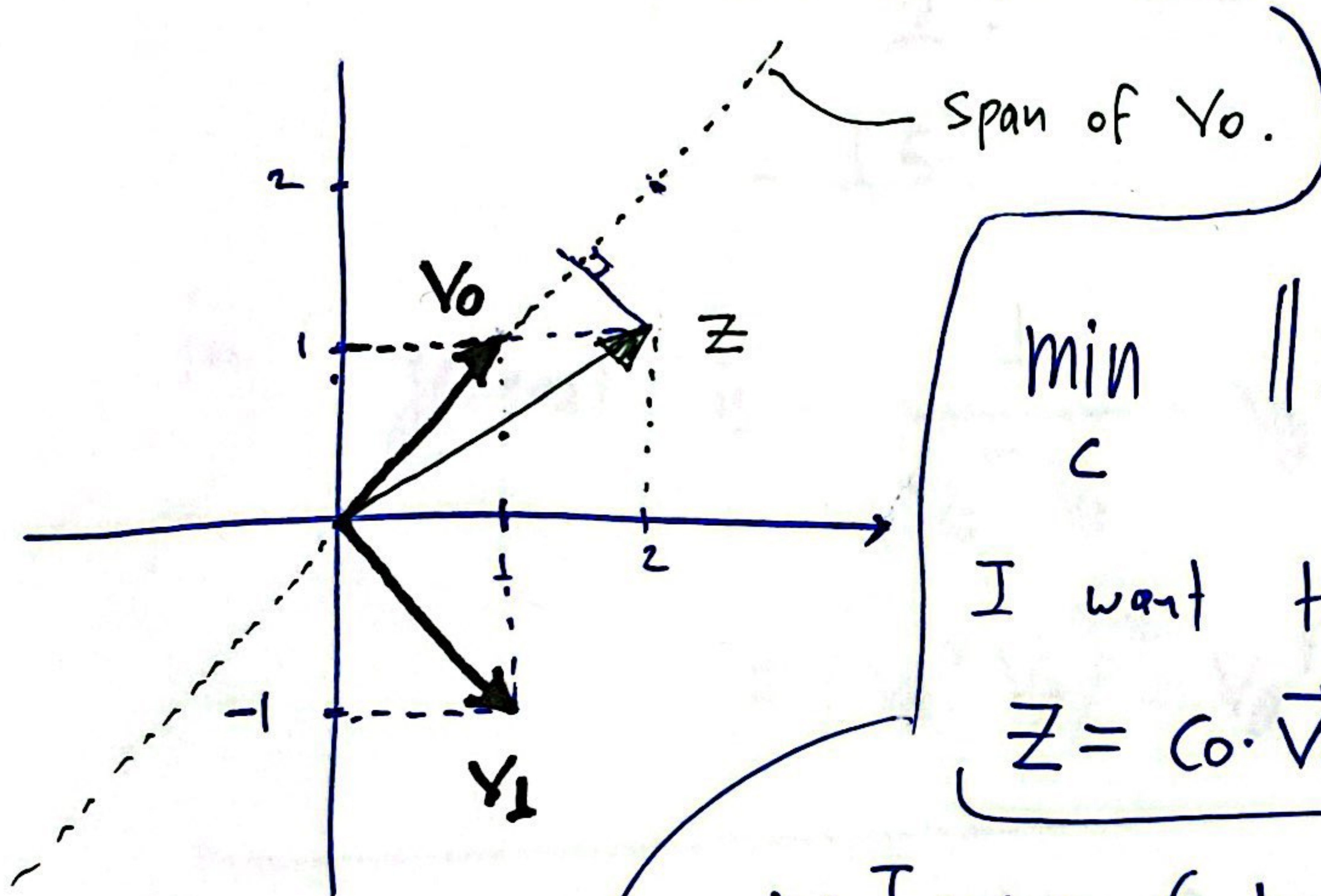


Vectors Projecting and Approximating.

$V_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. $V_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $Z = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Can you approximate Z as closely as possible by the span of V_0.

all vectors of the form $c \cdot V_0$. for any $c \in \mathbb{R}$



$$\min_{c} \|\vec{Z} - c \cdot \vec{V_0}\|^2.$$

I want to write

$$Z = c_0 \cdot \vec{V_0} + c_1 \cdot \vec{V_1}$$

$$V_0^T \cdot V_1 = (1 \ \ 1) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 - 1 = 0.$$

$V_0 \perp V_1$.

To project a vector Z on another vector V_0, normalize V_0 first and then do inner product.

$$V_0^N = \frac{V_0}{\|V_0\|_2} = \frac{V_0}{\sqrt{1^2+1^2}} = \frac{V_0}{\sqrt{2}}.$$

optimal coefficient $(V_0^N)^T \cdot Z = \frac{1}{\sqrt{2}} \cdot (1 \ \ 1) \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

$$= \frac{1}{\sqrt{2}} \cdot 3.$$

best approximation $\left((V_0^N)^T \cdot Z\right) \cdot V_0^N$.

3. $\underbrace{\left(V_0^{N^T} \cdot z\right)}_{\frac{1}{\sqrt{2}} \cdot 3} \cdot \underset{\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}}{\underset{\shortparallel}{V_0^N}} = \frac{3}{2} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$

hence $\beta^* = \frac{3}{2}.$ and the best approximation
$= 1{,}5$

in general is $\frac{1}{V_0^T \cdot V_0} \cdot V_0^T \cdot Z = \frac{1}{\|V_0\|_2^2} \cdot V_0^T \cdot Z.$

$= \left(V_0^T V_0\right)^{-1} \cdot V_0^T \cdot Z.$

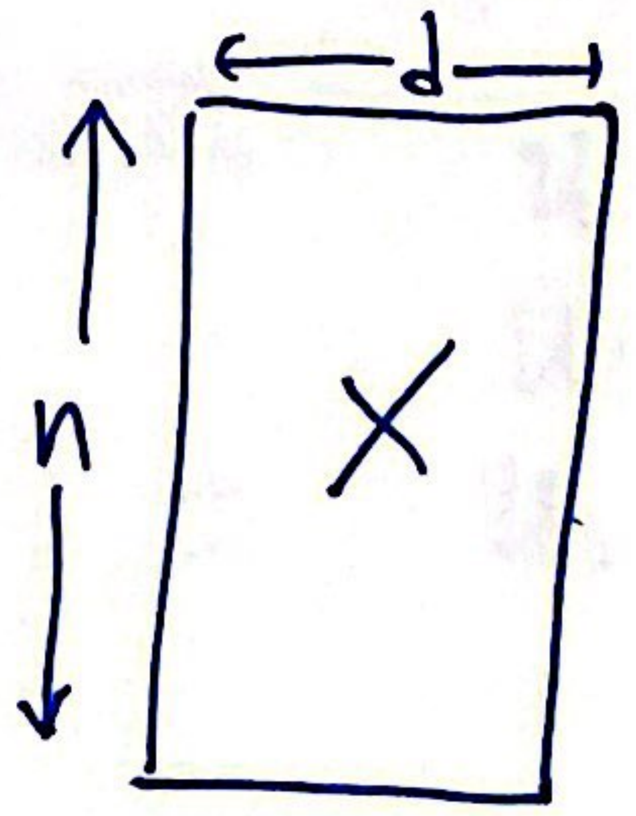
Example: I have 10 seconds of Bach Music sampled at 44KHz. $= X_1$, $X_2 = 10$ sec of Britney song. at 44kHz.

What is the <u>angle</u> of X_1, X_2.

$X_1^T \cdot X_2 = \|X_1\|_2 \cdot \|X_2\|_2 \cdot \cos\vartheta_{12}$.

---

④ Given a Matrix X:



The singular value decomposition (SVD) of X is a factorization $X = U \cdot D \cdot V^T$.

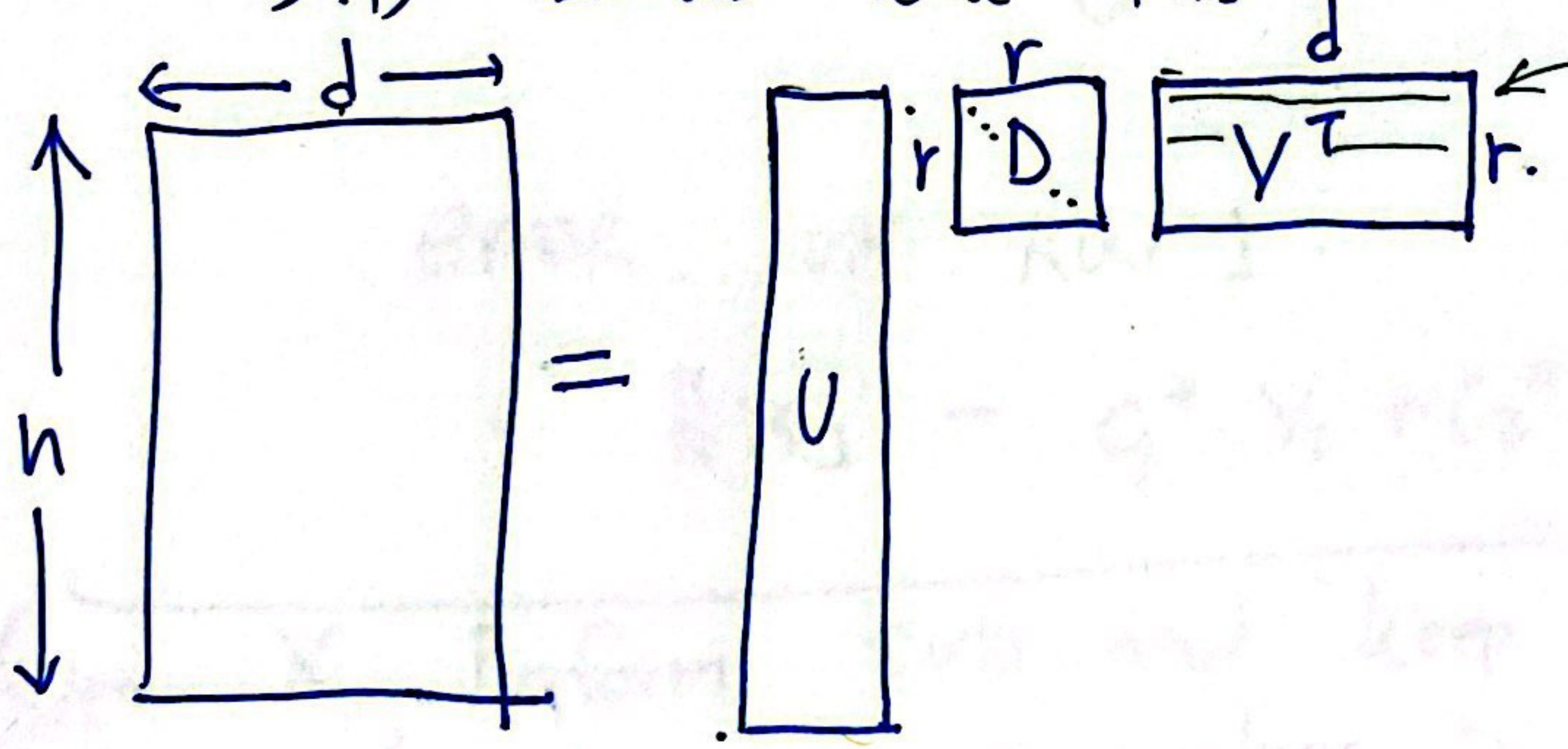
Where the cols of U, V are orthonormal.

$\left(\begin{matrix} U^T U = I. \\ V^T V = I. \end{matrix} \right)$ and D is diagonal. with nonnegative entries.

$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_R \\ | & | & & | \end{bmatrix}$ $u_i^T \cdot u_j = 0$ if $i \neq j$ orthogonal

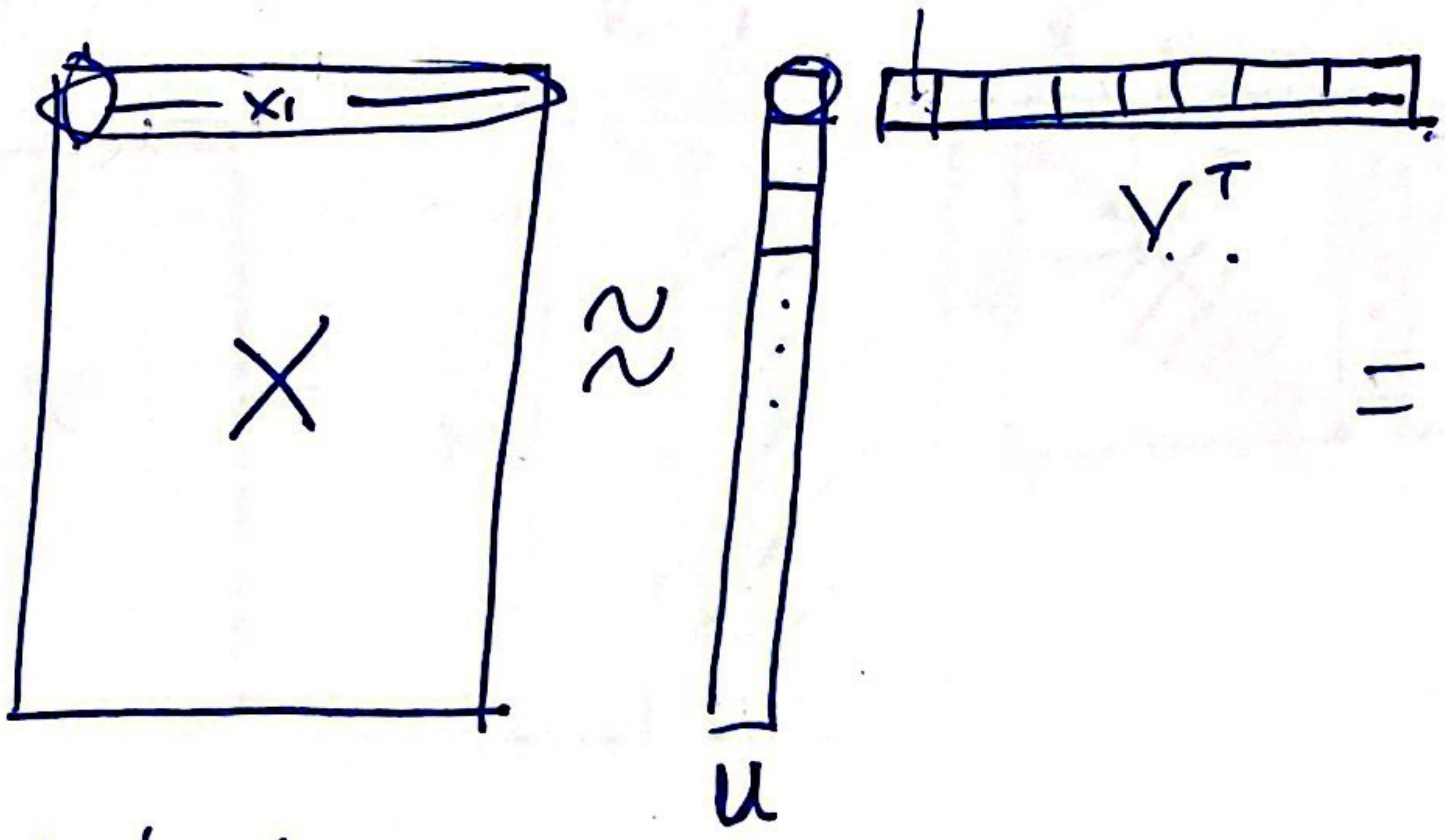
$u_i^T \cdot u_i = 1$ (normal / unit length.) .

SVD Looks like this:



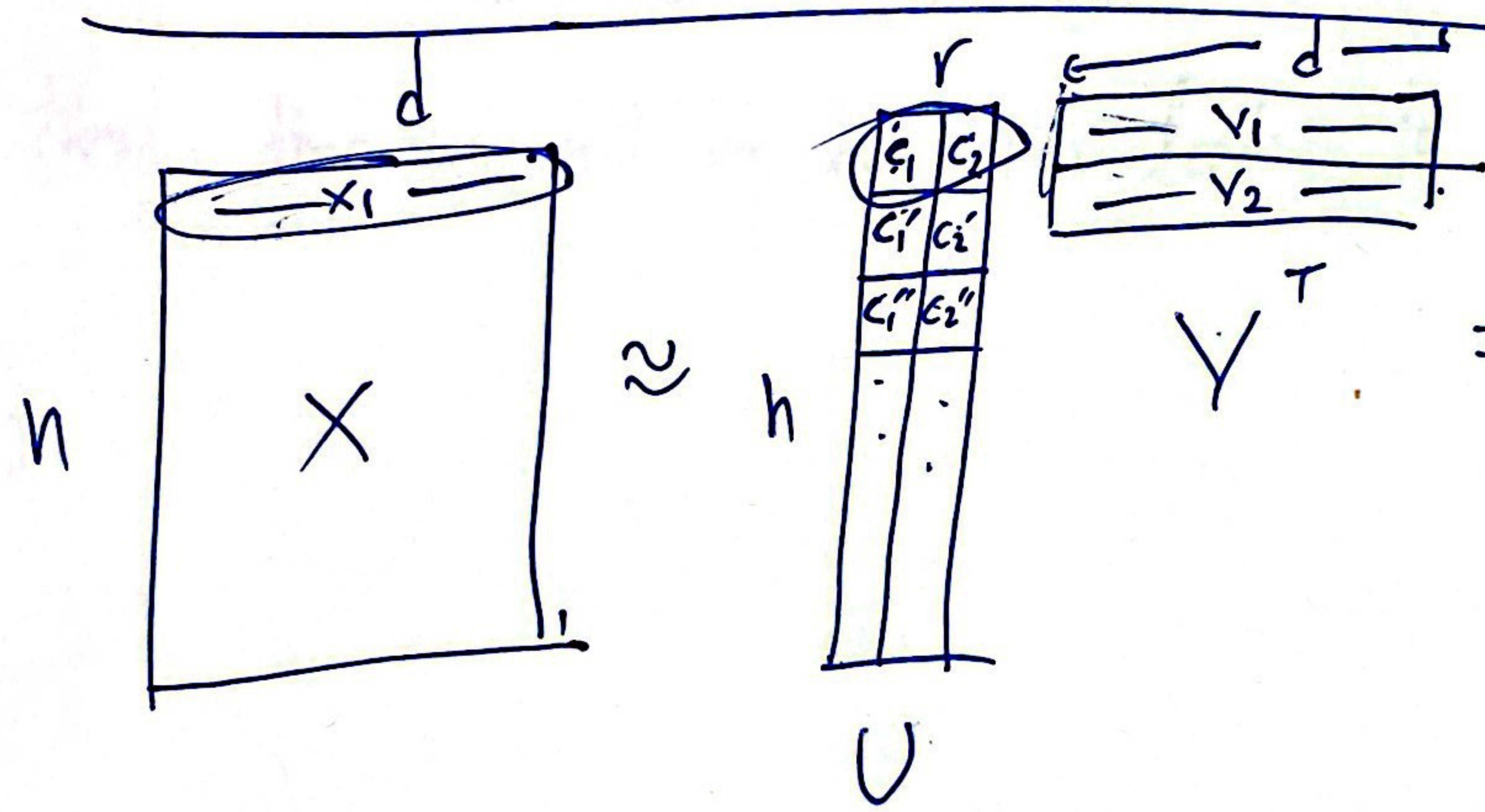
These Rows are special directions in the dataset.

(5)



$$= \begin{array}{l} u_1 \cdot V^T \\ u_2 \cdot V^T \\ u_3 \cdot V^T \end{array}$$

rank 1 approximation example

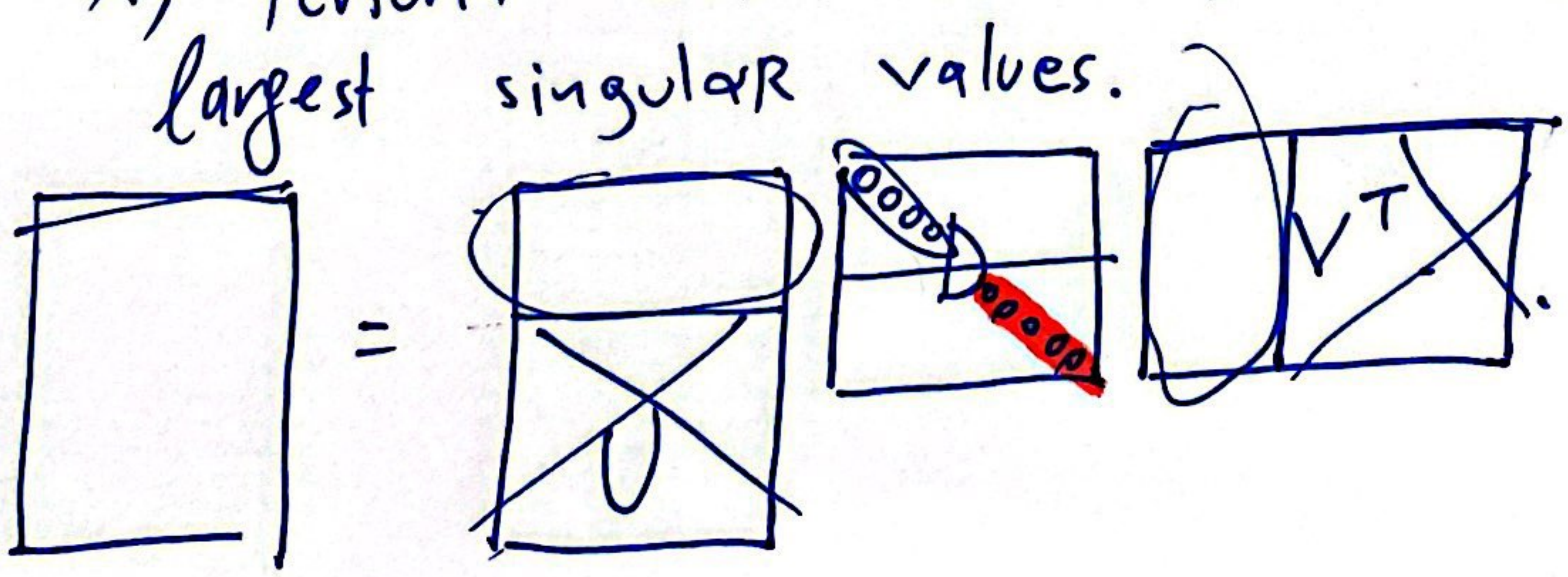


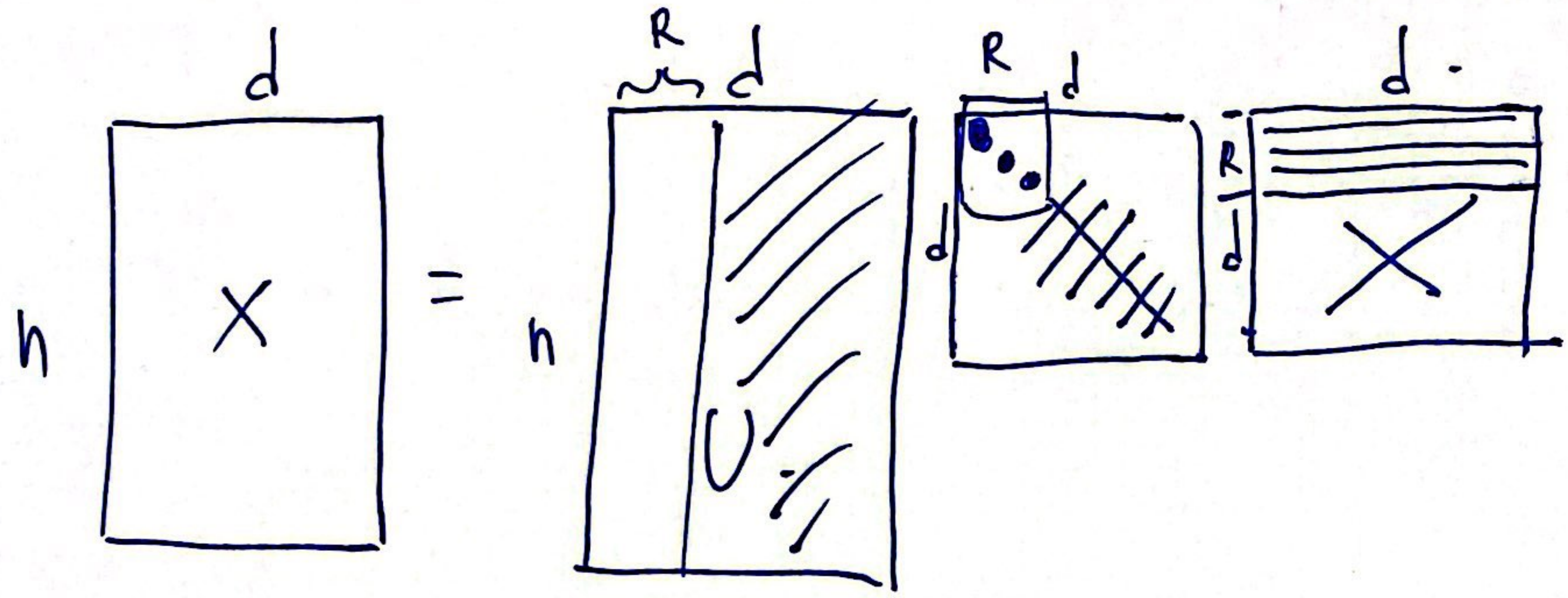
$$= \left[\begin{array}{l} c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2 \\ \hline \\ \end{array} \right.$$

error on row 1.

$$\| x_1 - c_1^* \cdot v_1 + c_2^* v_2 \|^2 .$$

Given X, perform SVD and keep the top r largest singular values.





· truncated SVD keeps the top R singular values and corresponding singular vectors.

Next time we will see how this leads to PCA.