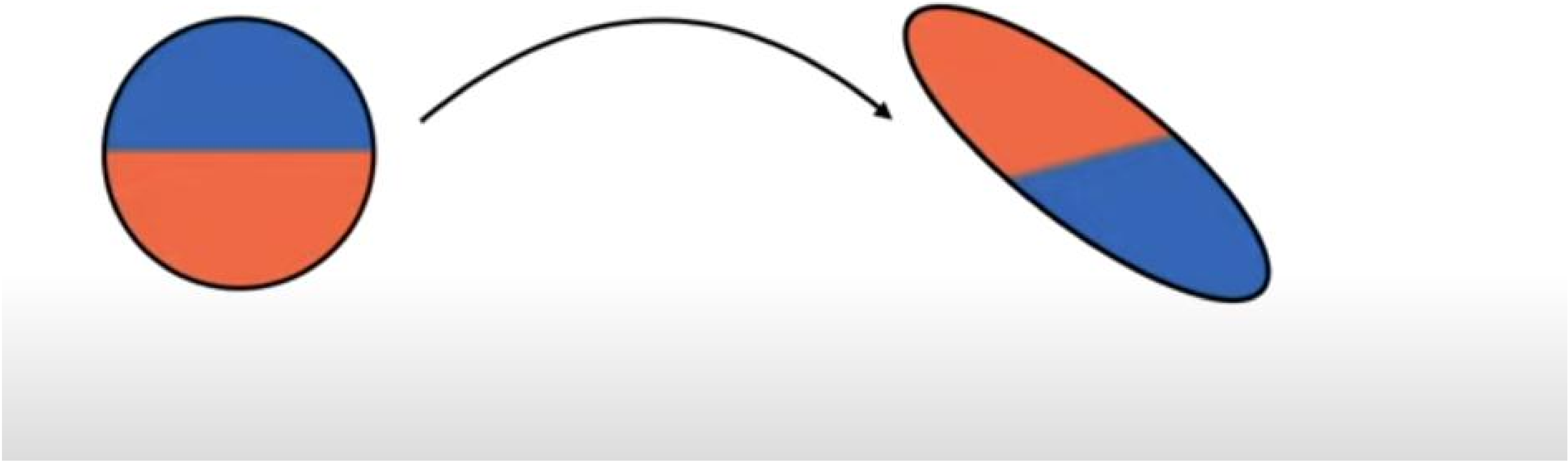
Puzzle



01

Giới thiệu

Thành viên nhóm

- 20520720 - Nguyễn Đỗ Quang
- 20520684 - Nguyễn Hải Nam
- 19521727 - Từ Trương Tuấn Kiệt
- Bùi Hạ Khánh (Không hoàn thành nhiệm vụ)

→

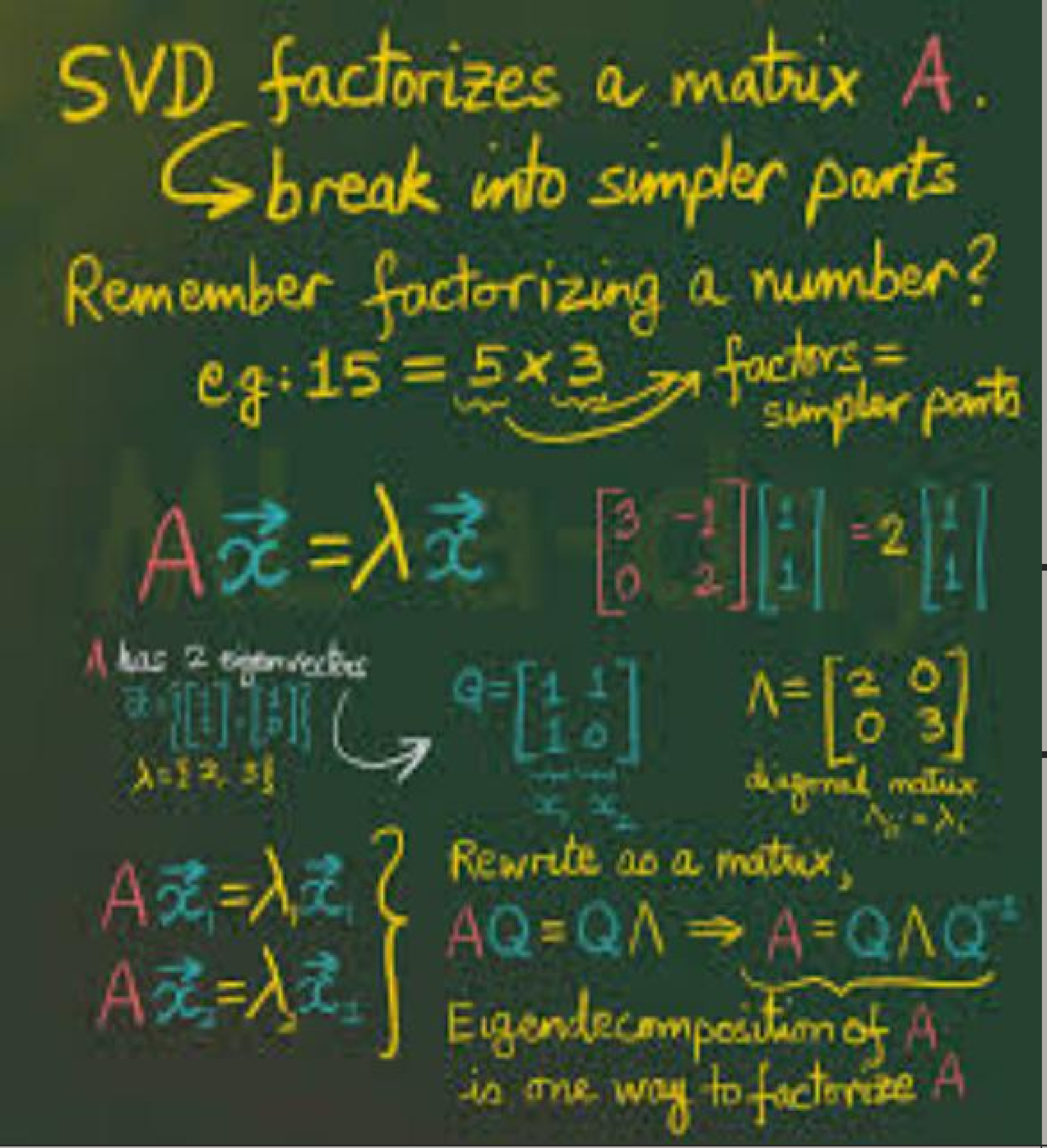
01

SVD - Singular value Decomposition

Today's Agenda

- Introduction
- Có thể bạn đã biết
- What is SVD ?
- Computing SVD

→



INTRODUCTION

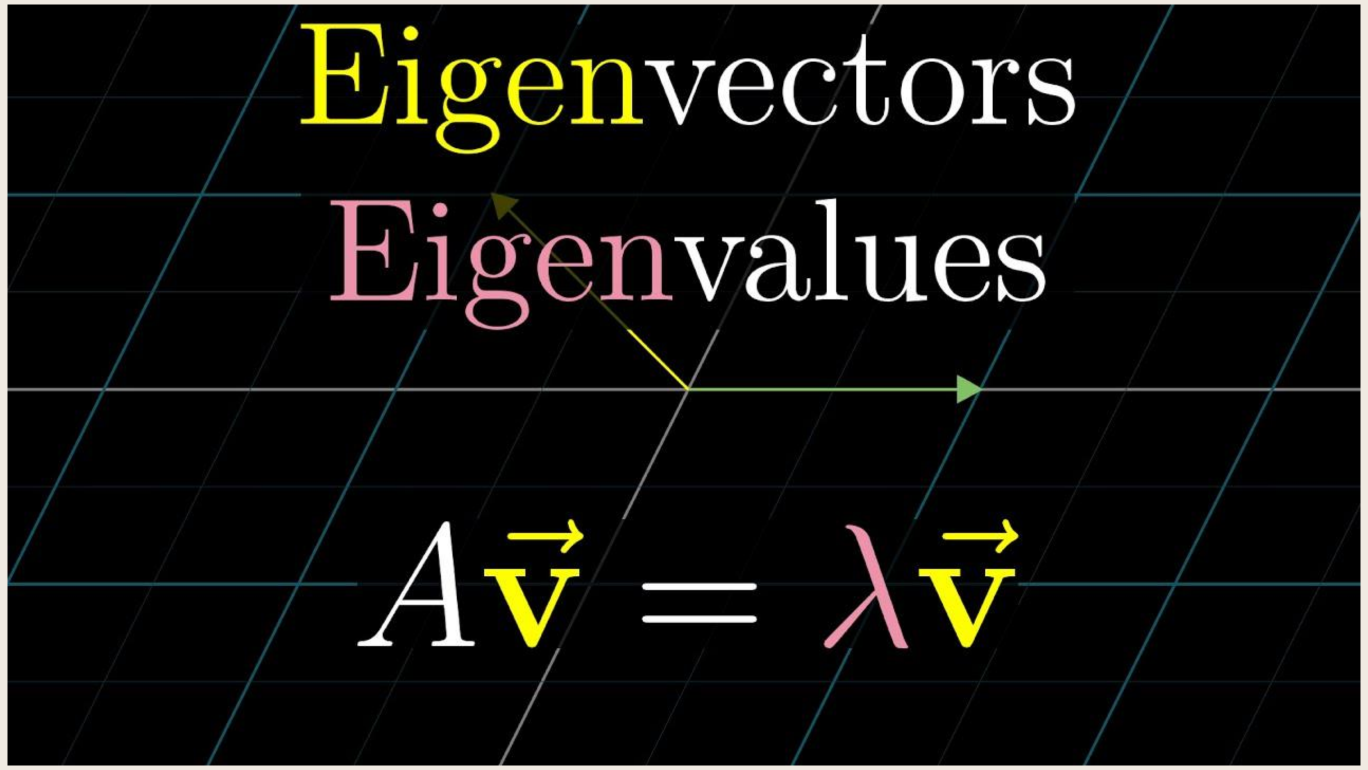
Professor Gilbert Strang calls ‘absolutely a high point of linear algebra.‘

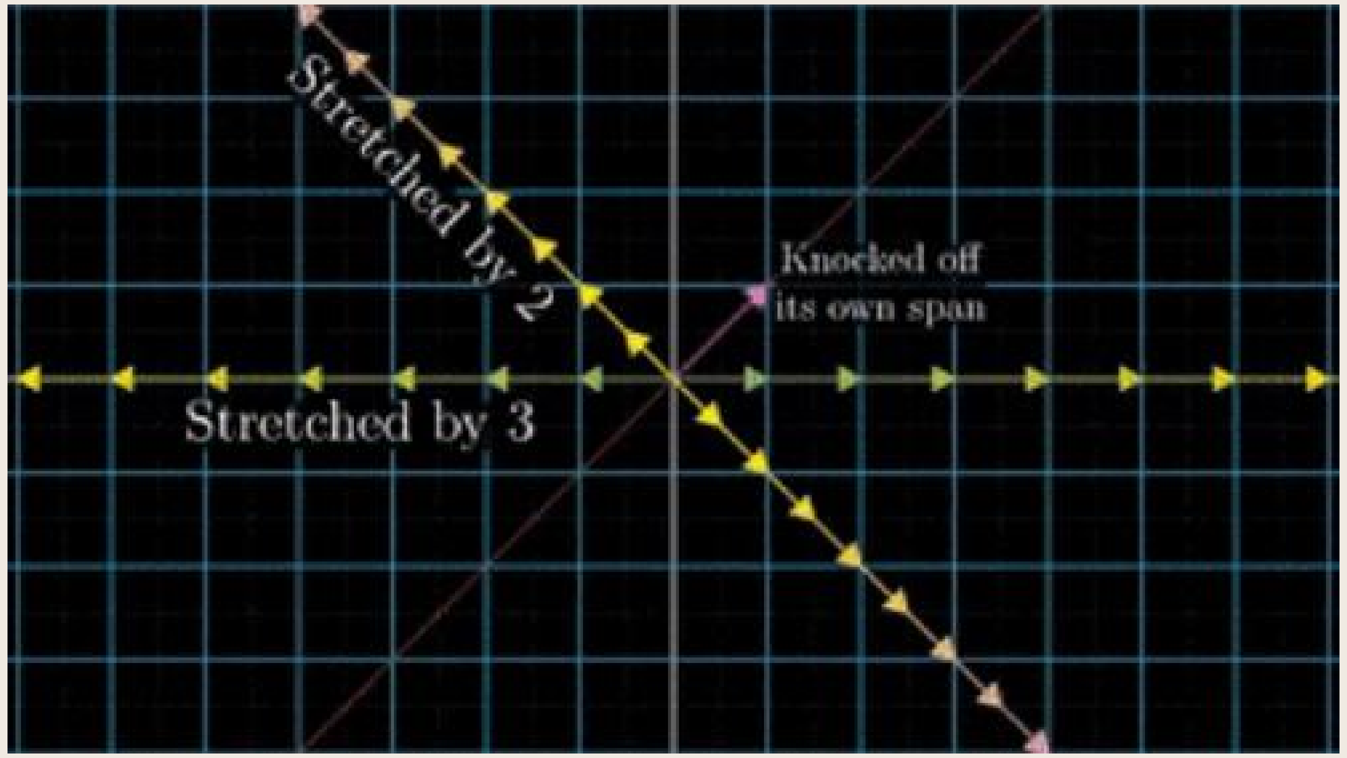
Phương pháp phân tích suy biến (singular value decomposition) được viết tắt là SVD là một trong những phương pháp thuộc nhóm matrix factorization. Phương pháp SVD đã được phát triển dựa trên những tính chất của ma trận trực giao và ma trận đường chéo để tìm ra một ma trận xấp xỉ với ma trận gốc.

Có thể bạn đã biết

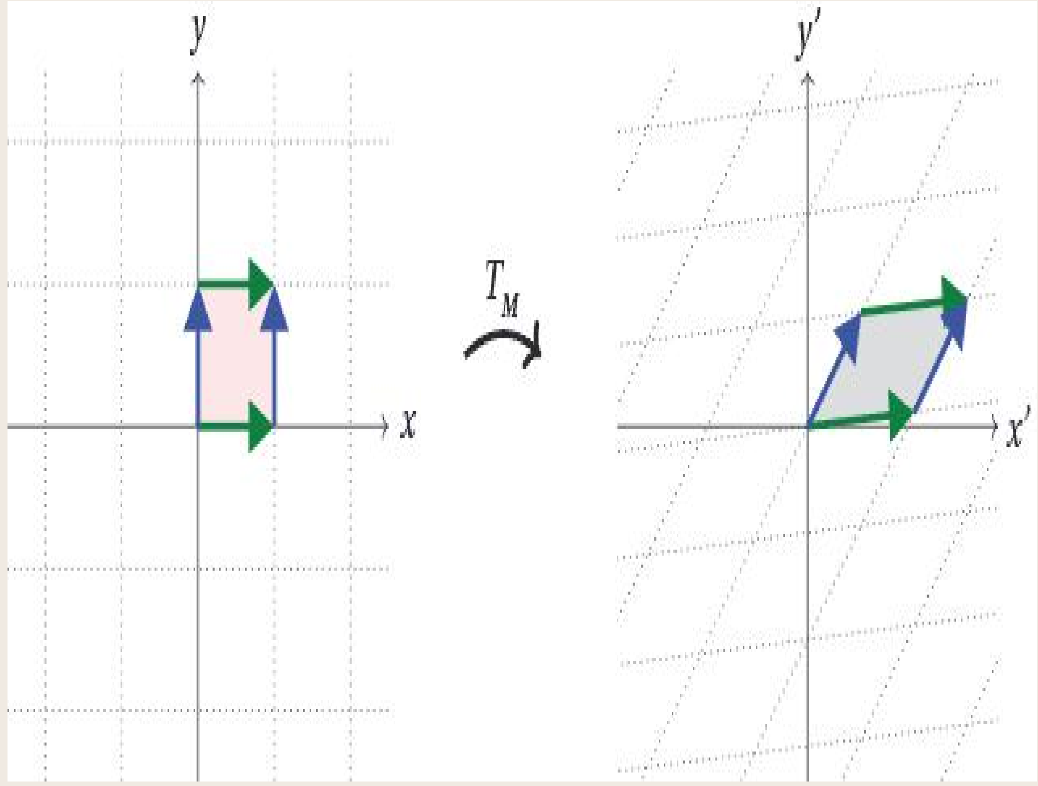
Orthogonal matrix	Orthonormal system	Diagonal matrix	Symmetric matrix

Eigenvector and eigenvalue





Có thể bạn đã biết

Linear transformation	rotation	scaling
		

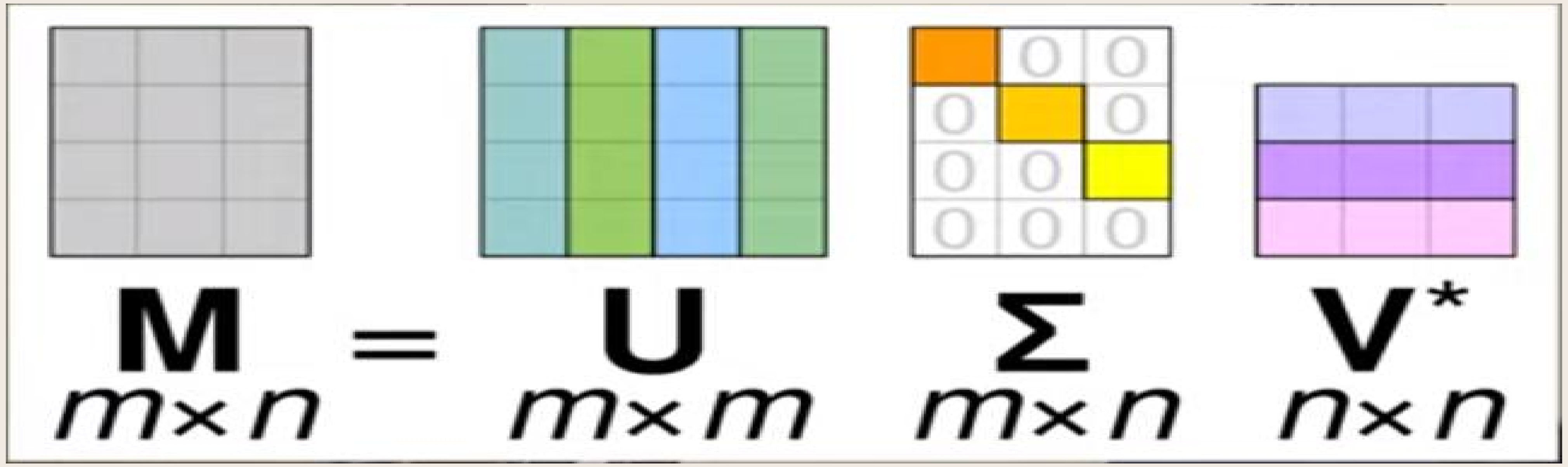
WHAT IS SVD - Singular value Decomposition ?

Singular value Decomposition

$$A = U\Sigma V^T$$

U, V^T: Ma trận trực giao(orthogonal matrix)

Σ: là ma trận đường chéo(diagonal Matrix)



EXAMPLE

Singular Values

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

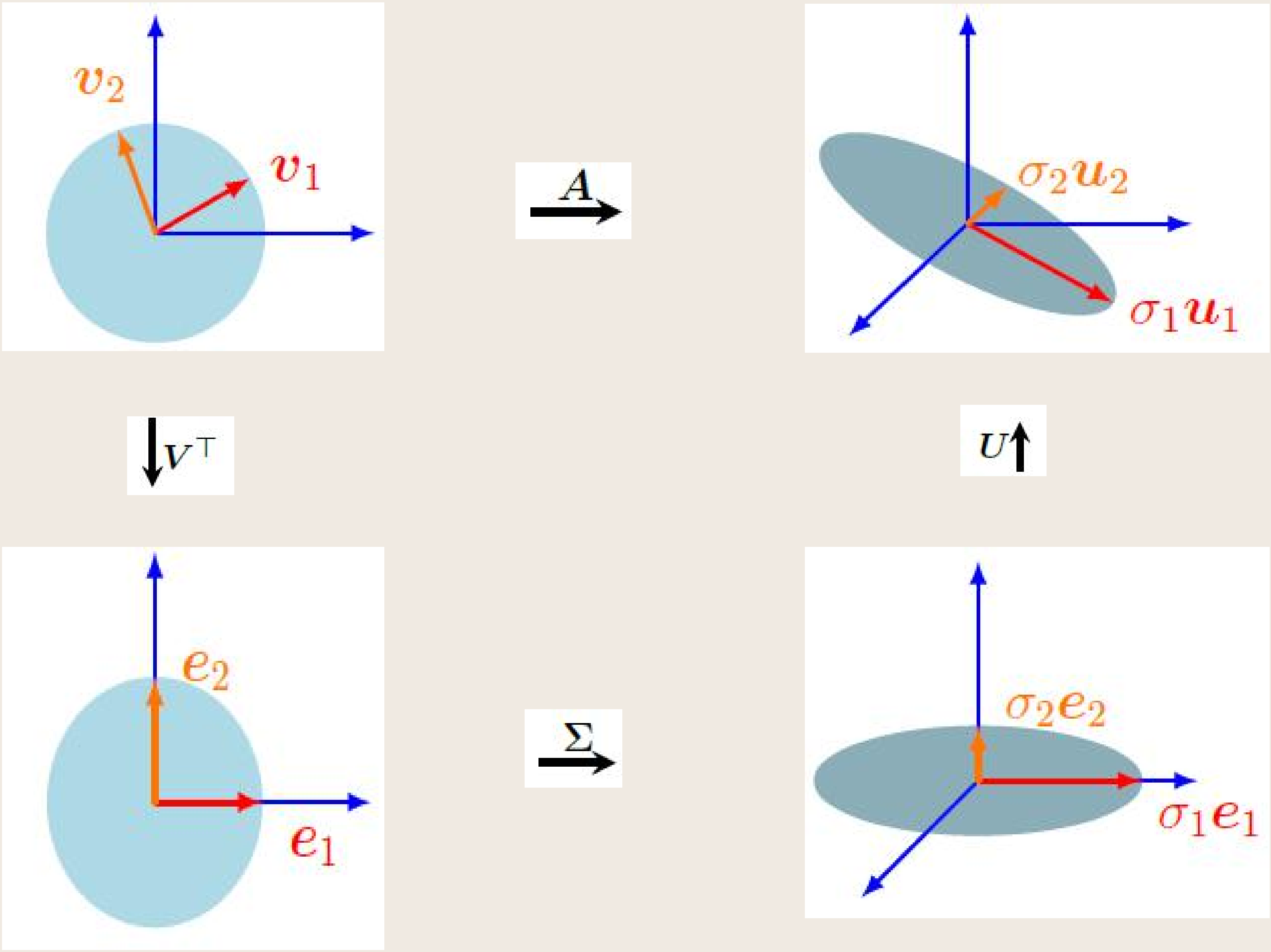
$$\begin{bmatrix} 0.85 & 0 & -0.53 \\ 0 & 1 & 0 \\ -0.53 & 0 & 0.85 \end{bmatrix} \begin{bmatrix} 1.62 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.62 \end{bmatrix} \begin{bmatrix} 0.53 & 0 & -0.85 \\ 0 & 1 & 0 \\ 0.85 & 0 & 0.53 \end{bmatrix}^T$$

Orthogonal matrix

Diagonal matrix

Orthogonal matrix

Về mặt hình học



Computing SVD

Ví dụ tính SVD của ma trận A:

$$A = \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

- Tìm ma trận chuyển vị:

$$A^T = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

- Tính $A^T A$:

$$AA^T = \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

$\longrightarrow$

Computing SVD

- Tìm các giá trị riêng và vecto riêng của AA^T

+ Giá trị riêng 8, vecto riêng: $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

+ Giá trị riêng 2, vecto riêng: $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

+ Giá trị riêng 0, vecto riêng: $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

$\longrightarrow$

Computing SVD

- Tìm căn bậc 2 của giá trị riêng khác 0

$\sigma_1 = \sqrt{8} = 2\sqrt{2}$
$\sigma_1 = \sqrt{2}$

=ı Ma trận
σ =

$$\begin{bmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Tìm U: Các cột của ma trận U là các vecto đơn vị đã được chuẩn hóa

$$U = \begin{bmatrix} \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{3} & -\frac{\sqrt{3}}{3} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

→

Computing SVD

Tìm v: $\upsilon_i = \frac{1}{\sigma_i} . A^T . u_i$

$$\upsilon_1 = \frac{1}{\sigma_1}\begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} u_1 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{6} \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{3}}{6} \end{bmatrix}$$

$$\upsilon_2 = \frac{1}{\sigma_2}\begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{3}}{3} \\ 0 \\ \frac{\sqrt{6}}{3} \end{bmatrix}$$

$\rightarrow$

Computing SVD

Tìm vecto trực giao với tất cả các vecto đã tìm được bằng cách tìm không gian rỗng (null space) của ma trận có hàng là vecto đã tìm được.

$$\begin{bmatrix} \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{3} & 0 & \frac{\sqrt{6}}{3} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ -1 \\ 1 \end{bmatrix}$$

Chuẩn hóa vecto vừa tìm được, ta được: $\upsilon_3 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{-1}{2} \\ \frac{1}{2} \end{bmatrix}$

$$\Rightarrow\ V = \{\upsilon_1, \upsilon_2, \upsilon_3\} = \begin{bmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{3}}{2} & 0 & -\frac{1}{2} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{6}}{3} & \frac{1}{2} \end{bmatrix}$$

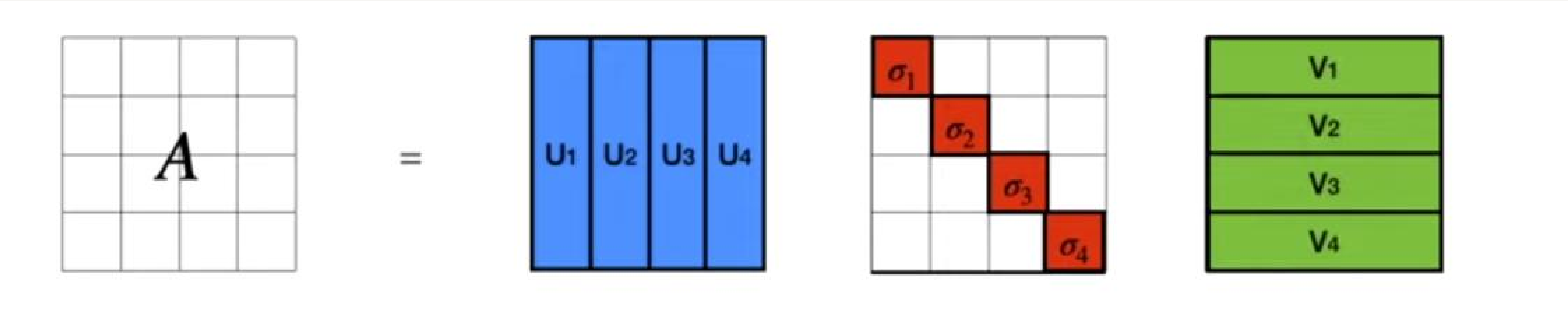
$\longrightarrow$

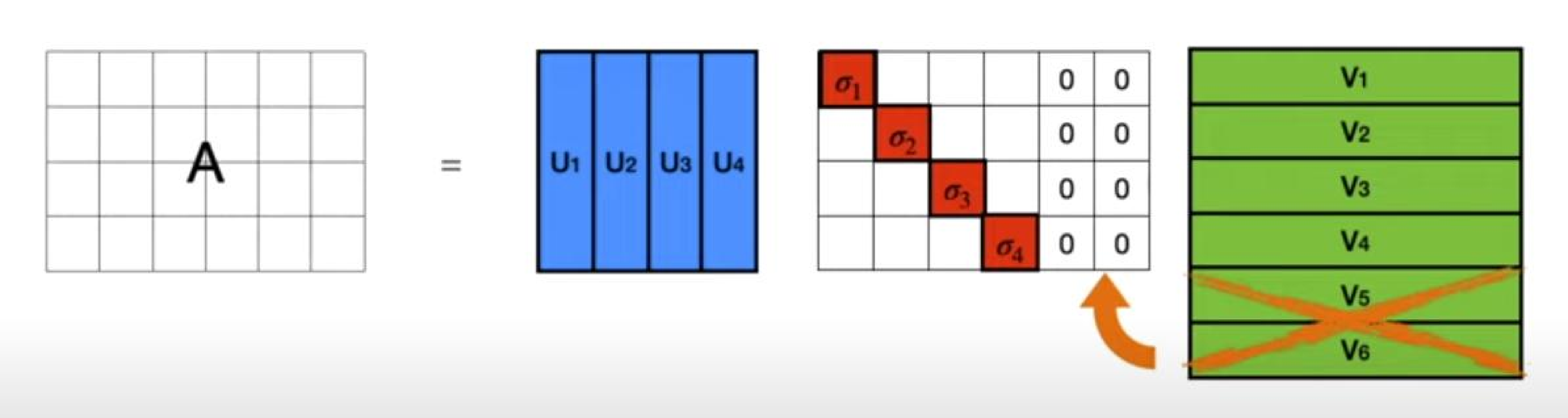
Computing SVD

$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\boldsymbol{T}}$

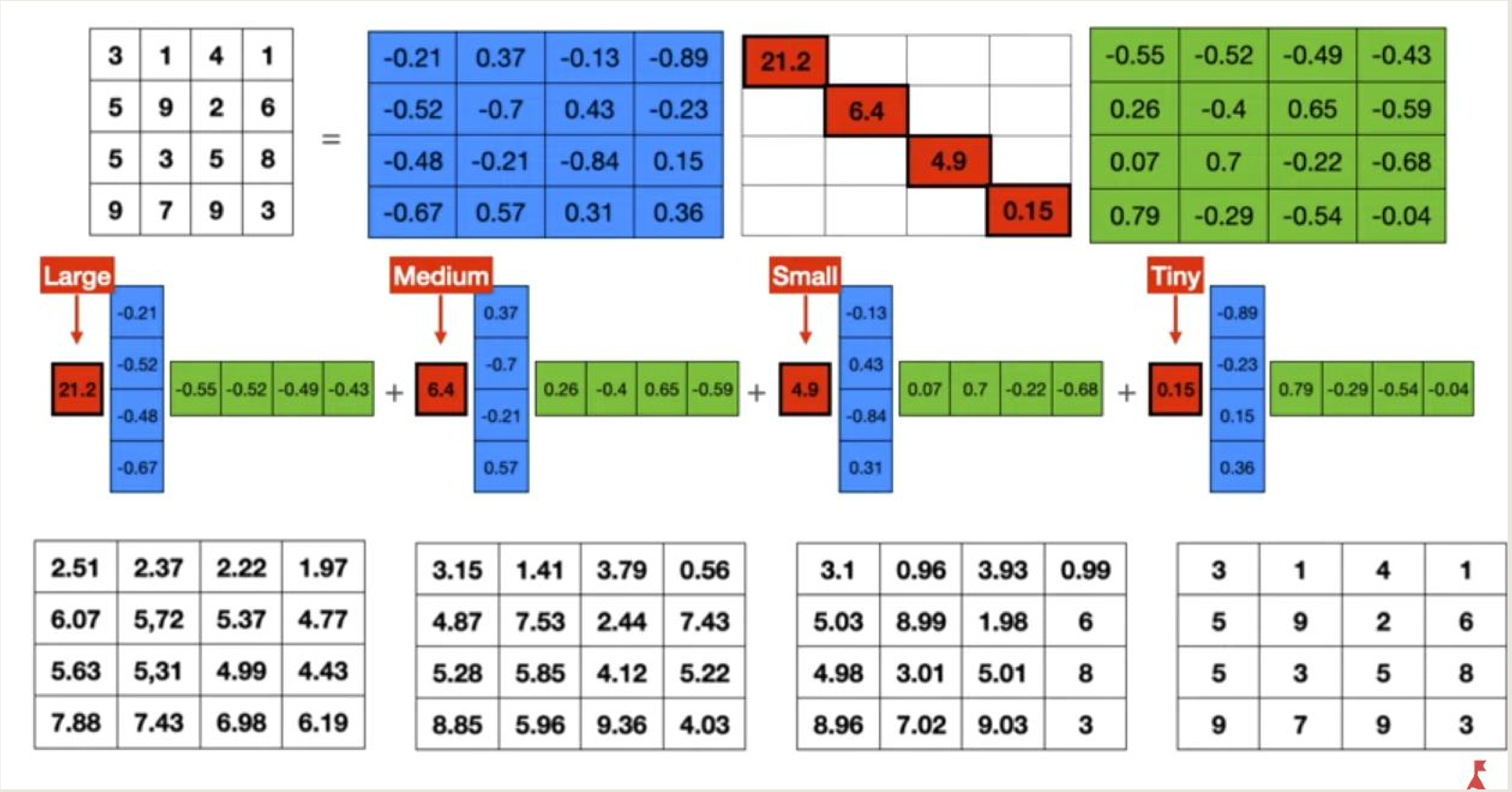
$$= \underbrace{\begin{bmatrix} \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{3} & -\frac{\sqrt{3}}{3} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} \end{bmatrix}}_{\mathrm{U}} \underbrace{\begin{bmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{3}}{2} & 0 & -\frac{1}{2} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{6}}{3} & \frac{1}{2} \end{bmatrix}^T}_{\mathrm{V}^{\mathrm{T}}}$$

SVD

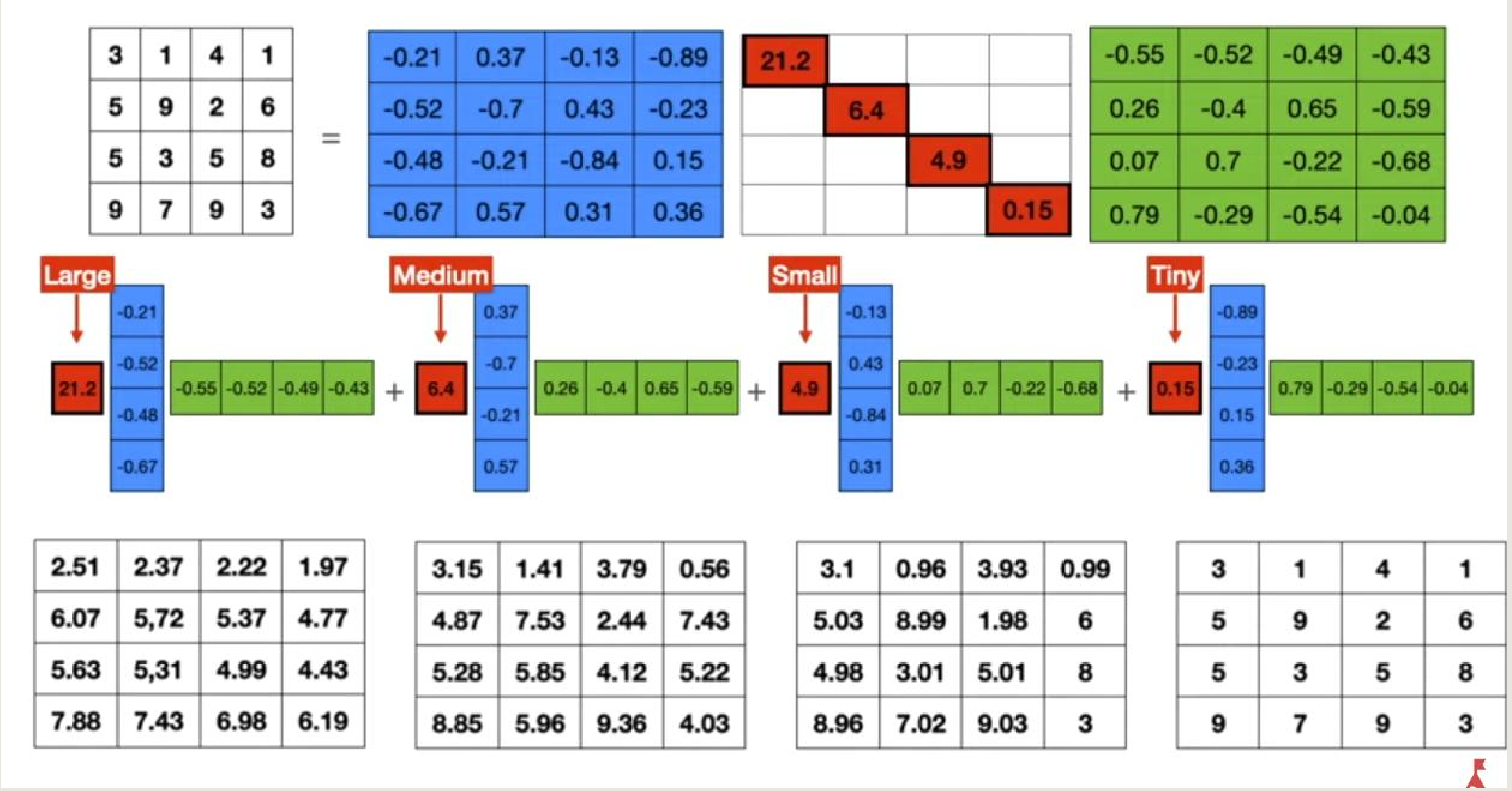

A
=
U1
U2
U3
U4
σ1
σ2
σ3
σ4
V1
V2
V3
V4


A
=
U1
U2
U3
U4
σ1
σ2
σ3
σ4
0 0
0 0
0 0
0 0
V1
V2
V3
V4
V5
V6

DECOMPOSITION

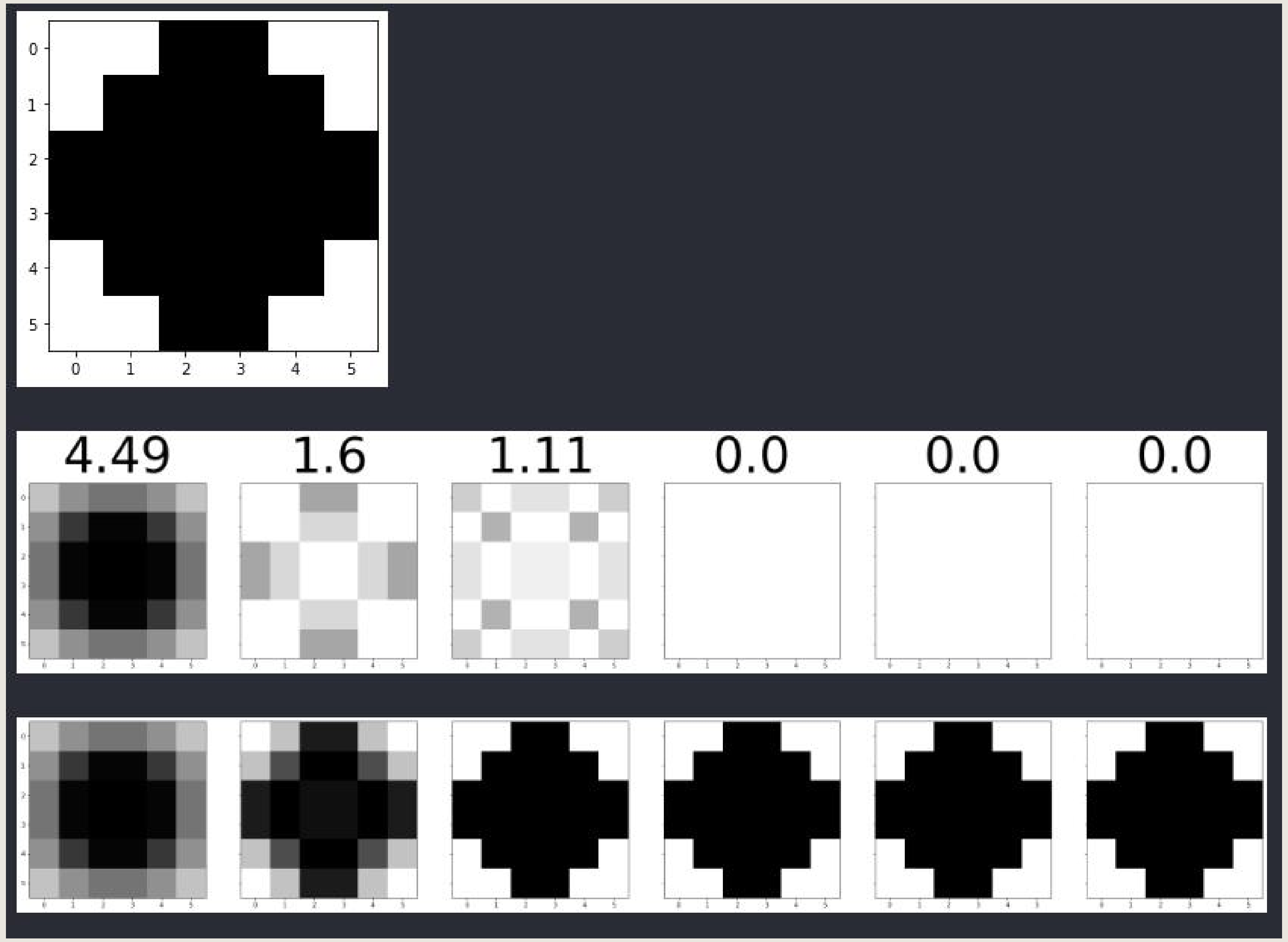


DECOMPOSITION



Trong trường hợp này giả sử mỗi ô dữ liệu mỗi ô dữ liệu là 1 byte, và nếu nếu chỉ sử dụng 2 vecto (cột thứ nhất và dòng thức nhất của ma trận U và V^T) ứng với singular value thứ nhất để lưu trữ, thì ta chỉ cần 9 bytes để lưu trữ dữ liệu thay vì 16 bytes.

DEMO IMAGE COMPRESSION



PCA – Principal components analysis

Group 1

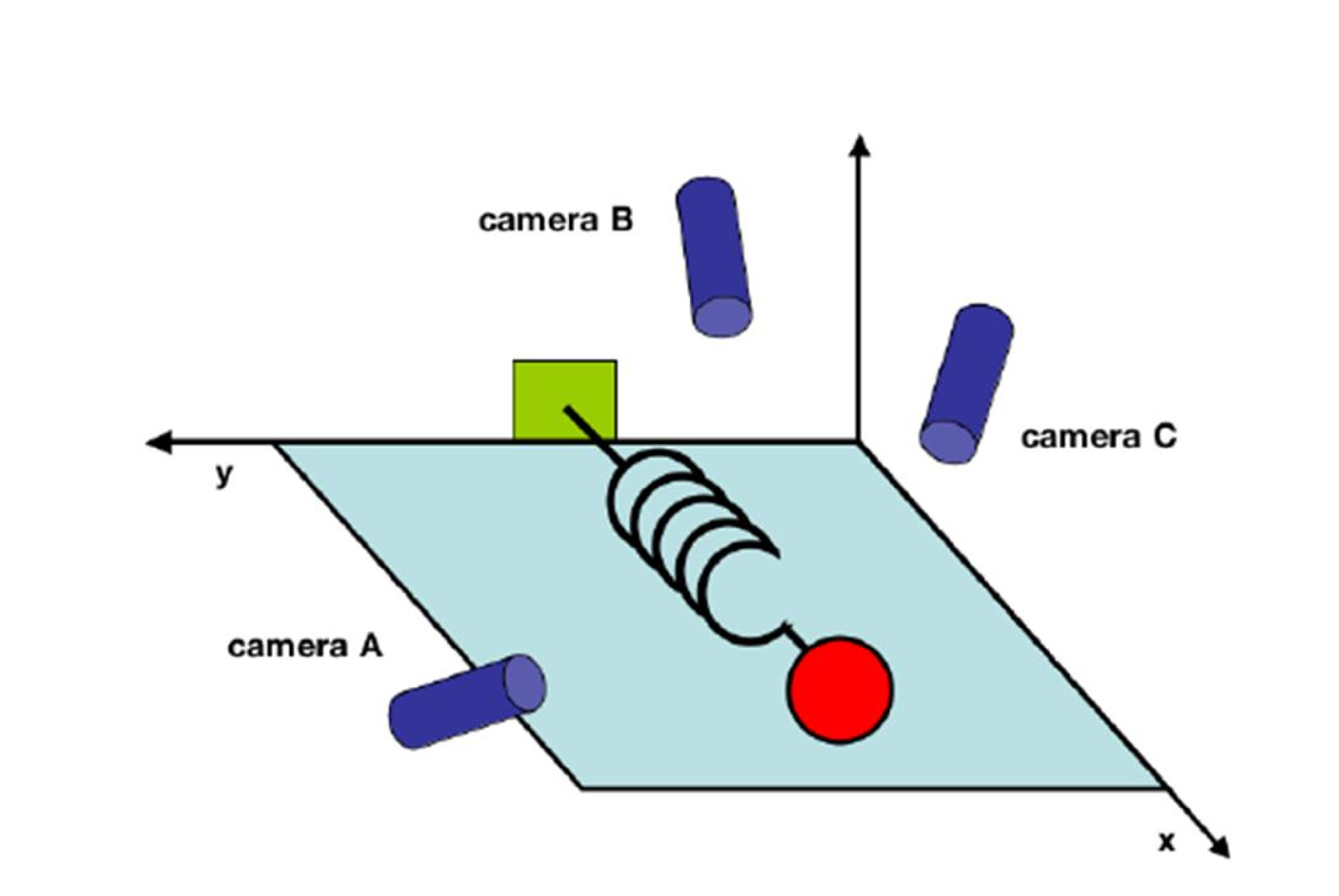
PCA – Principal components analysis

Today's Agenda

- What is Principal Component Analysis (PCA)?
- Computing PCA
- Example in PCA
- Application

→

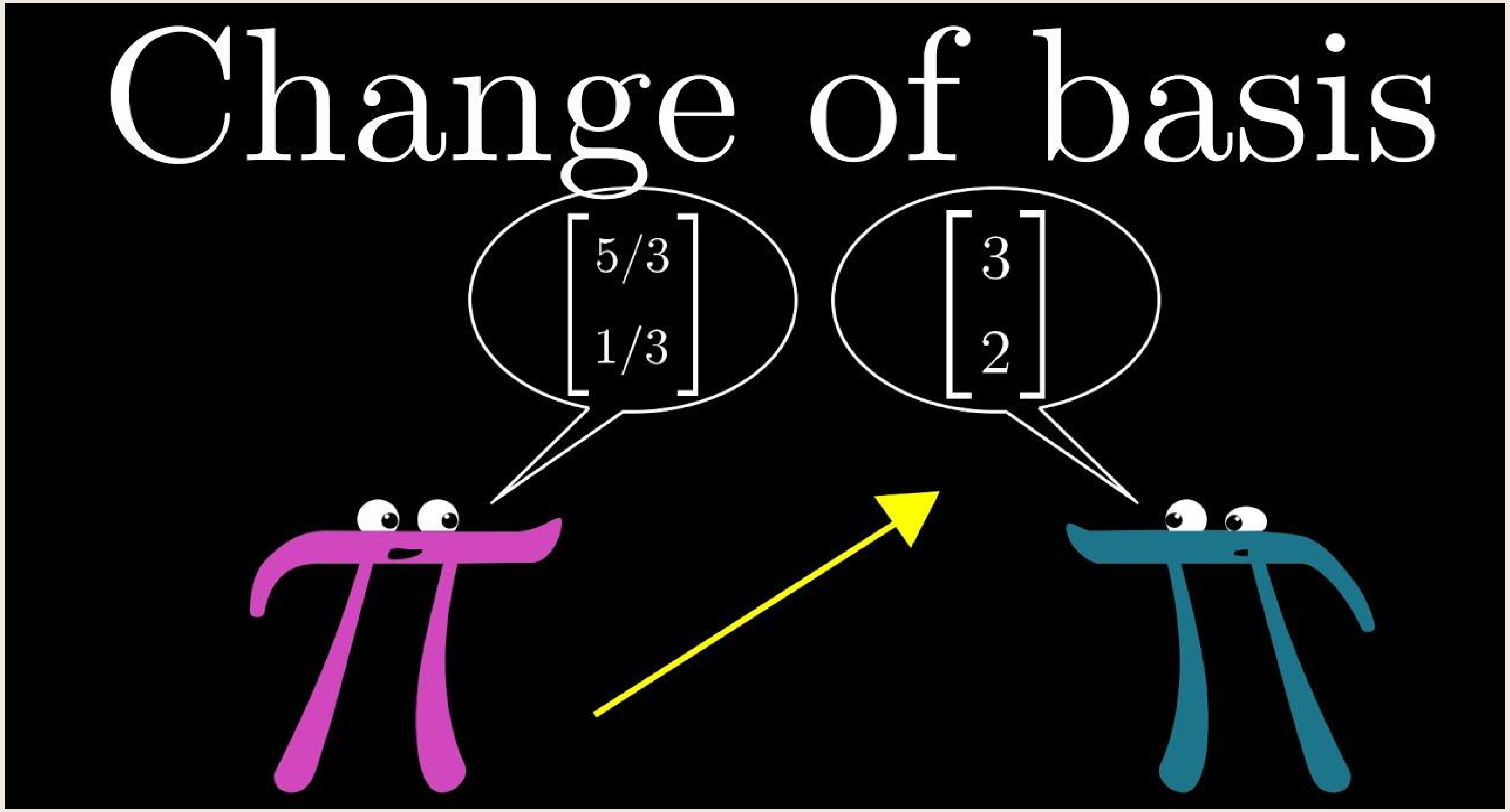
Toy example



6 dimensional column vector

$$\vec{X} = \begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \\ x_C \\ y_C \end{bmatrix}$$

Change of basis



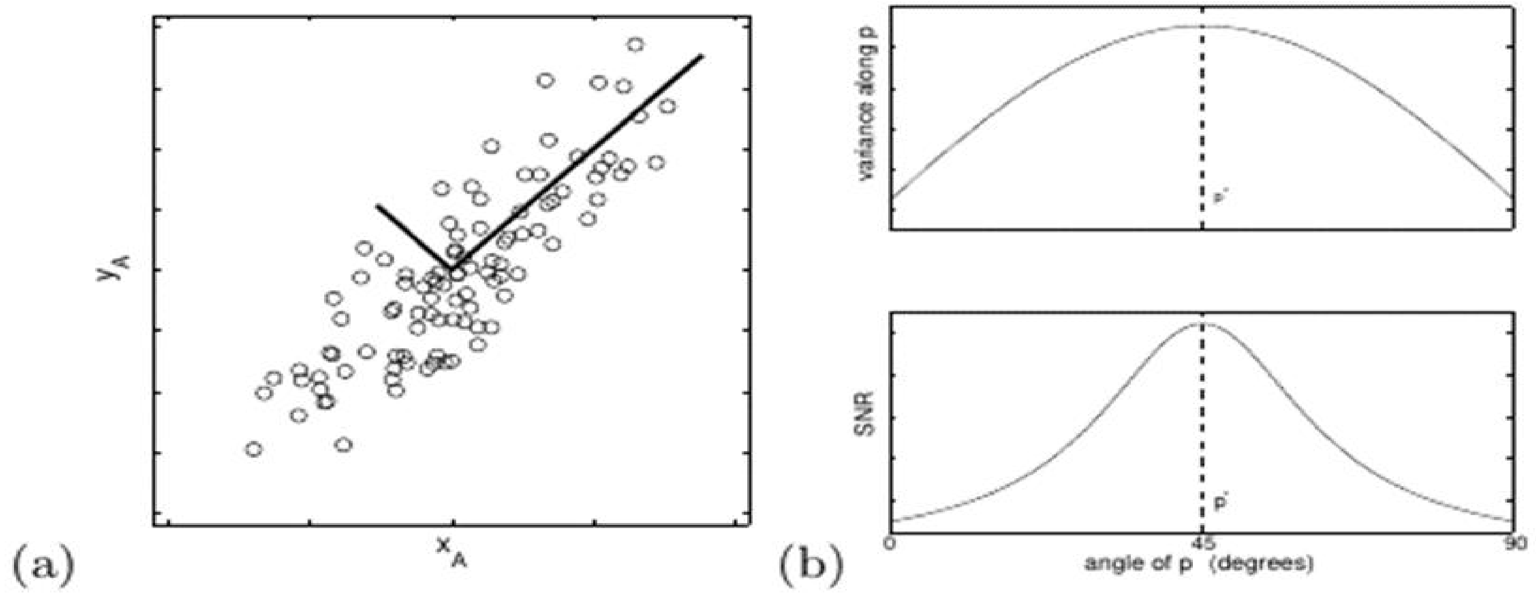
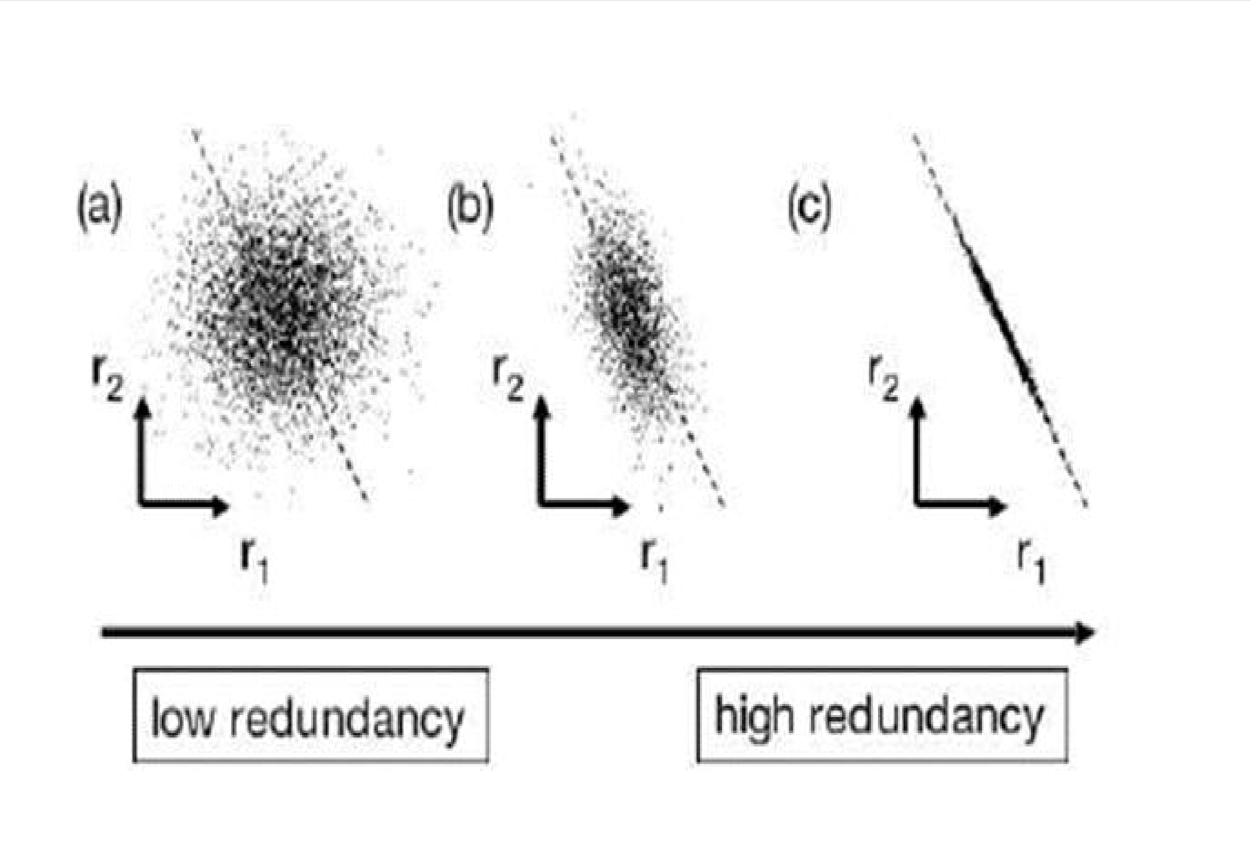
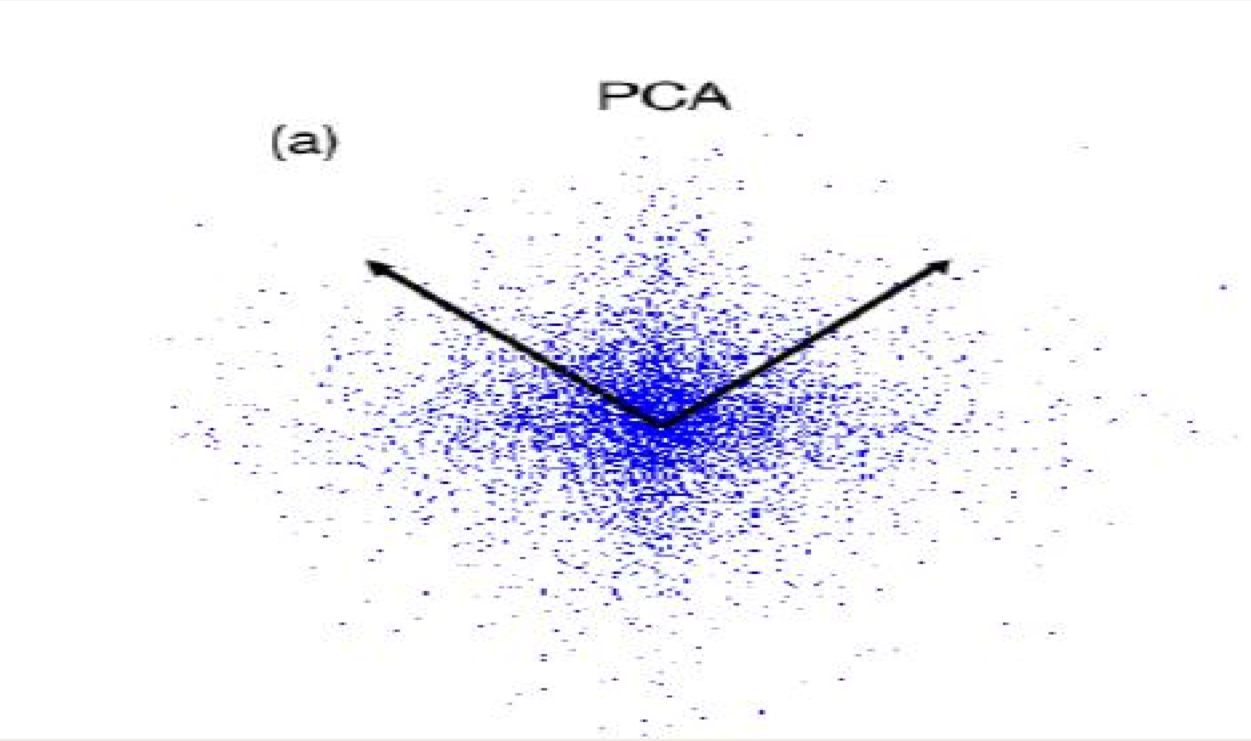


FIG. 2 (a) Simulated data of (x_A, y_A) for camera A. The signal and noise variances σ^2_{signal} and σ^2_{noise} are graphically represented by the two lines subtending the cloud of data. (b) Rotating these axes finds an optimal p^* where the variance and SNR are maximized. The SNR is defined as the ratio of the variance along p^* and the variance in the perpindicular direction.





PCA AND SVD

where $\boldsymbol{U} \in \mathbb{R}^{D\times D}$ and $\boldsymbol{V}^\top \in \mathbb{R}^{N\times N}$ are orthogonal matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times N}$ is a matrix whose only nonzero entries are the singular values $\sigma_{ii} \geqslant 0$. It then follows that

$$\boldsymbol{S} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\top = \frac{1}{N}\boldsymbol{U}\boldsymbol{\Sigma}\underbrace{\boldsymbol{V}^\top\boldsymbol{V}}_{=\boldsymbol{I}_N}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top = \frac{1}{N}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top . \tag{10.48}$$

Liên hệ giữa PCA AND SVD

Công thức tính covariance matrix hiệu chỉnh: $S = \frac{1}{n-1} X X^T$ (1)

Ta có có thể phân tích ma trận X sử dụng SVD như sau: $X = U \Sigma V^T$ (2)

Thế (2) vào (1) ta có: $S = \frac{1}{n-1} (U \Sigma V^T)(U \Sigma V^T)^T = \frac{1}{n-1} U \Sigma V^T V \Sigma U^T = \frac{1}{n-1} U \frac{\Sigma^2}{n-1} U^T$

Có nghĩa là mỗi singular vector của U tương ứng với mỗi gía trị $\lambda_i = \frac{{s_i}^2}{n-1}$

Ta có thể định nghĩa 1 ma trận Y = $\frac{1}{\sqrt{n-1}} X^T$ (Khi mà ở đó các cột của Y là 0 mean)

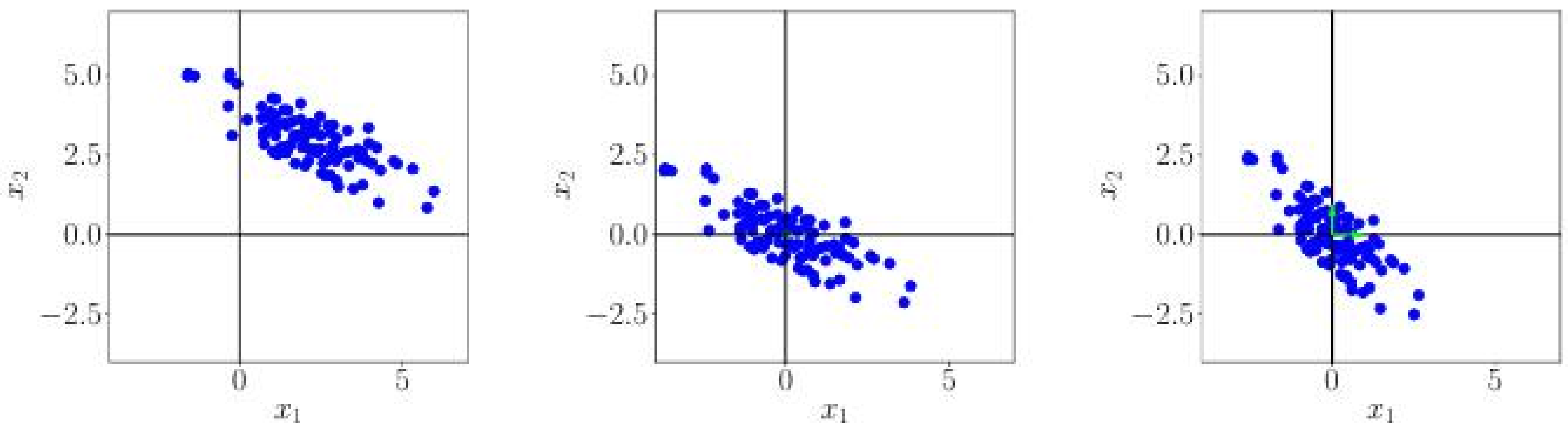
Ta có được:

$$\begin{aligned} \mathbf{Y}^T\mathbf{Y} &= \left(\frac{1}{\sqrt{n-1}}\mathbf{X}^T\right)^T \left(\frac{1}{\sqrt{n-1}}\mathbf{X}^T\right) \\ &= \frac{1}{n-1}\mathbf{X}^{TT}\mathbf{X}^T \\ &= \frac{1}{n-1}\mathbf{X}\mathbf{X}^T \\ \mathbf{Y}^T\mathbf{Y} &= \mathbf{C}_\mathbf{X} \end{aligned}$$

Trong đó $\boldsymbol{C_x}$ **là covariance matrix của X**

Khi mà ta xây dựng ma trận $Y^T Y$ bằng với covariance matrix của X, ta biết là các principal component của X cũng chính là các vector riêng của $\boldsymbol{C_x}$, khi mà ta sử dụng SVD để phân tích ma trận Y thì các cột của ma trận V chứa các vector riêng của ma trận $Y^T Y$=$\boldsymbol{C_x}$, Vì thế các **cột của ma trận V** chính là các *principal component* của X.

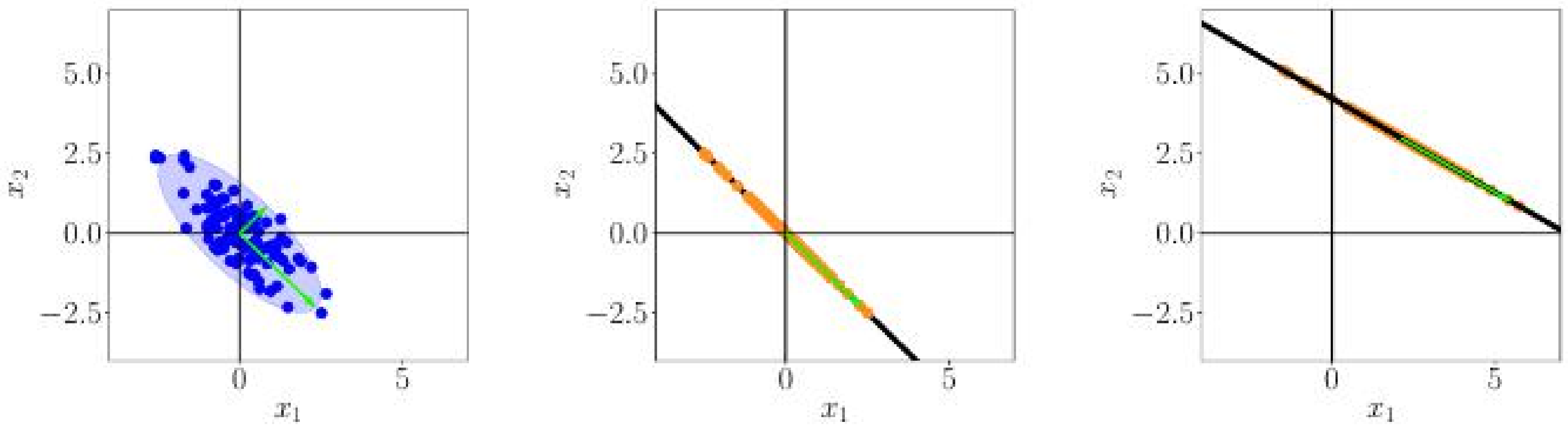
Key steps in pca



(a) Original dataset.

(b) Step 1: Centering by subtracting the mean from each data point.

(c) Step 2: Dividing by the standard deviation to make the data unit free. Data has variance 1 along each axis.



(d) Step 3: Compute eigenvalues and eigenvectors (arrows) of the data covariance matrix (ellipse).

(e) Step 4: Project data onto the principal subspace.

(f) Undo the standardization and move projected data back into the original data space from (a).

Figure 10.11 Steps of PCA. (a) Original dataset; (b) centering; (c) divide by standard deviation; (d) eigendecomposition; (e) projection; (f) mapping back to original data space.

What is Principal component analysis (pca)?

→

Problem:

- Working directly with high-dimensional data (such as images) come with difficulties

- Hard to analyze, expensive storage data vector

Example:

- A 640×480 pixel color image is a data point in a million-dimensional space, where every pixel responds to three dimensions, one for each color channel (red, green, blue).

What is Principal component analysis (pca)?

PRINCIPLE COMPONENT ANALYSIS

- Most common form of factor analysis
- Linear dimensionality reduction algorithm
- Describe the variation of **multiple correlated variables** by a group of new uncorrelated variables (**components**).

If 2 variables are (highly) correlated

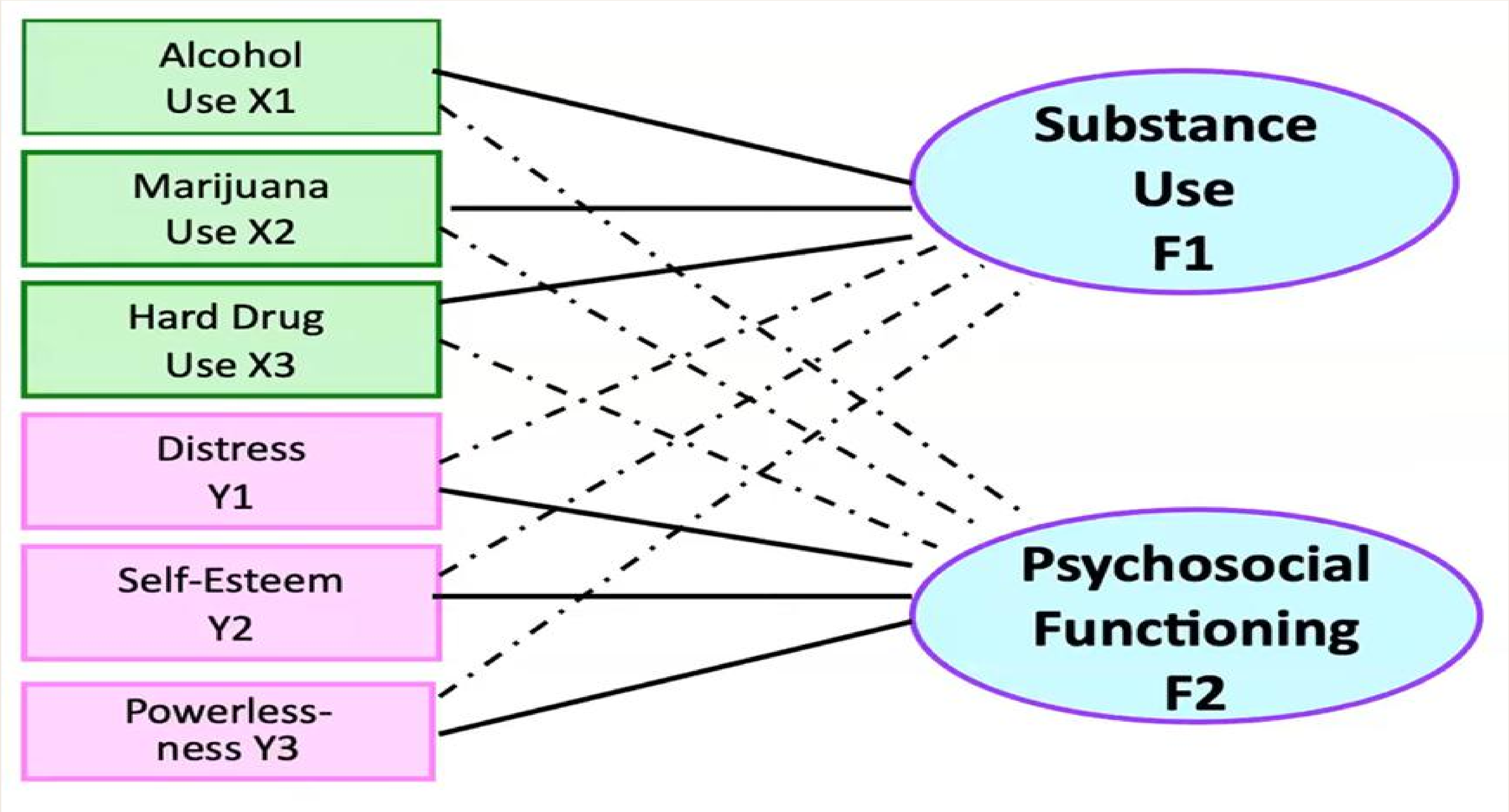
- They may reflect an underlying unobserved factor (*latent factor*)

-> Only need 1 variable

-> Latent variables can be known as ***'factors'*** or ***'principal components'***

What is Principal component analysis (pca)?

Example:



Computing PCA

- Mean subtraction
- Standardization
- Eigendecomposition of the covariance matrix
- Projection

→

Computing PCA

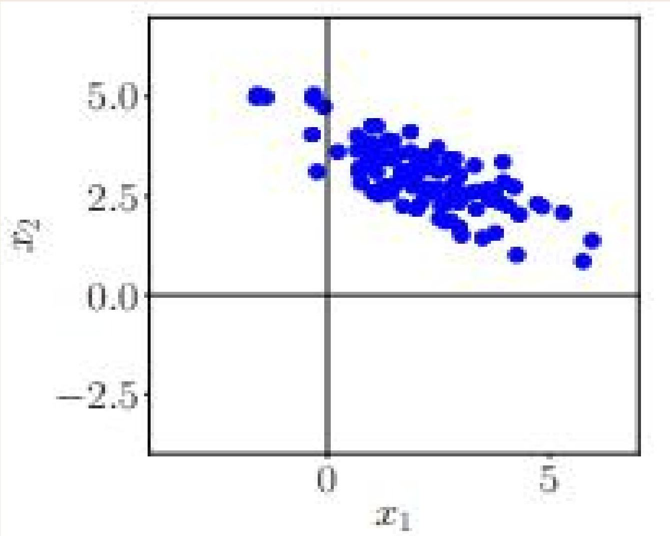
MEAN SUBTRACTION

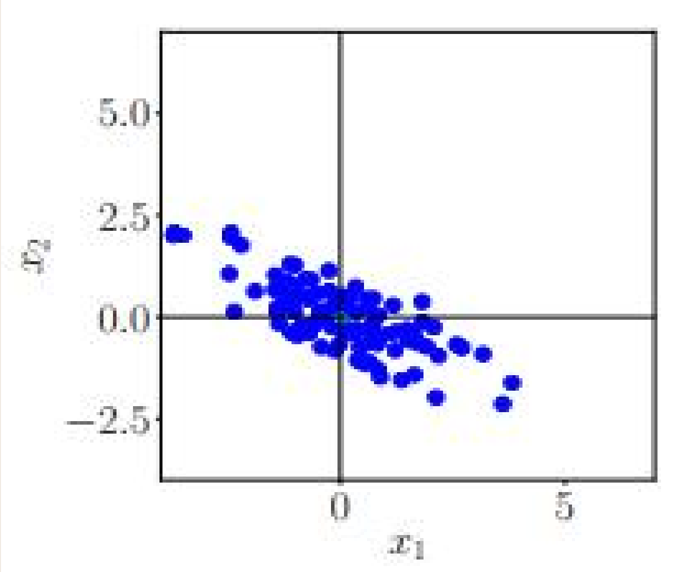
Centering the data by computing the mean $\boldsymbol{\mu}$ of the dataset and subtracting it from every single data point

The mean $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Subtracting data points to $\boldsymbol{\mu}$: $x_i = x_i - \mu$ where $x_i \in R^D$





Computing PCA

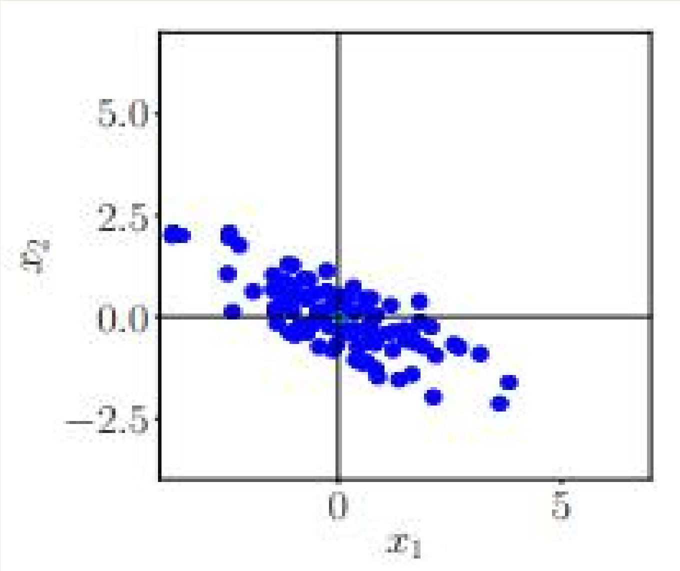
STANDARDIZATION

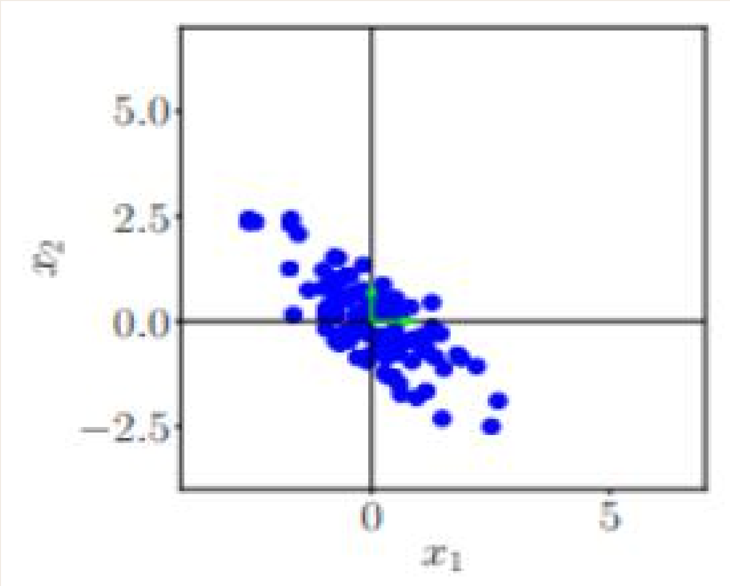
Divide the data points by the standard deviation $\boldsymbol{\sigma\_d}$ of the dataset for every dimension d = 1, . . . , D. Data has variance 1 along each axis.

The standard deviation:

$$\boldsymbol{\sigma_d} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

Standardization of the dataset: $x_i = \dfrac{x_i}{\sigma_d}$ where $x_i \in R^D$





Computing PCA

EIGENDECOMPOSITION OF THE COVARIANCE MATRIX

Compute the data covariance matrix and its eigenvalues and corresponding eigenvectors

The covariance(X, Y):

$$cov(X, Y) = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(y_i - \overline{Y})}{N}$$

The covariance matrix:

$$\begin{bmatrix} Var(x) & Cov(x, y) \\ Cov(y, x) & Var(y) \end{bmatrix}$$

Example: The covariance matrix of a 2 x 2 will be:

$$\begin{bmatrix} Var(x_1) & \dots & Cov(x_1, x_n) \\ . & \dots & . \\ Cov(x_n, x_1) & \dots & Var(x_n) \end{bmatrix}$$

→

Computing PCA

EIGENDECOMPOSITION OF THE COVARIANCE MATRIX

Compute the data covariance matrix and its eigenvalues and corresponding eigenvectors

Eigenvectors and eigenvalues:

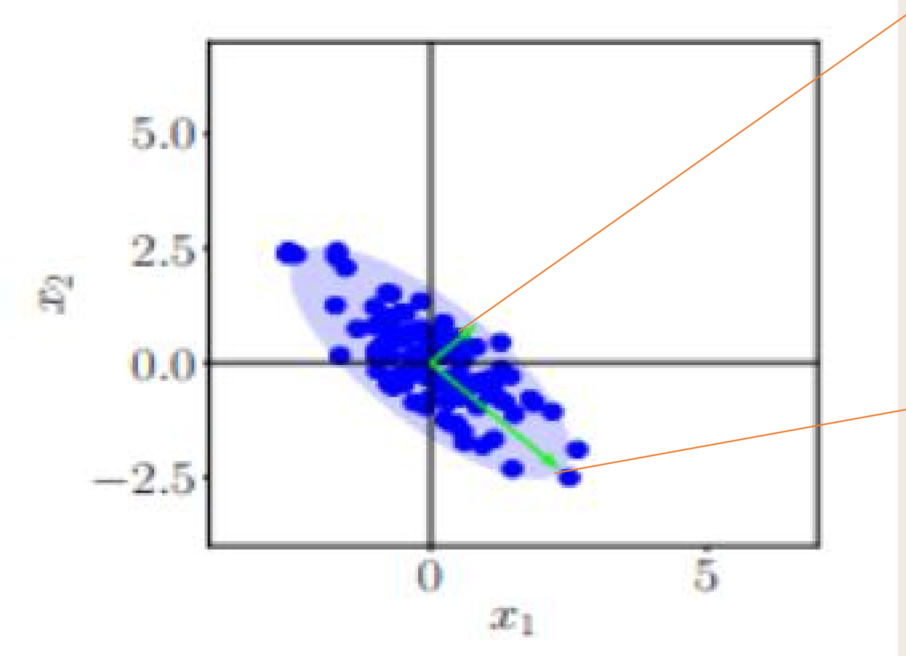
$$A\vec{x} = \lambda\vec{x} \qquad where\ \vec{x} : eigenvector$$

λ : eigenvalue

A : covariance matrix

Using eigenvalues and eigenvectors, we can find the main axes of our data. The first main axis (also called "first principal component") is the axis in which the data varies the most. The second main axis (also called "second principal component")

Second principal component axis (PC2)



First principal component axis (PC1)

Computing PCA

PROJECTION

Project the data points to the new axes $(\mathrm{PC}1, \mathrm{PC}2, \ldots)$ and put it back to the original data space by undo standardization

We obtain the projection as:

$$\widetilde{x}_* = BB^T x_*$$

With coordinate:

$$\widetilde{z}_* = B \quad x_*$$

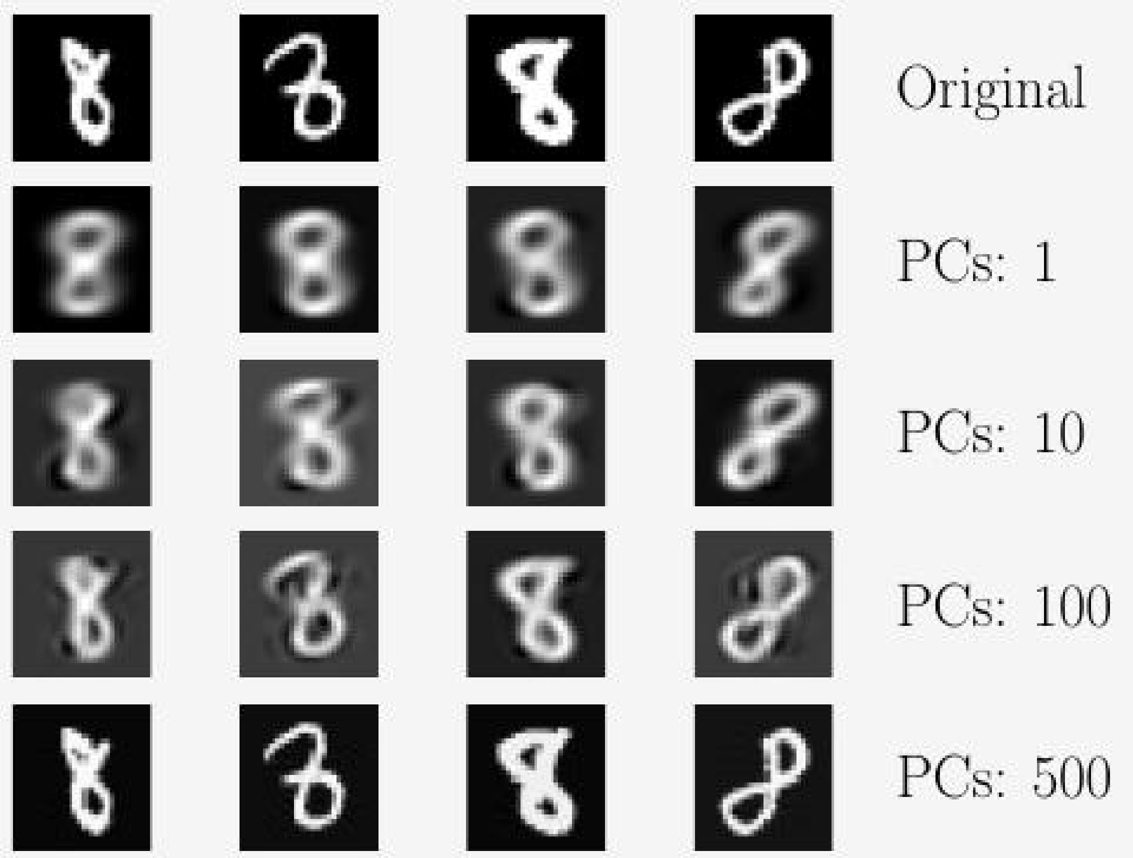
$$where\ B\ is\ the\ matrix\ that\ contains\ eigenvectors\ that\ are\ associated$$
$$with\ the\ largest\ eigenvalues\ of\ the\ data\ covariance\ matrix\ as\ column$$

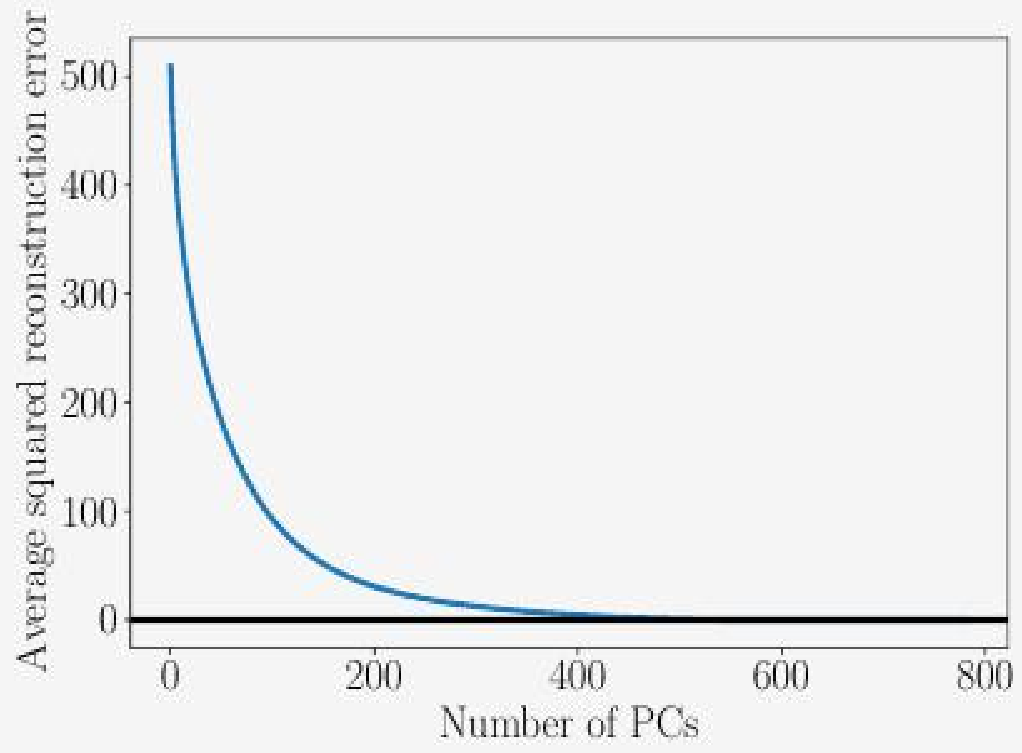
Undo standardization:

$$\widetilde{x}_*^{(\mathrm{d})} = \widetilde{x}_*^{(\mathrm{d})}\sigma_d + \mu_d \quad with\ d = 1,2,\ldots,D$$

→

DEMO


Original
PCs: 1
PCs: 10
PCs: 100
PCs: 500


Average squared reconstruction error
500
400
300
200
100
0
0
200
400
600
800
Number of PCs

Look at these EXAMPLE

We do experiment on Iris Dataset

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

Is everything clear?

Feel free to make this an open discussion for questions or clarifications before proceeding.

And we’re done for the day! Thank you

Use this space for announcements, homeworks, or ways students can approach you if ever they have questions.