Warm-up as you log in

1. https://www.sporcle.com/games/MrChewypoo/minimalist_disney
2. https://www.sporcle.com/games/Stanford0008/minimalist-cartoons-slideshow
3. https://www.sporcle.com/games/MrChewypoo/minimalist

Announcements

Assignments

- HW7 (online)
  - Due Tue 3/31, 11:59 pm
- HW8 (written + programming)
  - Out this week
  - Due Tue 4/7, 11:59 pm

Introduction to Machine Learning

Dimensionality Reduction
PCA

Instructor: Pat Virtue

Outline

- **Supervised vs Unsupervised Learning**
- **Dimensionality Reduction**
  - High-dimensional data
  - Learning (low dimensional) representations
- **Principal Component Analysis (PCA)**
  - Examples: 2D and 3D
  - Data for PCA
  - PCA Definition
  - Objective functions for PCA
  - PCA, Eigenvectors, and Eigenvalues

Supervised vs Unsupervised Learning

Supervised Learning

$$D = \{ y^{(i)}, \vec{x}^{(i)} \}_{i=1}^{N}$$

label / output

Classification
Regression

$$\hat{y} = h(x)$$

$$y^{(i)} \longleftrightarrow \hat{y}^{(i)} = h(\vec{x}^{(i)})$$

Unsupervised Learning

$$D = \{ \vec{x}^{(i)} \}_{i=1}^{N}$$

Clustering

Dimensionality Reduction

$$\vec{x}^{(i)} \longleftrightarrow g(f(\vec{x}^{(i)}))$$

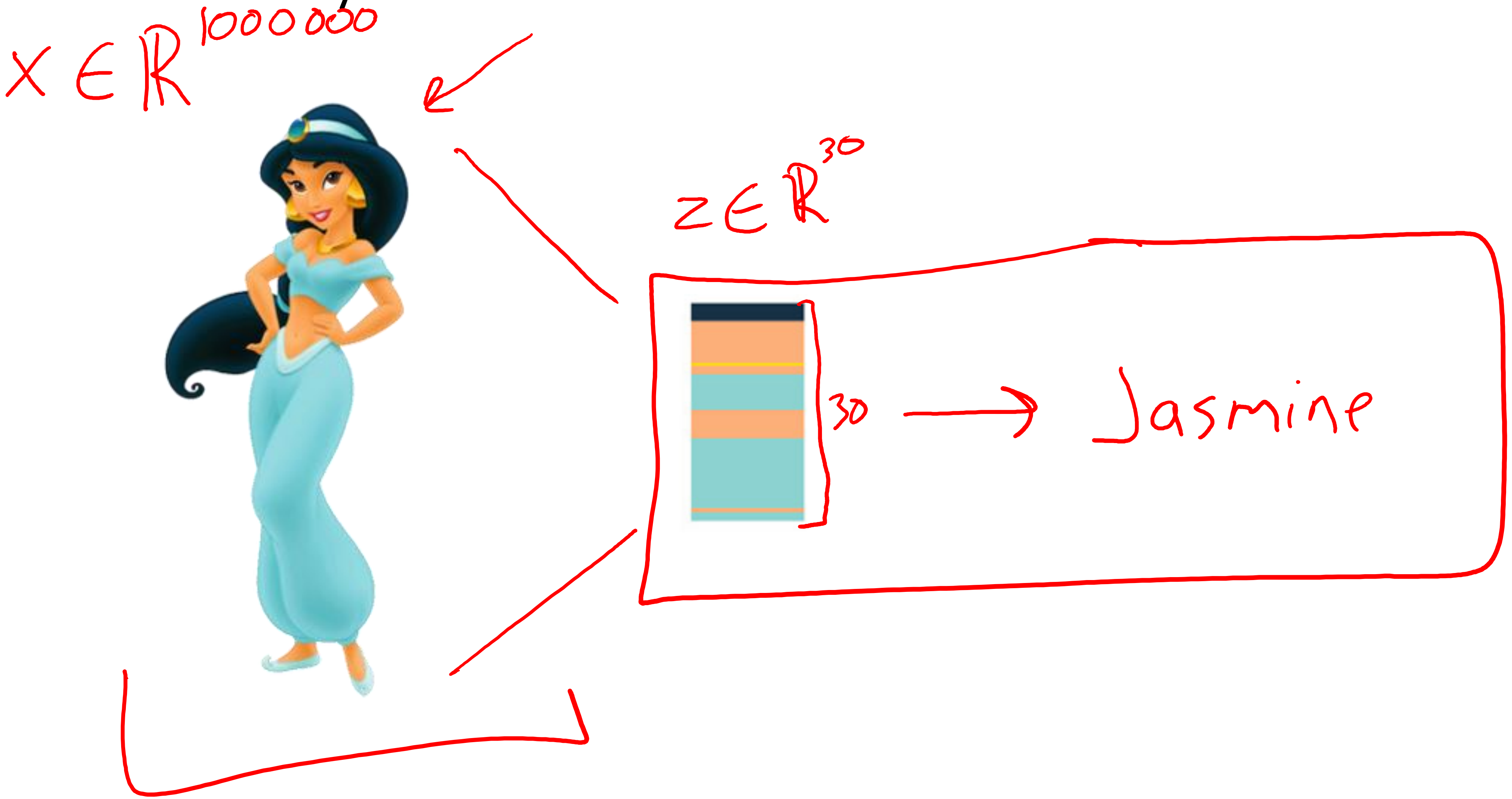
Warm-up as you log in

1. https://www.sporcle.com/games/MrChewypoo/minimalist_disney

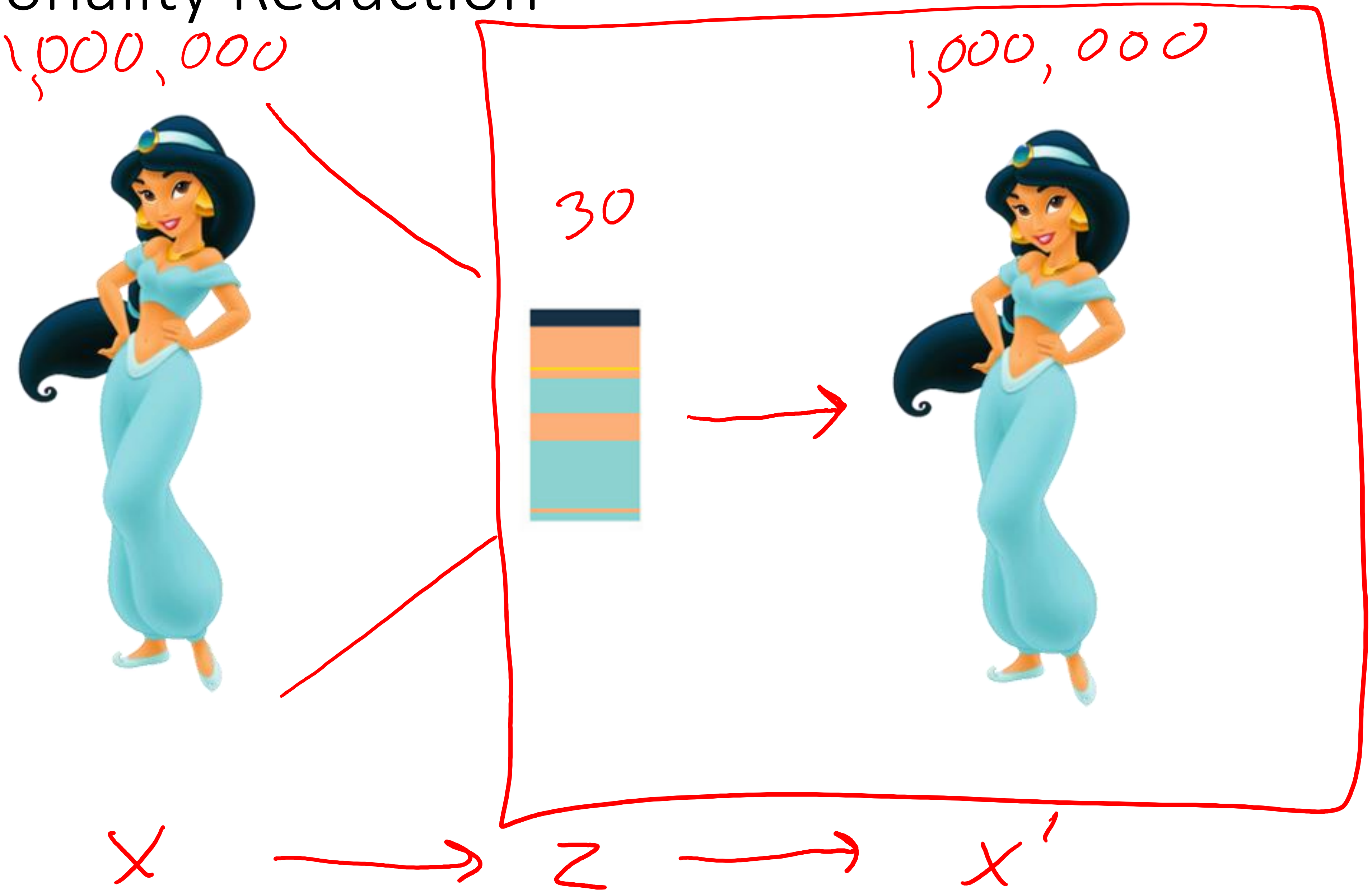
2. https://www.sporcle.com/games/Stanford0008/minimalist-cartoons-slideshow

3. https://www.sporcle.com/games/MrChewypoo/minimalist

Dimensionality Reduction



Dimensionality Reduction



Dimensionality Reduction

For each $x^{(i)} \in \mathbb{R}^M$ find representation $u^{(i)} \in \mathbb{R}^K$ where $K \ll M$

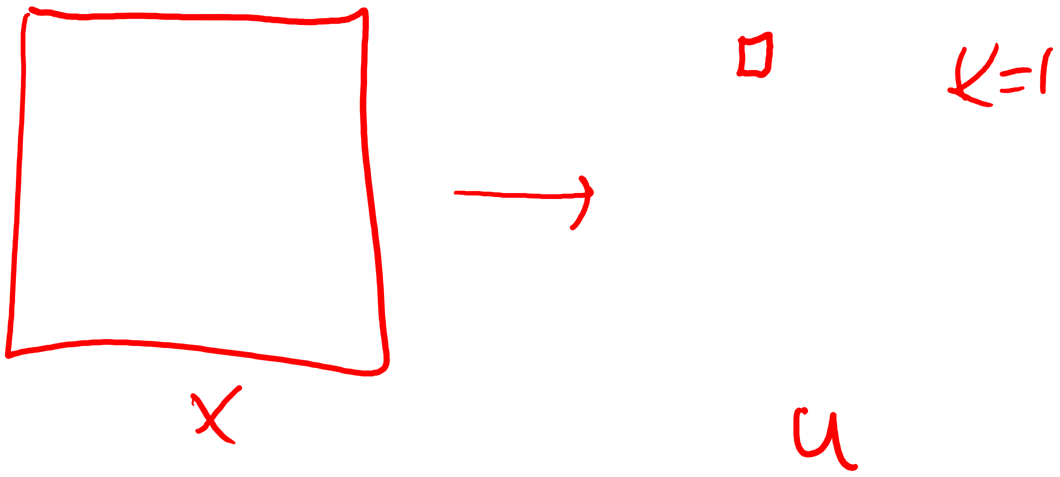
$$\vec{u} = f(\vec{x})$$

$$\vec{x}' = g(\vec{u})$$

$$\frac{1}{M} \|\vec{x} - \vec{x}'\|_2^2$$

$$\frac{1}{N} \frac{1}{M} \|X - X'\|_F^2$$

Reconstruction error



High Dimension Data

Examples of high dimensional data:

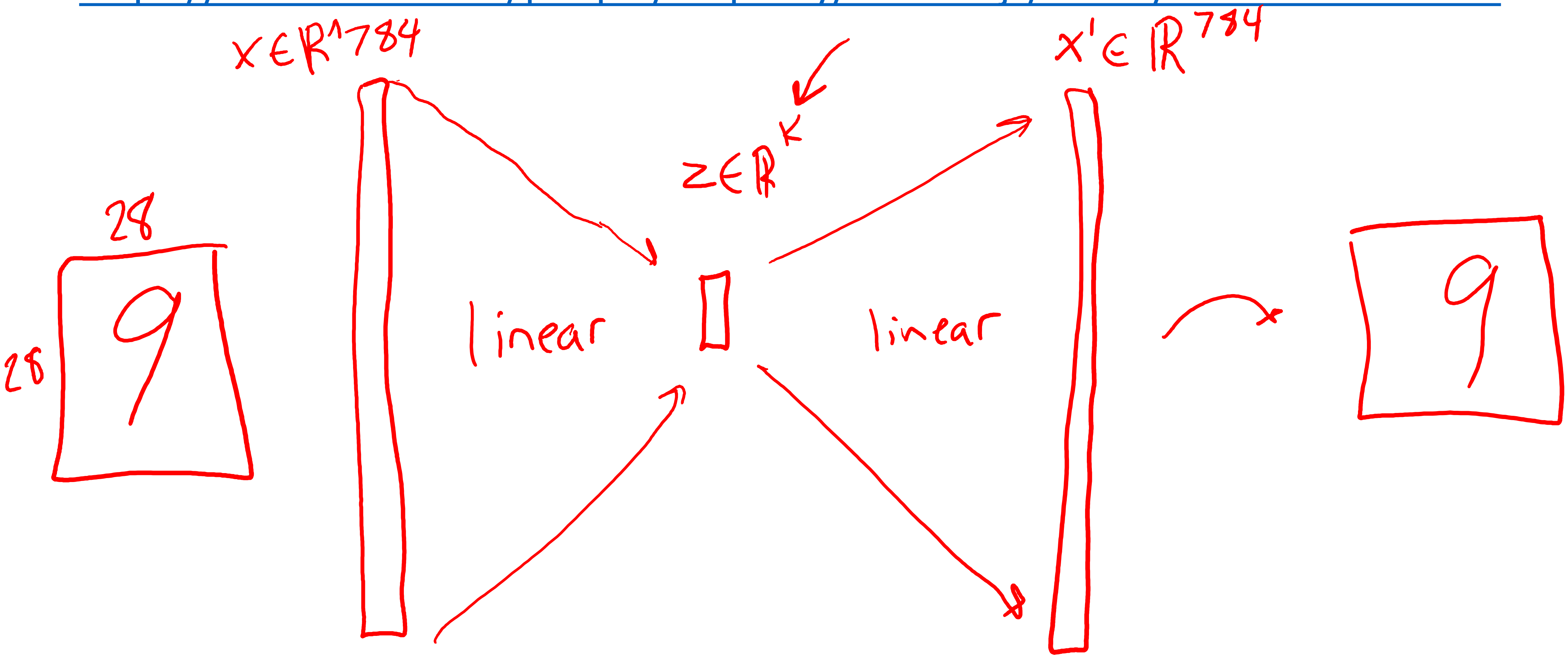
– High resolution images (millions of pixels)

1,000 x 1,000
1 byte / pixel
↓
0.5 MB

Dimensionality Reduction

http://timbaumann.info/svd-image-compression-demo/

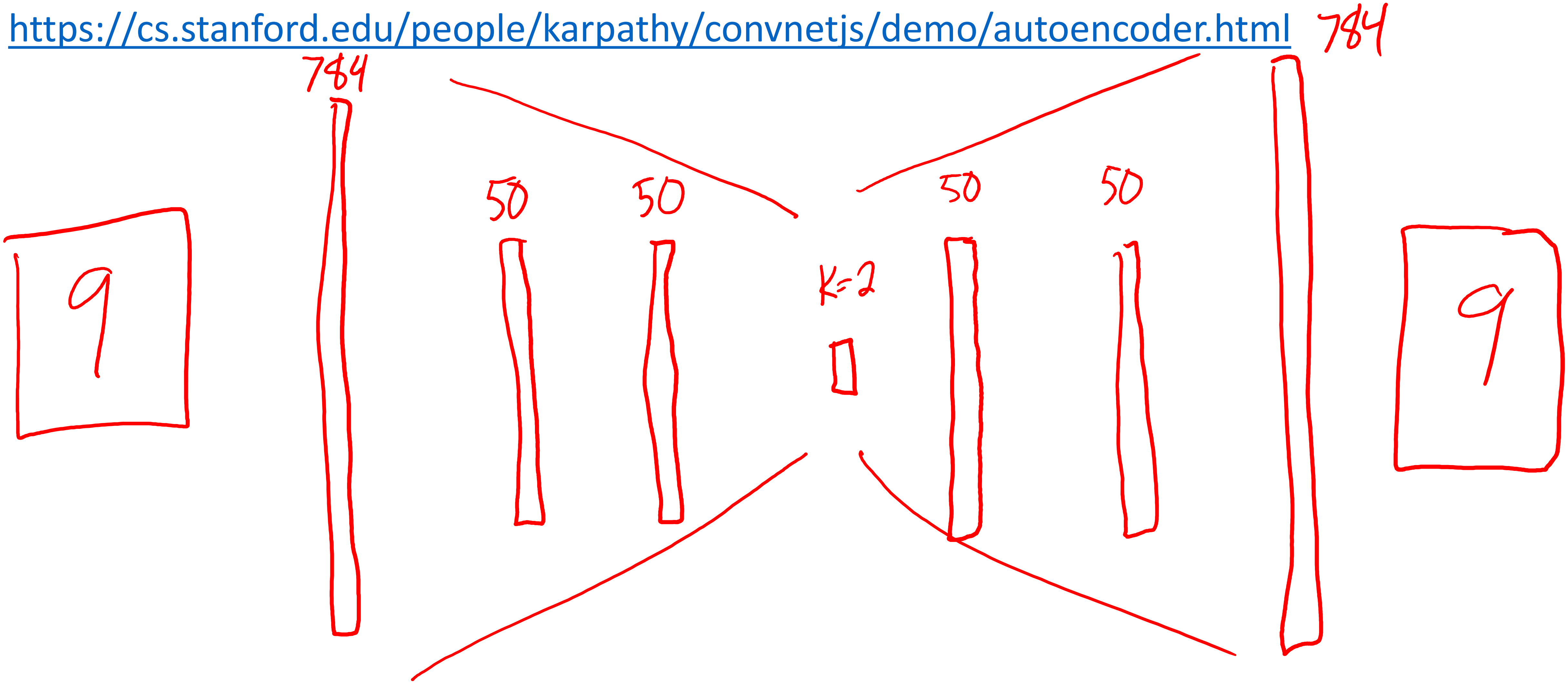
https://cs.stanford.edu/people/karpathy/convnetjs/demo/autoencoder.html



Dimensionality Reduction

http://timbaumann.info/svd-image-compression-demo/

https://cs.stanford.edu/people/karpathy/convnetjs/demo/autoencoder.html



High Dimension Data

Examples of high dimensional data:

– Multilingual News Stories
  (vocabulary of hundreds of thousands of words)


Hilo Tribune.

The Tribune Prints the News

We Lead—Others Follow

KILAUEA CRATER CONTINUES IN ERUPTION

FOUR DAYS OF ACTIVITY

ANOTHER OUTBREAK OCCURS

HONOLULU GETS APPROPRIATIONS FOR HARBOR AND LIGHTHOUSE

River and Harbor Bill Passes.

Postmaster Prant Confirmed.

Latest Sugar Quotations.

Fire At New Orleans.


叻報


Македония

PRAVA BRAĆA

Истински братя

DEKLARACIJA.

Декларация


குடிஅரசு

"KUDI ARASU"

High Dimension Data

Examples of high dimensional data:

– Brain Imaging Data (100s of MBs per scan)

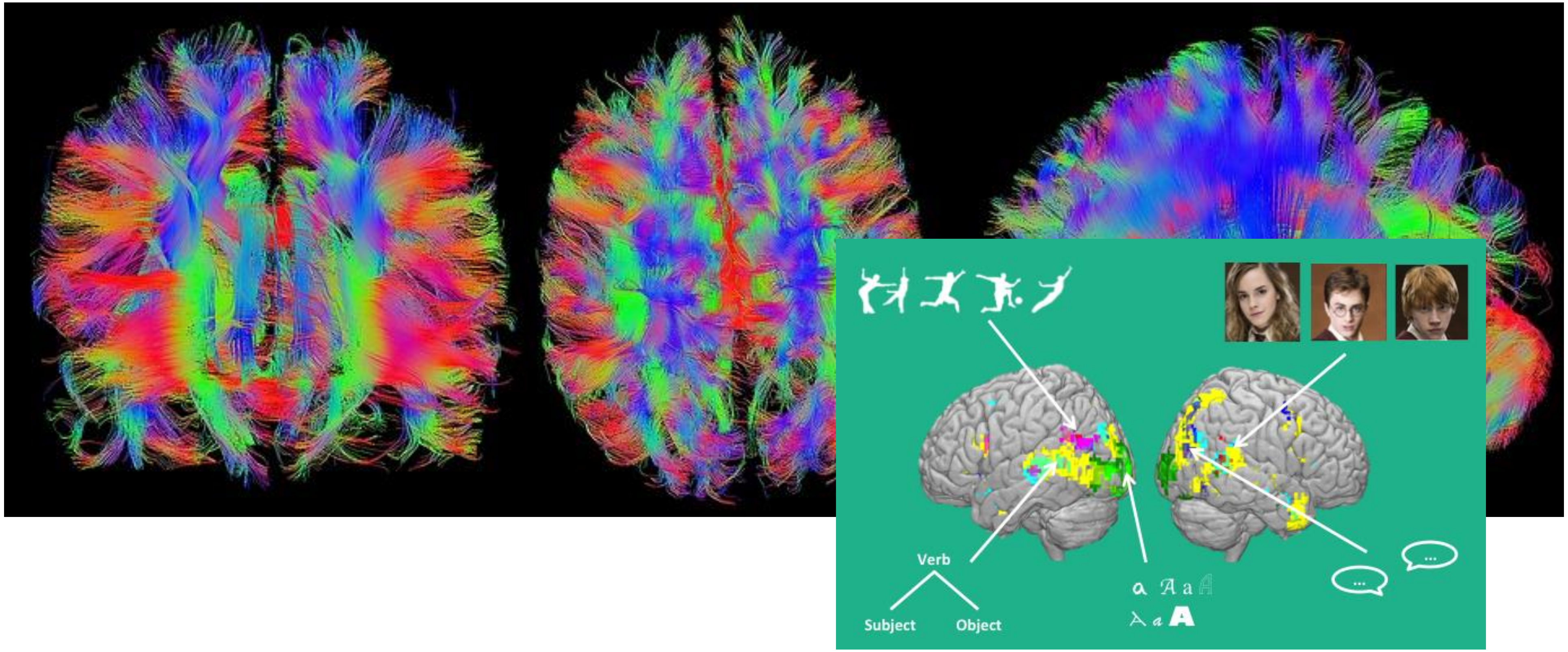


Image from (Wehbe et al., 2014)

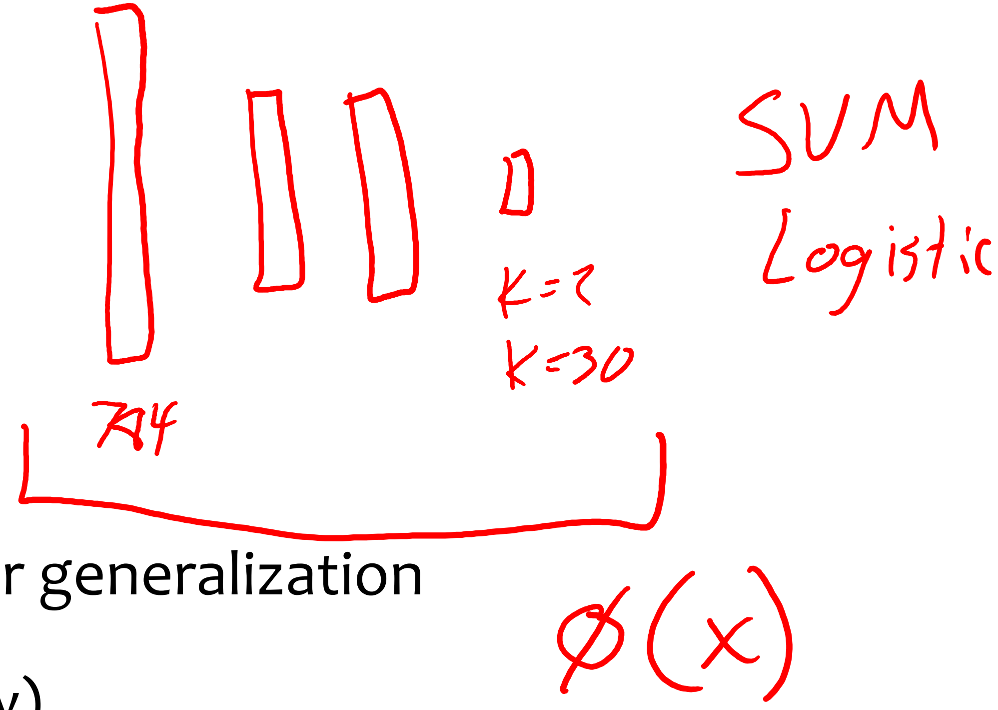
Image from https://pixabay.com/en/brain-mrt-magnetic-resonance-imaging-1728449/

Learning Representations

PCA, Kernel PCA, ICA: Powerful unsupervised learning techniques for extracting hidden (potentially lower dimensional) structure from high dimensional datasets.

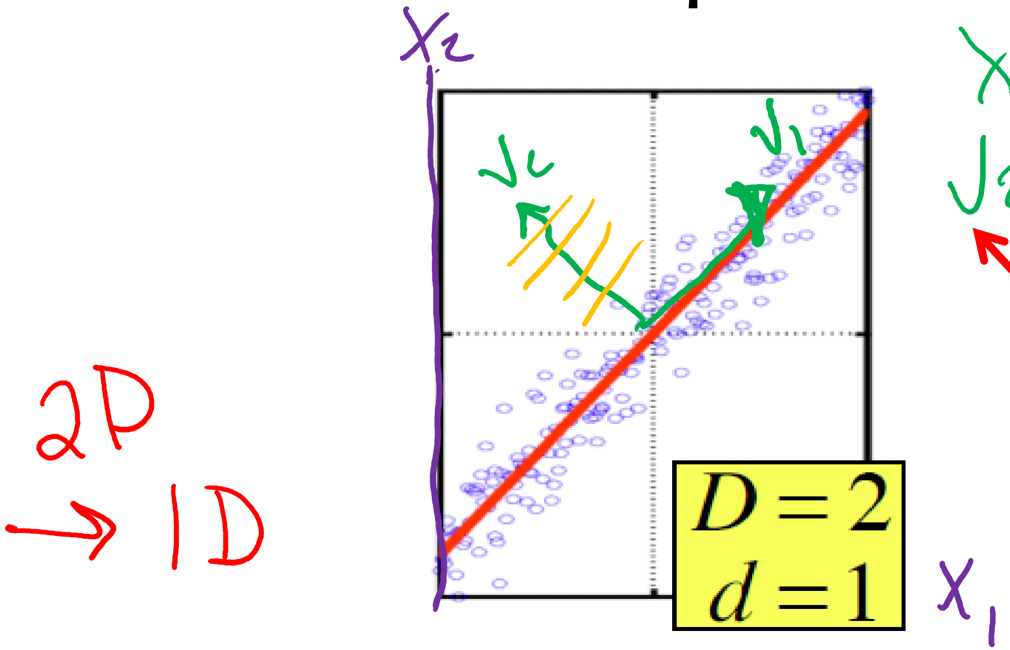
Useful for:

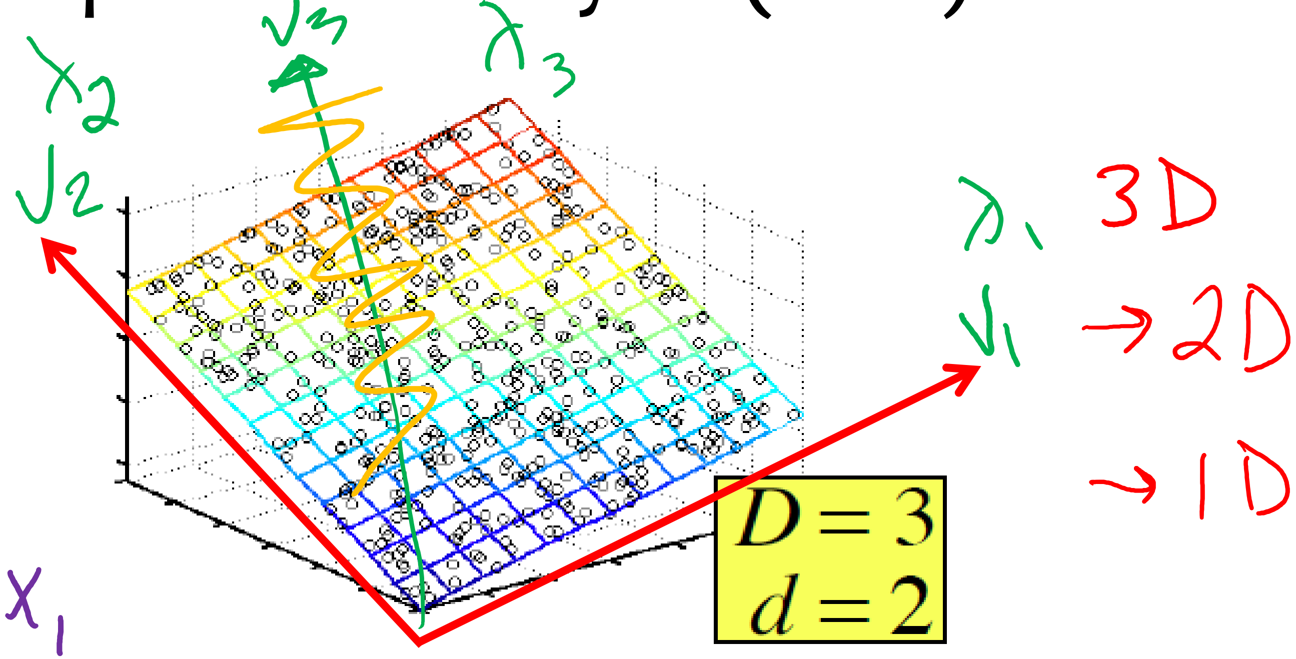
- Visualization
- More efficient use of resources (e.g., time, memory, communication)
- Statistical: fewer dimensions → better generalization
- Noise removal (improving data quality)
- Further processing by machine learning algorithms



PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA)



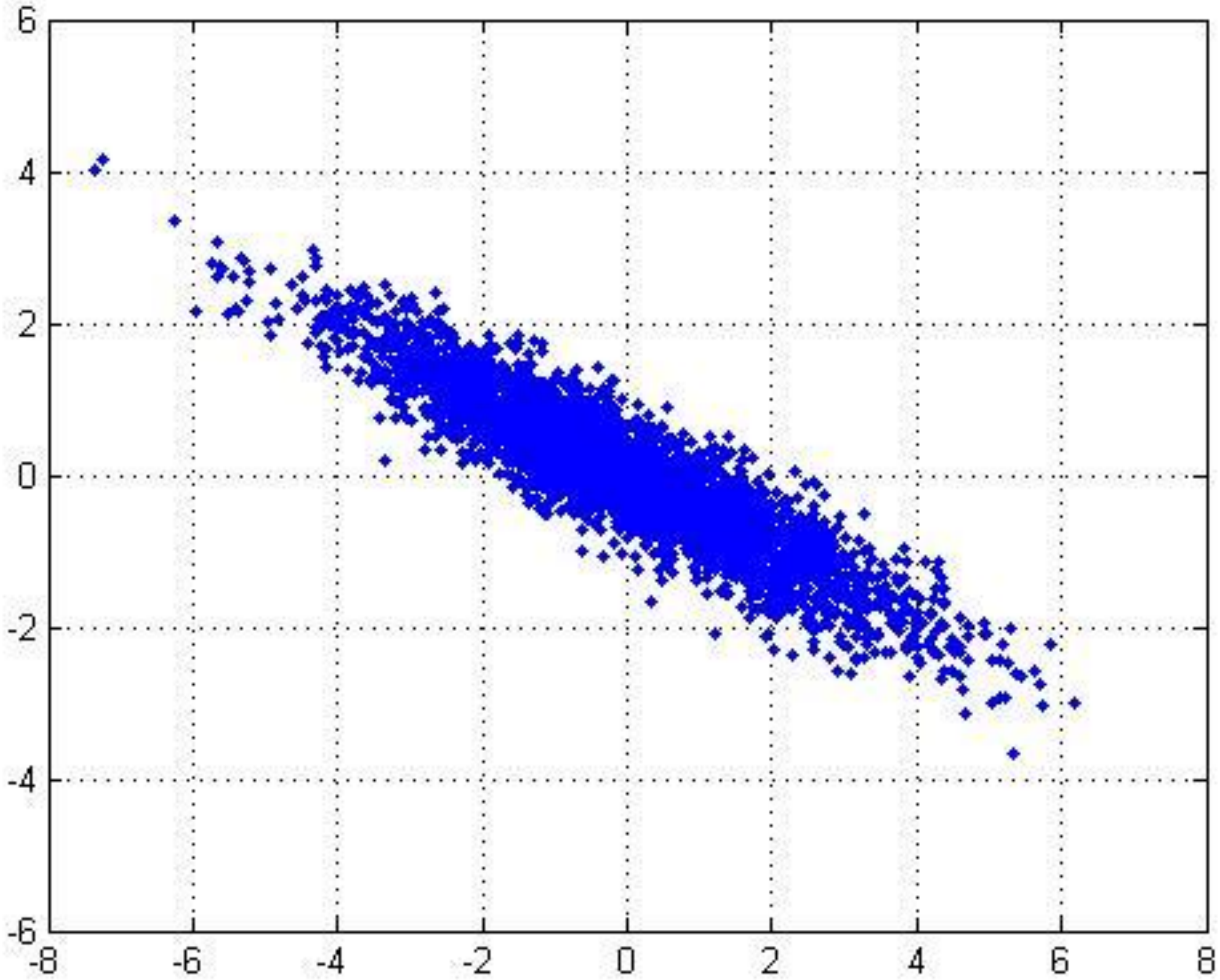


In case where data lies on or near a low d-dimensional linear subspace, axes of this subspace are an effective representation of the data.

Identifying the axes is known as Principal Components Analysis, and can be obtained by using classic matrix computation tools (Eigen or Singular Value Decomposition).

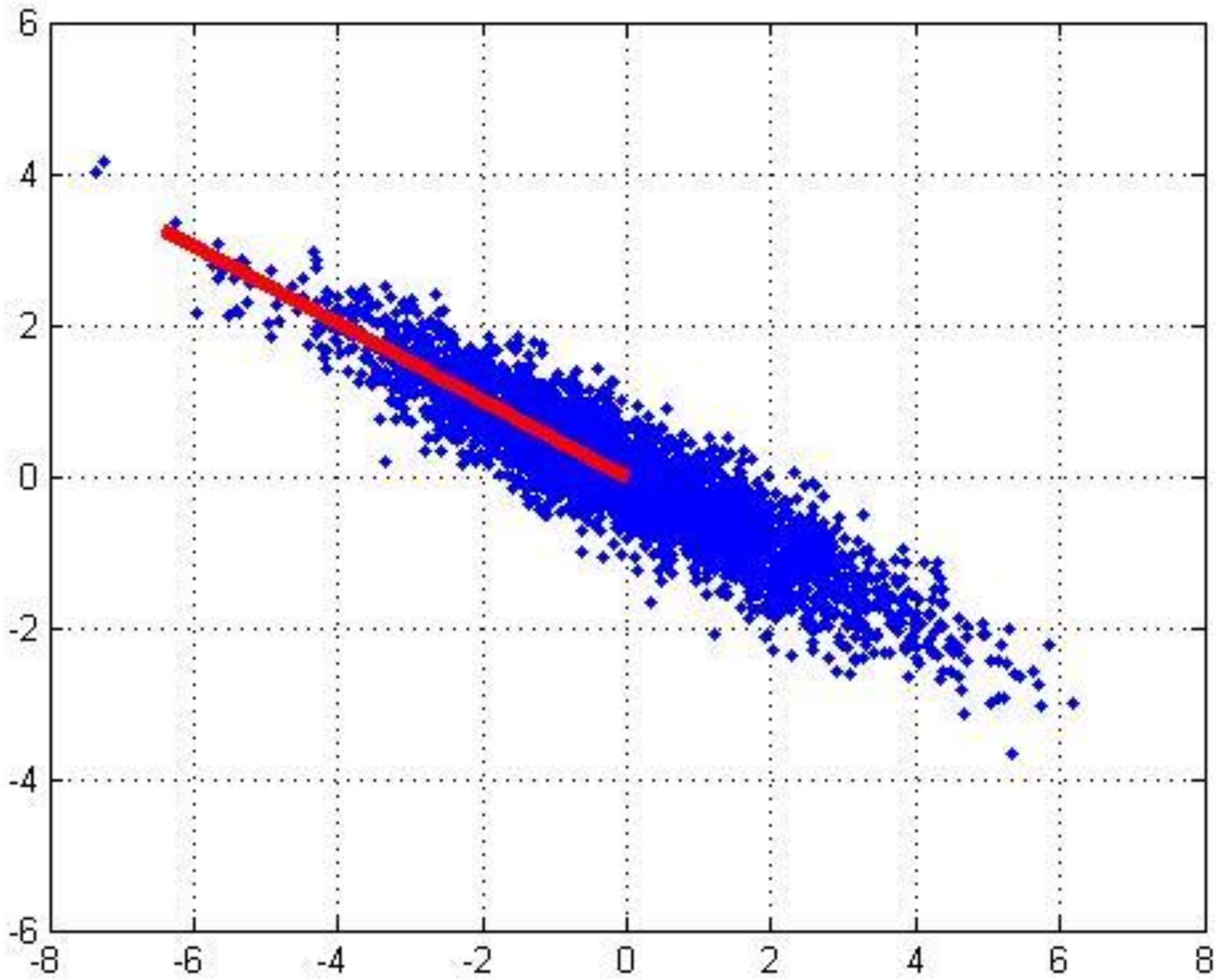
Slide from Nina Balcan

2D Gaussian dataset



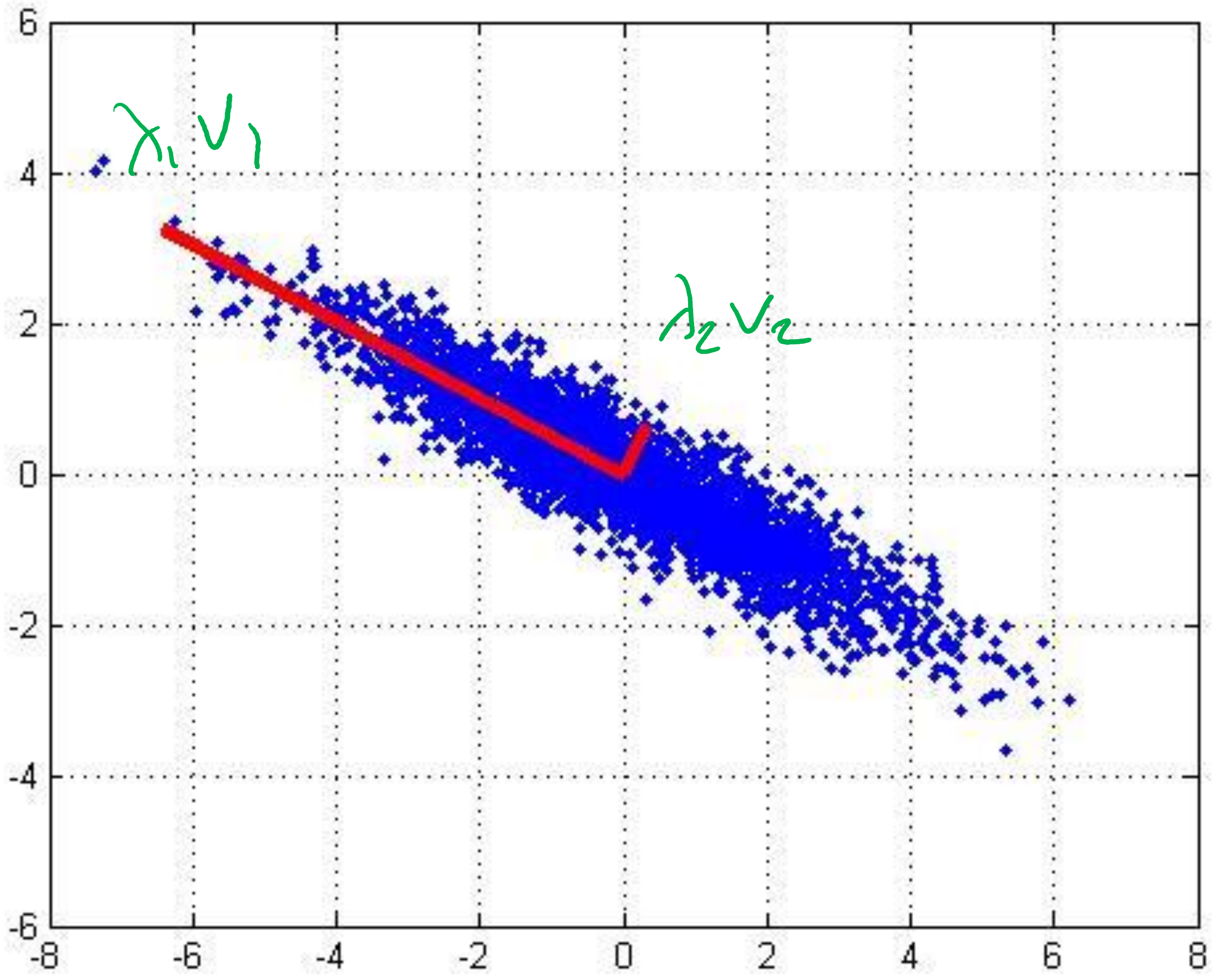
Slide from Barnabas Poczos

1st PCA axis



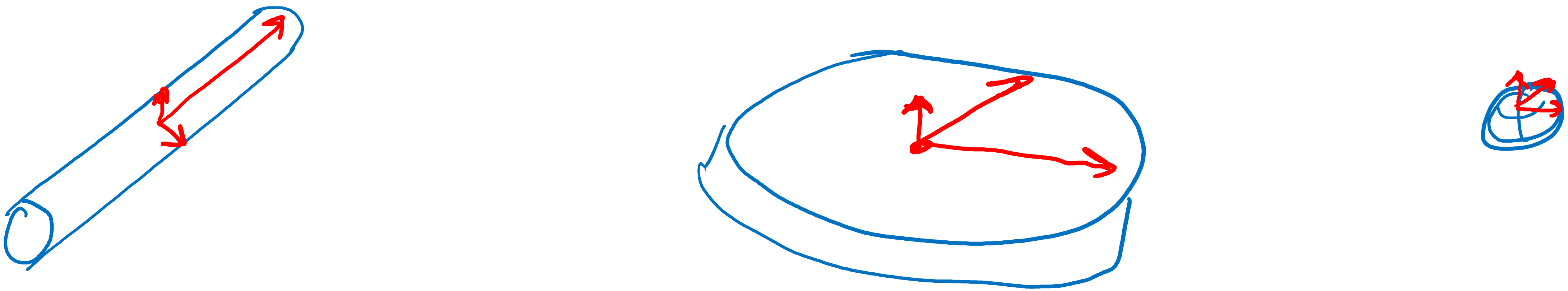
Slide from Barnabas Poczos

2nd PCA axis



Slide from Barnabas Poczos

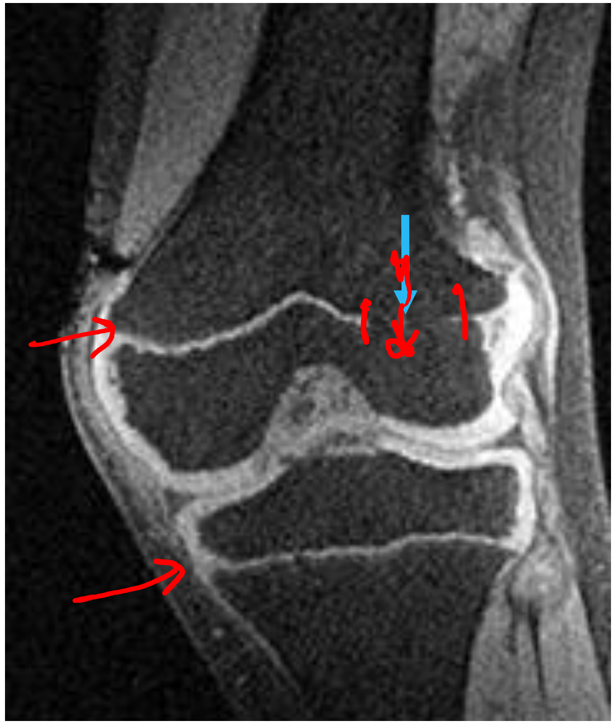
PCA Axes



Growth Plate Imaging

Growth Plate Disruption and Limb Length Discrepancy

8 year-old boy with previous fracture and
4cm leg length discrepancy

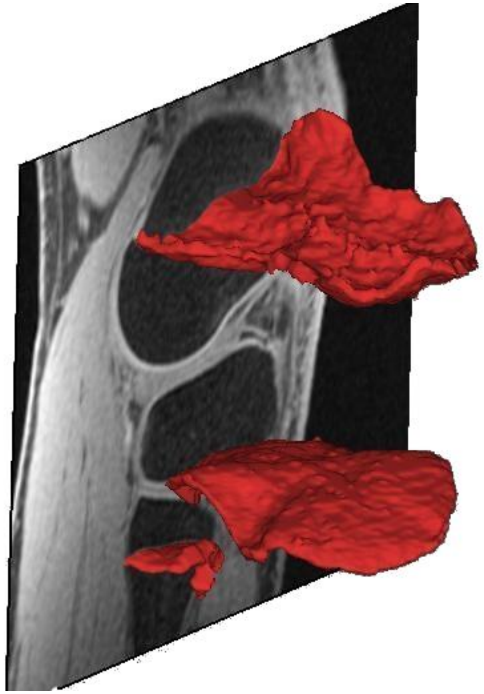


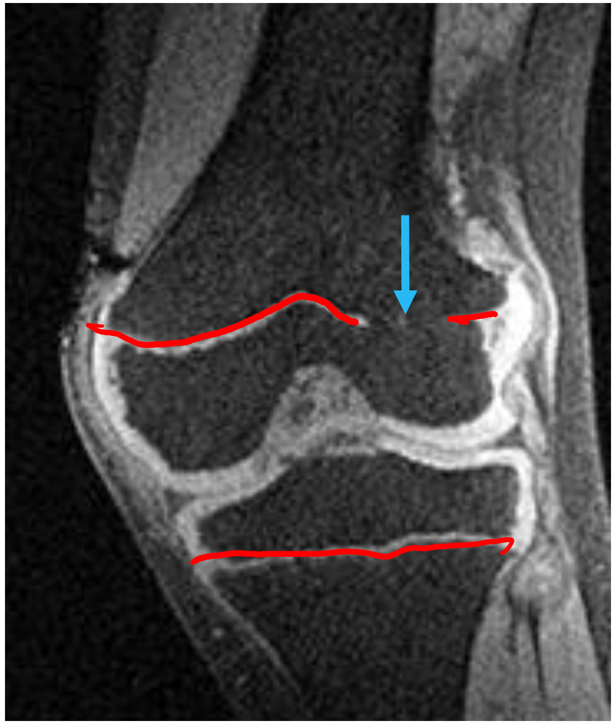
Images Courtesy
H. Potter, H.S.S.

Growth Plate Imaging

Growth Plate Disruption and Limb Length Discrepancy

8 year-old boy with previous fracture and
4cm leg length discrepancy



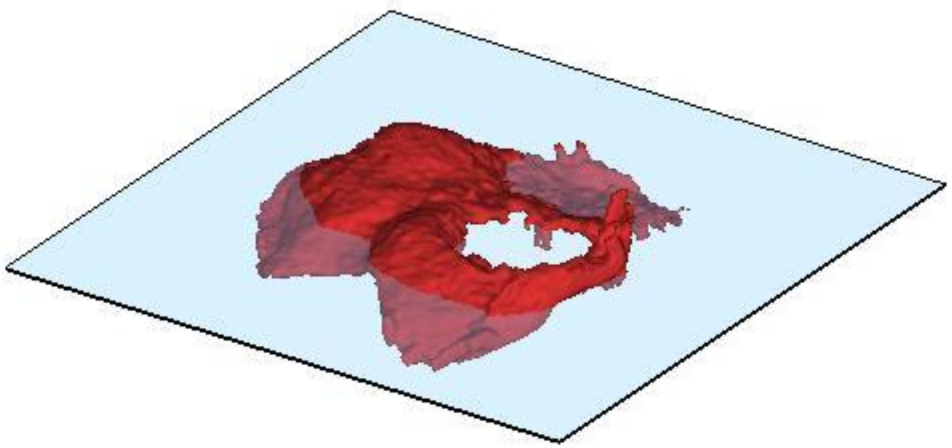


Images Courtesy
H. Potter, H.S.S.

imagination at work

Growth Plate Imaging

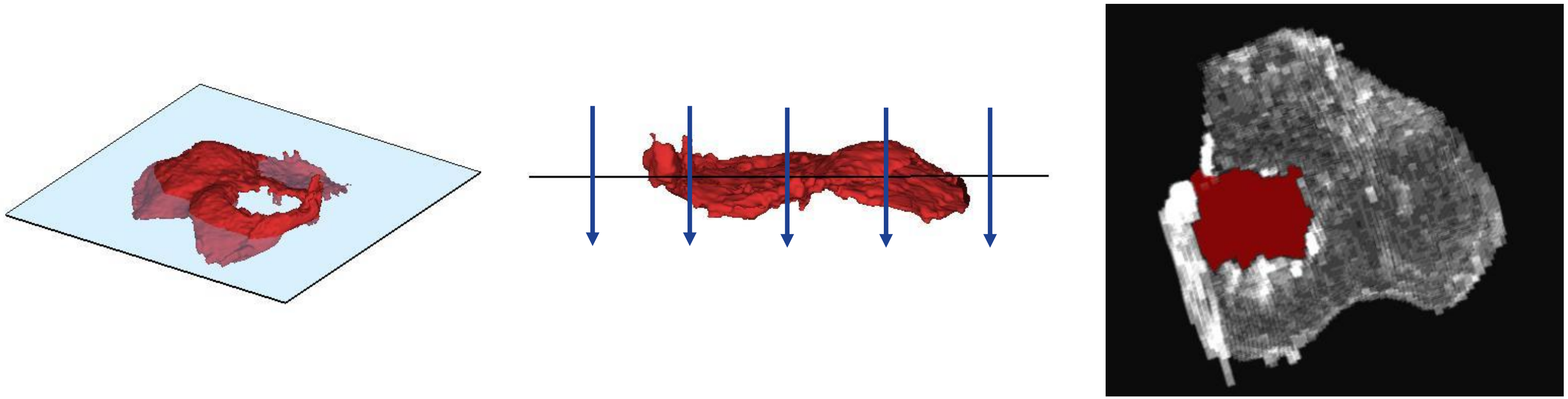
Area Measurement



imagination at work

Growth Plate Imaging

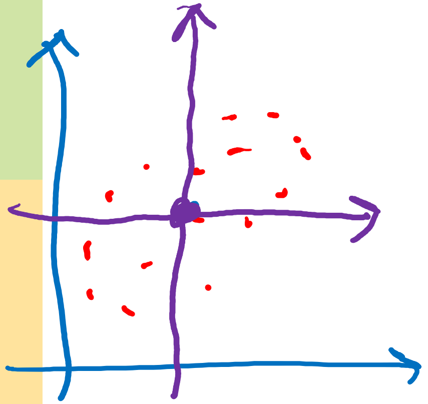
Area Measurement



Flatten Growth Plate to Enable 2D Area Measurement

Data for PCA

$$\mathcal{D} = \{\mathbf{x}^{(i)}\}_{i=1}^{N} \qquad \mathbf{X} = \begin{bmatrix} (\mathbf{x}^{(1)})^T \\ (\mathbf{x}^{(2)})^T \\ \vdots \\ (\mathbf{x}^{(N)})^T \end{bmatrix}$$



We assume the data is **centered**

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{(i)} = \mathbf{0}$$

Q: What if your data is **not** centered?

A: Subtract off the sample mean

Slide from Matt Gormley

Sample Covariance Matrix

$K = XX^T$

$N \times N$

The sample covariance matrix is given by:

$$\Sigma_{jk} = \frac{1}{N} \sum_{i=1}^{N} (x_j^{(i)} - \mu_j)(x_k^{(i)} - \mu_k)$$

Since the data matrix is centered, we rewrite as:

$M \times M$

$$\mathbf{\Sigma} = \frac{1}{N} \mathbf{X}^T \mathbf{X}$$

$$\mathbf{X} = \begin{bmatrix} (\mathbf{x}^{(1)})^T \\ (\mathbf{x}^{(2)})^T \\ \vdots \\ (\mathbf{x}^{(N)})^T \end{bmatrix}$$

Slide from Matt Gormley

Piazza Poll 1

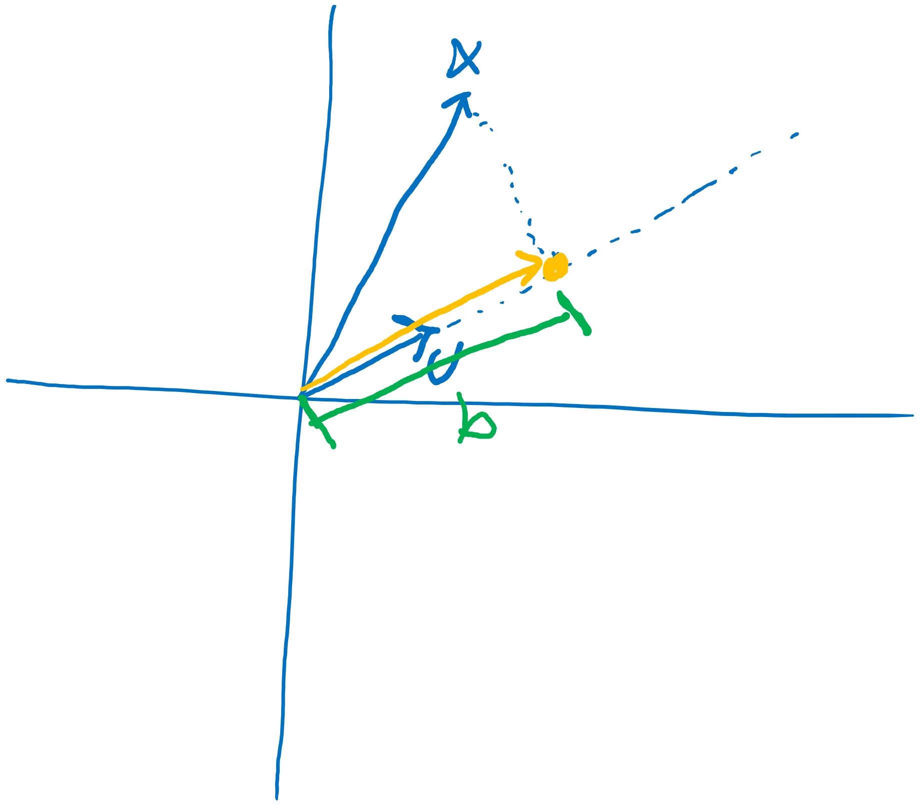
What is the projection of point $\boldsymbol{x}$ onto vector $\boldsymbol{v}$, assuming that $||\boldsymbol{v}||_2 = 1$?

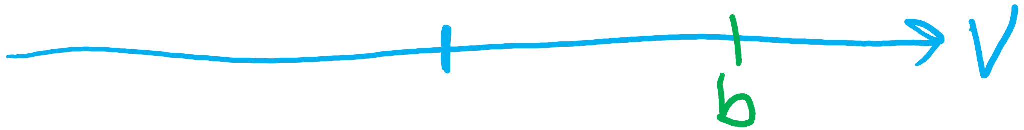
A. $\boldsymbol{v}\boldsymbol{x}$

B. $\boldsymbol{v}^T\boldsymbol{x}$ = b

C. $(\boldsymbol{v}^T\boldsymbol{x})\boldsymbol{v}$ = $\vec{c}$

D. $\boldsymbol{v}^T\boldsymbol{x}\,\boldsymbol{x}^T\boldsymbol{v}$





Principle Component Analysis PCA)

- Sketch of PCA algorithm
- Two PCA objective functions

Rotation of Data (and back)

1. For any orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{M \times M}$
2. Rotate to new space: $\boldsymbol{z}^{(i)} = \boldsymbol{V}\boldsymbol{x}^{(i)} \quad \forall i$
3. (Un)rotate back: $\boldsymbol{x}'^{(i)} = \boldsymbol{V}^T \boldsymbol{z}^{(i)}$

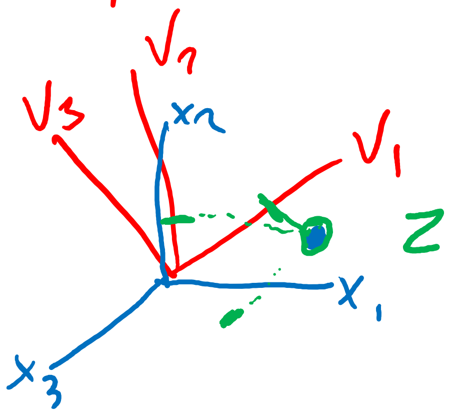
$$V = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k \\ \vdots \\ v_M^T \end{bmatrix} \quad x$$

V is orthogonal

$$V^T V = I$$

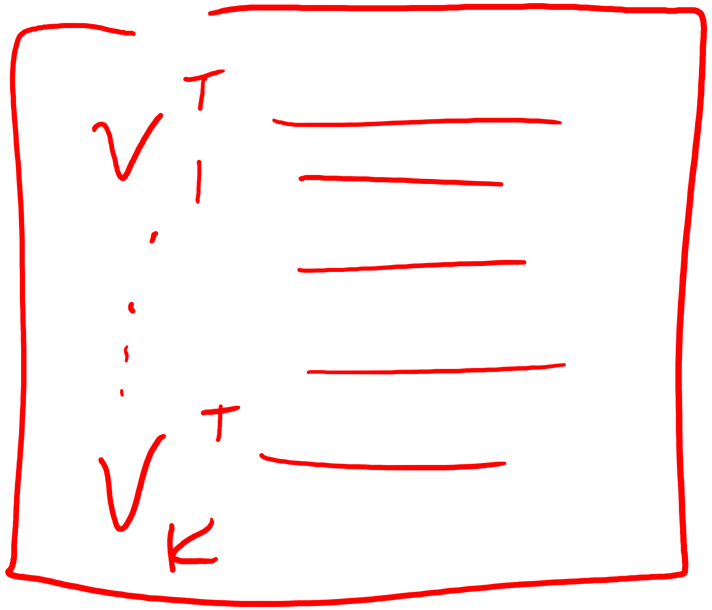
$$V^T = V^{-1}$$

$$v_1^T x$$



Sketch of PCA

1. Randomly select $\boldsymbol{V} \in \mathbb{R}^{K \times M}$
2. Project down: $\boldsymbol{z}^{(i)} = \boldsymbol{V}\boldsymbol{x}^{(i)} \quad \forall i$
3. Reconstruct up: $\boldsymbol{x}'^{(i)} = \boldsymbol{V}^T \boldsymbol{z}^{(i)}$



Sketch of PCA

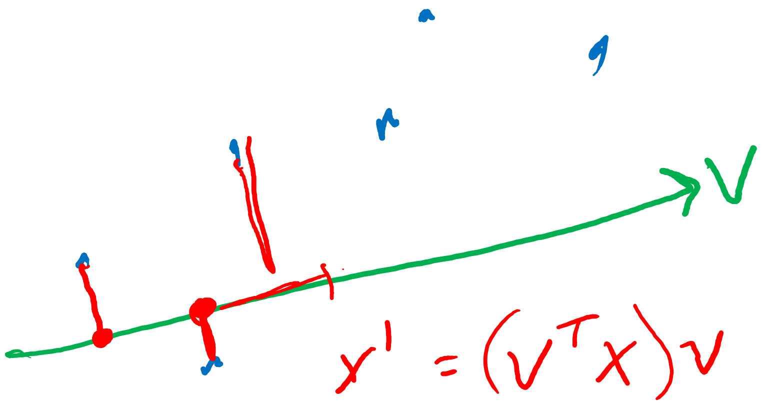
1. Randomly select $\boldsymbol{V} \in \mathbb{R}^{K \times M}$
2. Project down: $\boldsymbol{z}^{(i)} = \boldsymbol{V}\boldsymbol{x}^{(i)} \quad \forall i$
3. Reconstruct up: $\boldsymbol{x}'^{(i)} = \boldsymbol{V}^T \boldsymbol{z}^{(i)}$

Definition of PCA

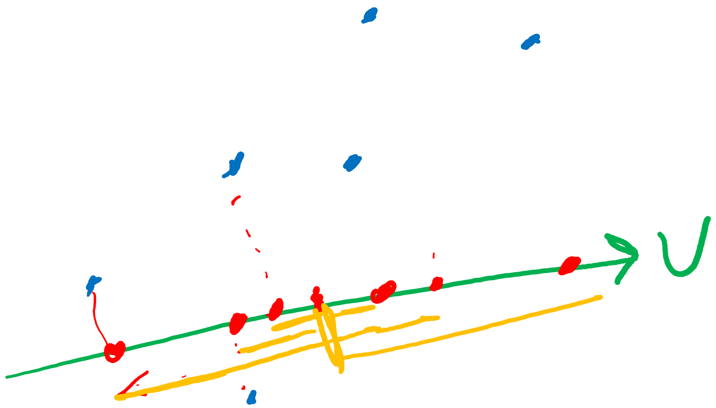
1. Select v_1 that best explains data
2. Select next v_j that (v_2)
   i. Is orthogonal to $v_1, \ldots, v_{j-1}$
   ii. Best explains remaining data
3. Repeat 2 until desired amount of data is explained $x - x'$

Select "Best" Vector

Reconstruction Error vs Variance of Projection



$x' = (v^T x) v$



Reconstruction error

$$\| x^{(i)} - x'^{(i)} \|_2^2$$

$$v^* = \underset{v}{\operatorname{argmin}} \| x^{(i)} - (v^T x) v \|_2^2$$

s.t. $\|v\|_2 = 1$

Variance of Projection

$$v^* = \underset{v}{\operatorname{argmax}} \sum_{i=1}^{N} (v^T x^{(i)})^2$$

st $\|v\|_2 = 1$