

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

Unsupervised Learning

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 $K \ll M$

High Dimension Data

Examples of high dimensional data:

- High resolution images (millions of pixels)



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)



High Dimension Data

Examples of high dimensional data:

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

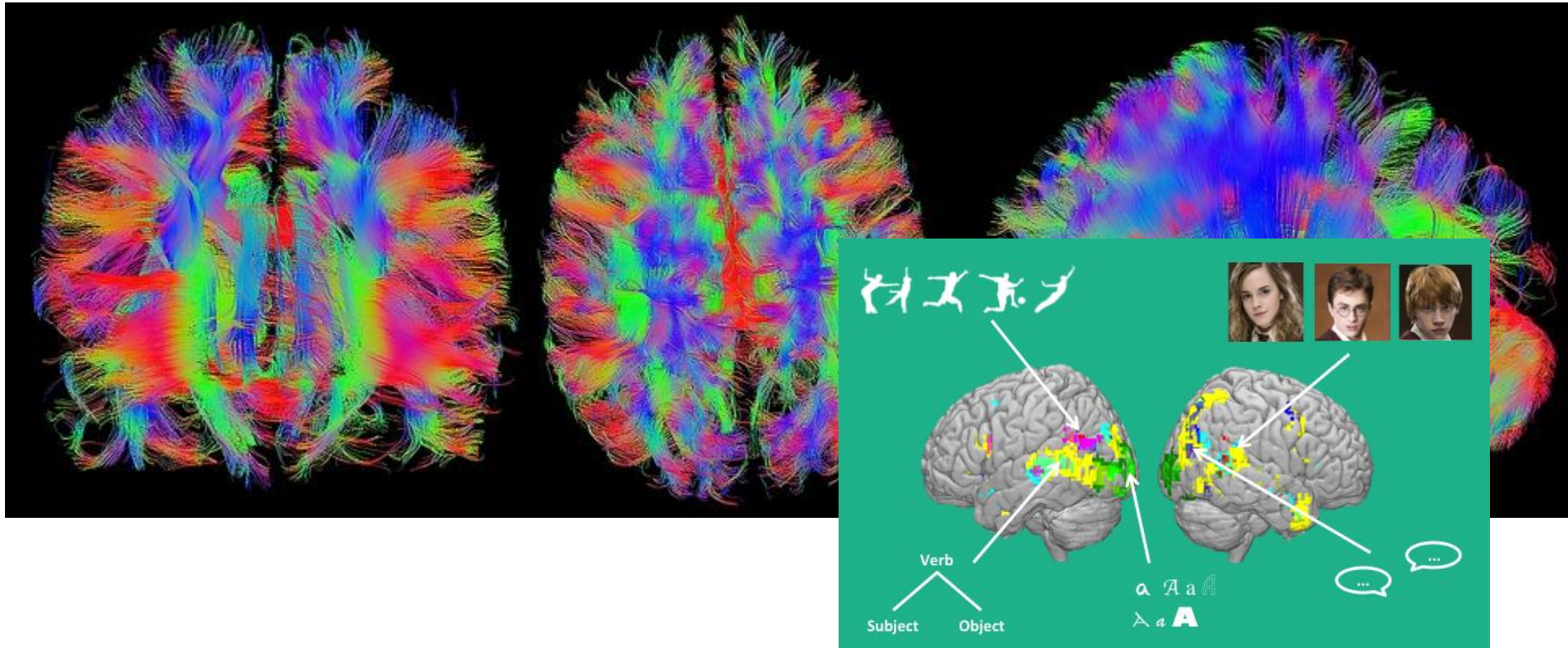


Image from (Wehbe et al., 2014)

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

Learning Representations

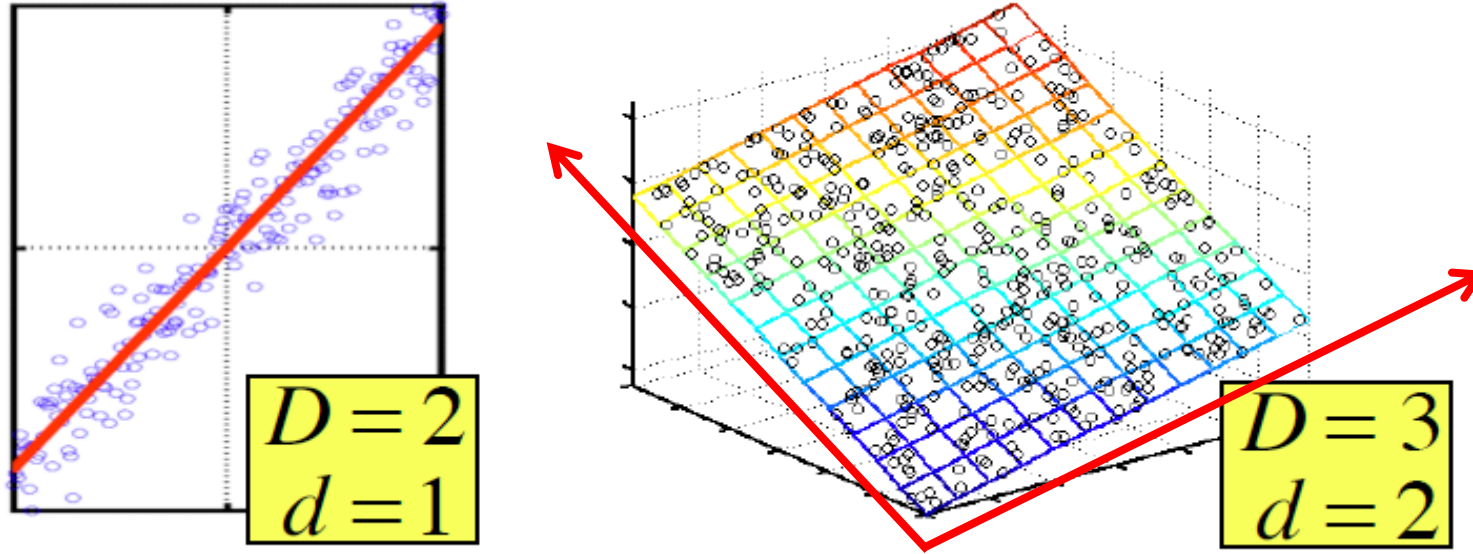
PCA, Kernel PCA, ICA: Powerful unsupervised learning techniques for extracting hidden (potentially lower dimensional) structure from high dimensional datasets.

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)

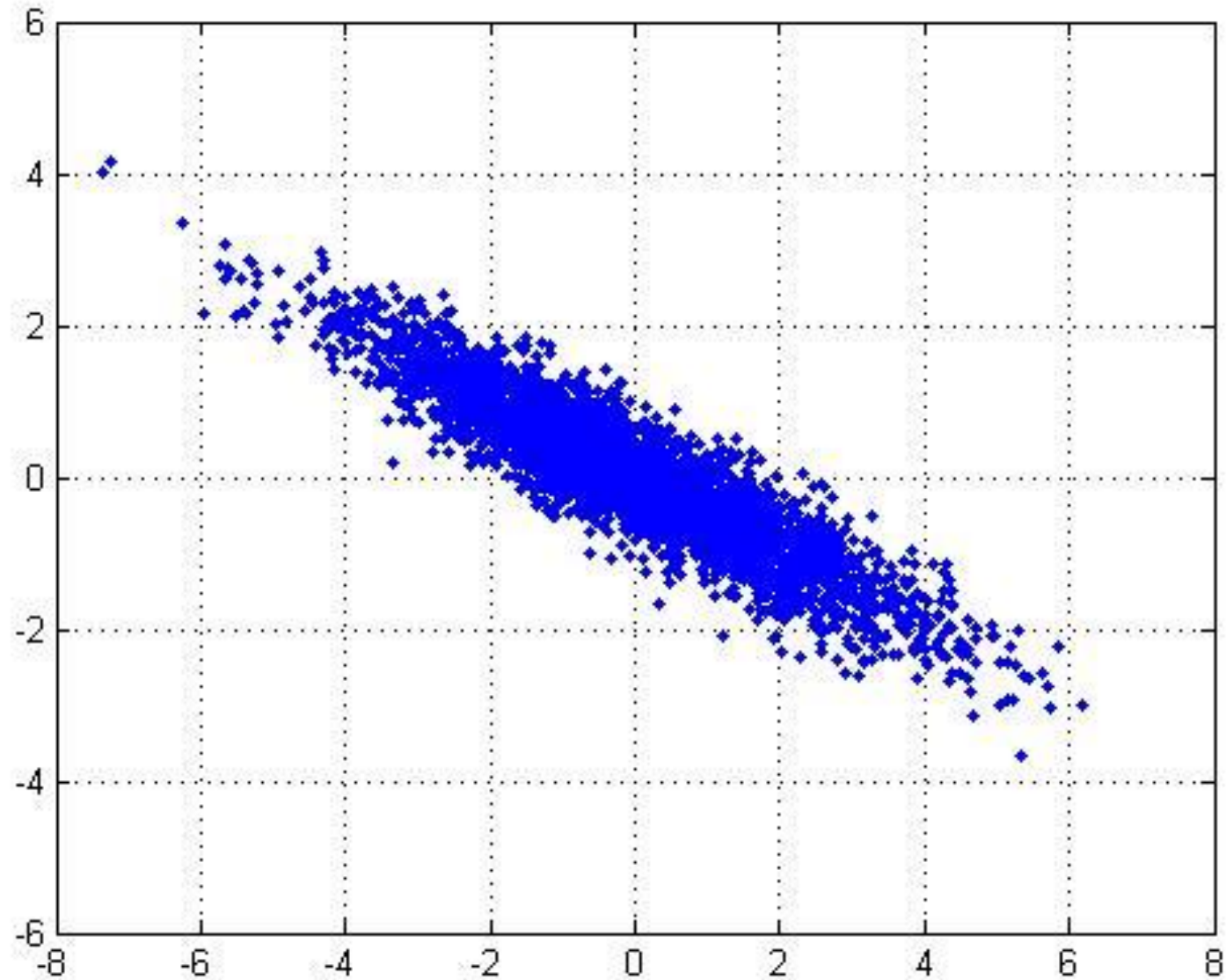
Principal Component Analysis (PCA)



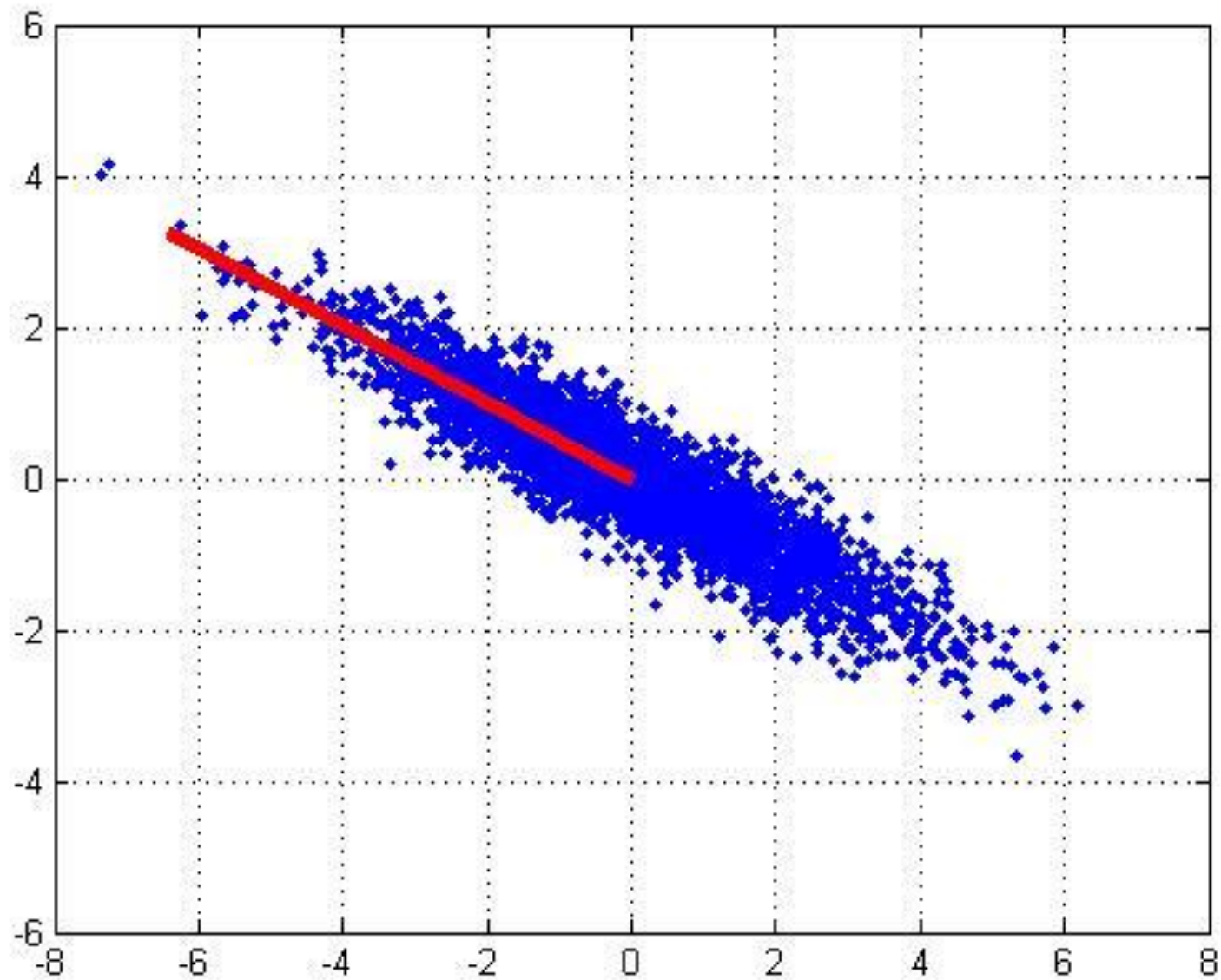
In case where data lies on or near a low d -dimensional linear subspace, axes of this subspace are an effective representation of the data.

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

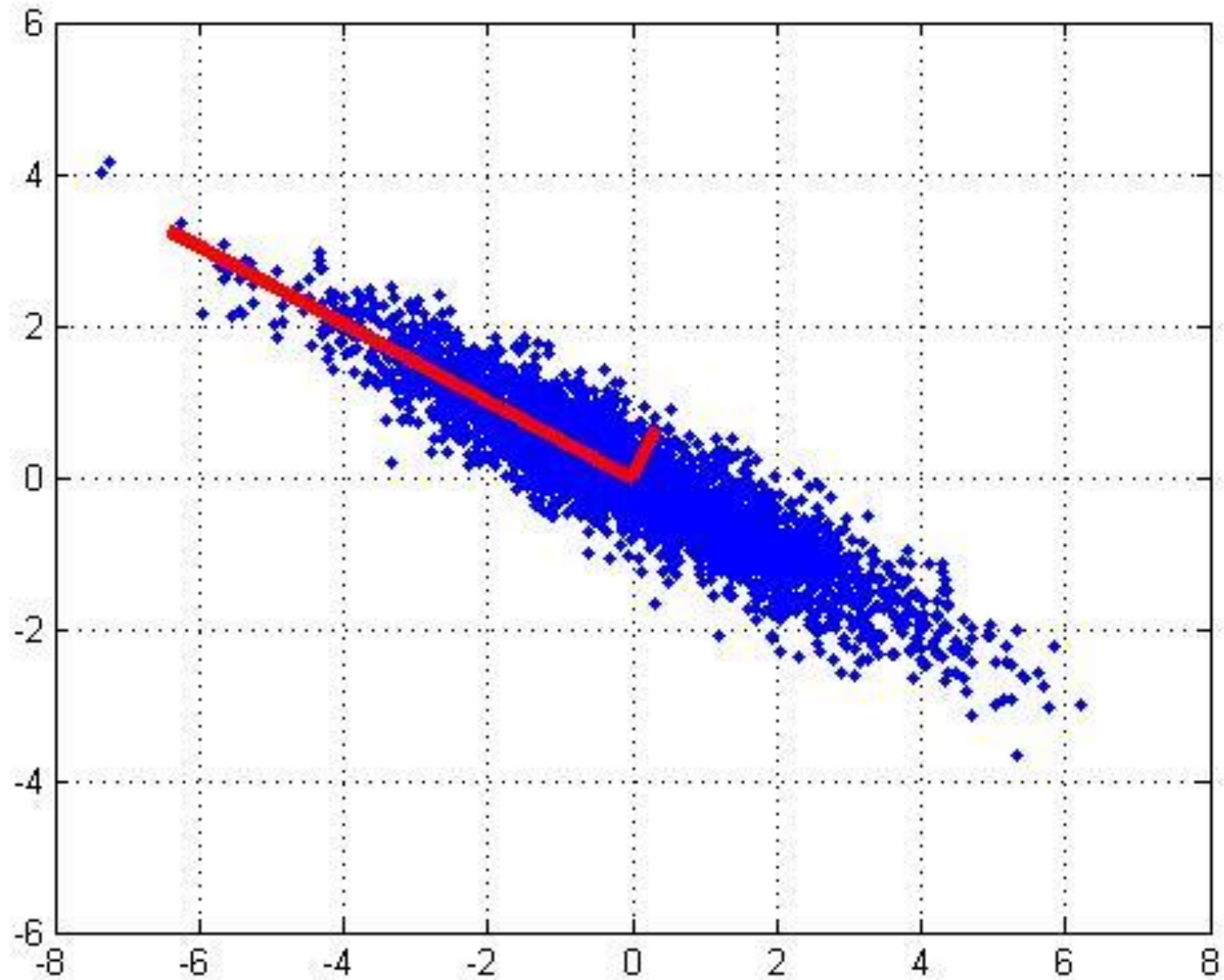
2D Gaussian dataset



1st PCA axis



2nd PCA axis



PCA Axes

Growth Plate Imaging

Growth Plate Disruption and Limb Length Discrepancy



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

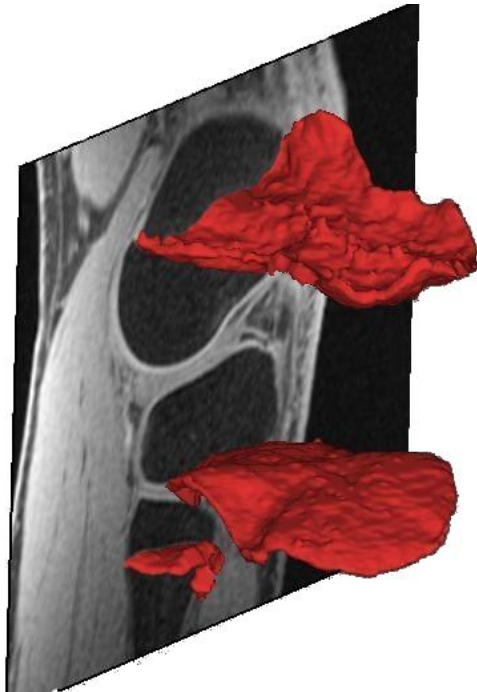


Images Courtesy
H. Potter, H.S.S.

Growth Plate Imaging

Growth Plate Disruption and Limb Length Discrepancy

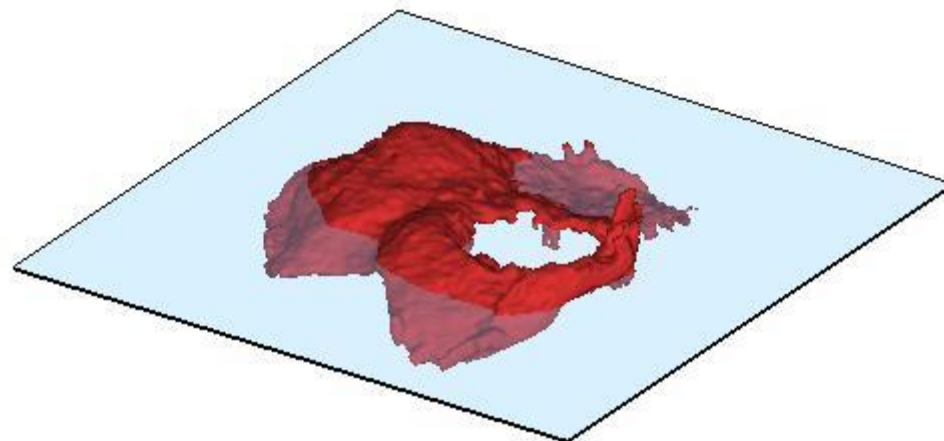
8 year-old boy with previous fracture and
4cm leg length discrepancy



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

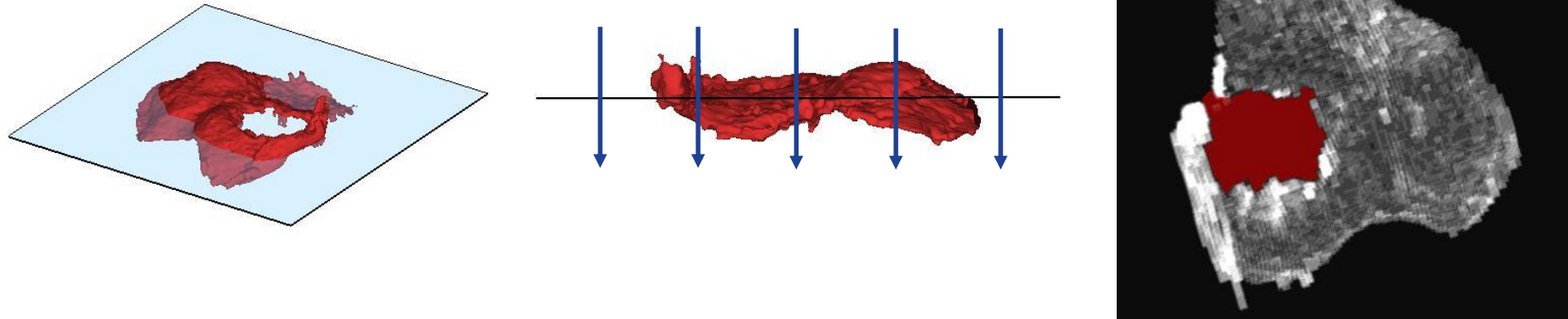
Growth Plate Imaging

Area Measurement



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

Sample Covariance Matrix

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:

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

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

Piazza Poll 1

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

A. $\mathbf{vx}$

B. $\mathbf{v}^T \mathbf{x}$

C. $(\mathbf{v}^T \mathbf{x})\mathbf{v}$

D. $\mathbf{v}^T \mathbf{x} \mathbf{x}^T \mathbf{v}$

Principle Component Analysis (PCA)

- Sketch of PCA algorithm
- Two PCA objective functions

Rotation of Data (and back)

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

Sketch of PCA

1. Randomly select $V \in \mathbb{R}^{K \times M}$
2. Project down: $\mathbf{z}^{(i)} = V\mathbf{x}^{(i)} \quad \forall i$
3. Reconstruct up: $\mathbf{x}^{(i)} = V^T \mathbf{z}^{(i)}$

Sketch of PCA

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

Definition of PCA

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

Select “Best” Vector

Reconstruction Error vs Variance of Projection