CS 540 Introduction to Artificial Intelligence

Linear Algebra & PCA

University of Wisconsin-Madison

Spring 2025

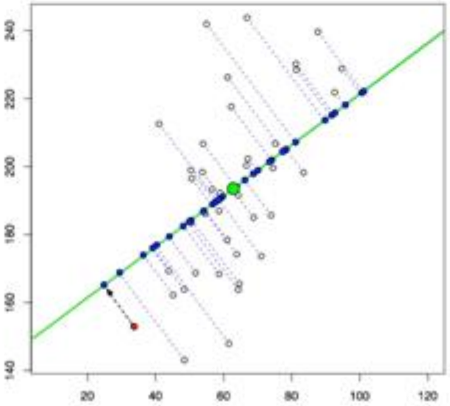
Announcements

- **HW 1 online:**
  - Writing assignment---nothing too stressful
  - Deadline **Tuesday Feb. 4th 11:59PM**
- **HW 2**:
  - Probability
  - Deadline **Thursday Feb. 6th 11:59PM**
- Class roadmap:

Linear Algebra and PCA	Mostly Foundations
Logic	Mostly Foundations
NLP	
Machine Learning: Introduction	
Machine Learning: Unsupervised Learning I	

Outline

- Basics: vectors, matrices, operations
- Dimensionality reduction
- Principal Components Analysis (PCA)



Lior Pachter

Matrix Inverses

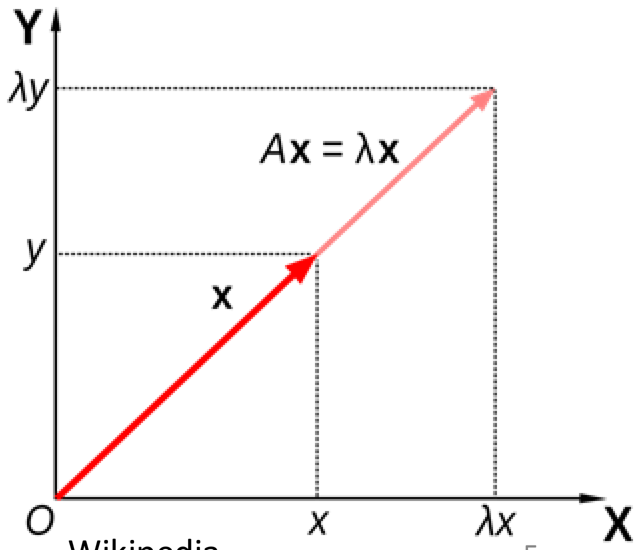
- If for *A* there is a *B* such that $AB = BA = I$
  - Then *A* is invertible/nonsingular, B is its inverse
  - Some matrices are **not** invertible!

  - Usual notation: A^{-1}

$$\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} = I$$

Eigenvalues & Eigenvectors

- For a square matrix *A*, solutions to $Av = \lambda v$
  - *v* (nonzero) is a vector: **eigenvector**
  - λ is a scalar: **eigenvalue**
  - Intuition: A is a linear transformation;
  - Can stretch/rotate vectors;
  - E-vectors: only stretched (by e-vals)

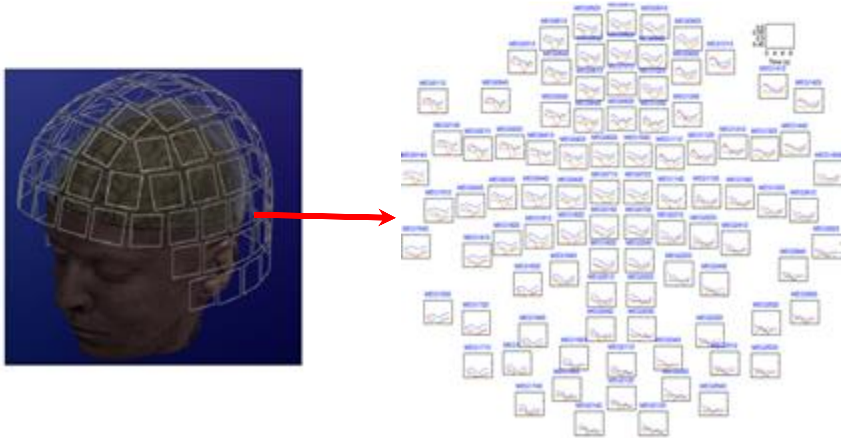


Wikipedia

Dimensionality Reduction

- Vectors store features. Lots of features!
  - Document classification: thousands of words per doc
  - Netflix surveys: 480189 users x 17770 movies
  - MEG Brain Imaging: 120 locations x 500 time points x 20 objects

	movie 1	movie 2	movie 3
Tom	5	?	?
George	?	?	3
Susan	4	3	1
Beth	4	3	?



Dimensionality Reduction

Reduce dimensions

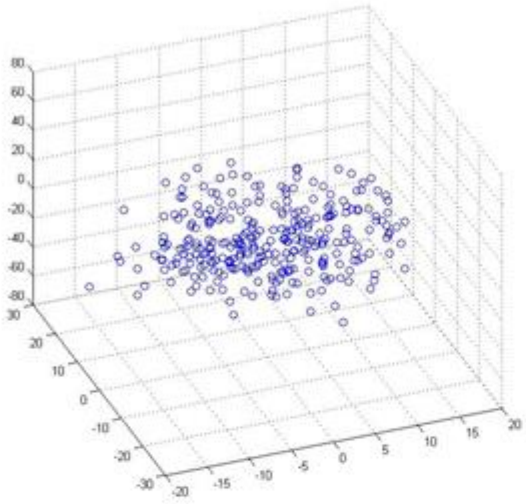
- Why?
  - Lots of features redundant
  - Storage & computation costs

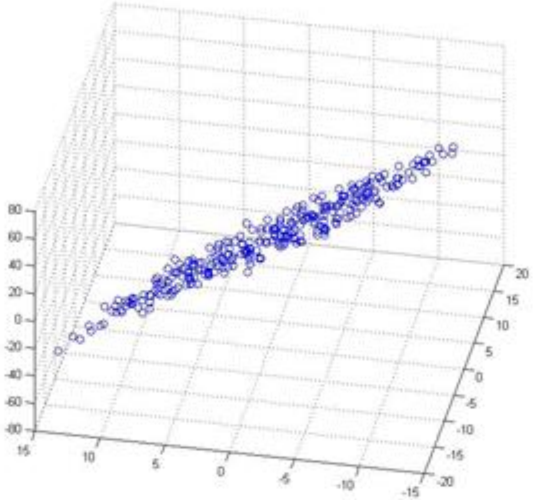
CreativeBlog

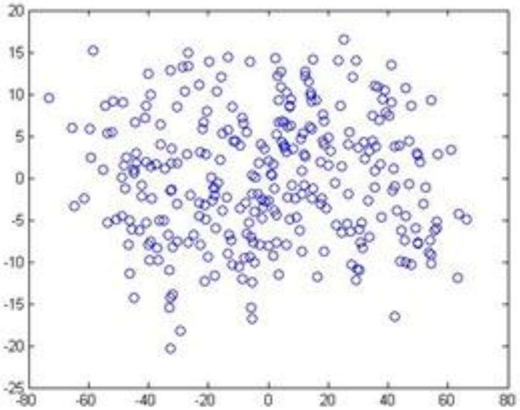
- Goal: take $x \in \mathbb{R}^d \rightarrow x \in \mathbb{R}^r$ for $r << d$
  - But, minimize information loss

Dimensionality Reduction

Examples: 3D to 2D







Andrew Ng

Break & Quiz

Q 2.1: What is the inverse of $A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$

A: $A^{-1} = \begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix}$

B: $A^{-1} = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{2} & 0 \end{bmatrix}$

C: Undefined / *A* is not invertible

Break & Quiz

Q 2.1: What is the inverse of $A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$

A: $A^{-1} = \begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix}$

B: $A^{-1} = \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{2} & 0 \end{bmatrix}$

C: Undefined / A is not invertible

$$AA^{-1} = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 * a + c * 2 & 0 * b + 2 * d \\ 3 * a + c * 0 & 3 * b + 0 * d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$2c = 1$$
$$3a = 0$$
$$2d = 0$$
$$3b = 1$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 1/3 \\ 1/2 & 0 \end{bmatrix}$$

Break & Quiz

Q 2.2: What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

A. -1, 2, 4
B. 0.5, 0.2, 1.0
C. 0, 2, 5
D. 2, 5, 1

Break & Quiz

Q 2.2: What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

A. -1, 2, 4
B. 0.5, 0.2, 1.0
C. 0, 2, 5
D. 2, 5, 1

Break & Quiz

Q 2.2: What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

A. -1, 2, 4
B. 0.5, 0.2, 1.0
C. 0, 2, 5
D. 2, 5, 1

Solution #1: You may recall from a linear algebra course that the eigenvalues of a diagonal matrix (in which only diagonal entries are non-zero) are just the entries along the diagonal. Hence D is the correct answer.

Break & Quiz

Q 2.2: What are the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

A. -1, 2, 4
B. 0.5, 0.2, 1.0
C. 0, 2, 5
D. 2, 5, 1

Solution #2: Use the definition of eigenvectors and values: $Av = \lambda v$

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} v_1 \\ v_2 \\ v_3 \end{matrix} = \begin{bmatrix} 2v_1 + 0v_2 + 0v_3 \\ 0v_1 + 5v_2 + 0v_3 \\ 0v_1 + 0v_2 + 1v_3 \end{bmatrix} = \begin{bmatrix} 2v_1 \\ 5v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} \lambda v_1 \\ \lambda v_2 \\ \lambda v_3 \end{bmatrix}$$

Since A is a 3x3 matrix, A has 3 eigenvalues and so there are 3 combinations of values for λ and v that will satisfy the above equation. The simple form of the equations suggests starting by checking each of the standard basis vectors* as v and then solving for λ. Doing so gives D as the correct answer.

*Each standard basis vector $e_i \in \mathbb{R}^n$ is the vector in which all components are zero except component i is 1.

Break & Quiz

Q 2.3: Suppose we are given a dataset with n=10000 samples with 100-dimensional binary feature vectors. Our storage device has a capacity of 50000 bits. What's the lowest compression ratio we can use?

A. 20X
B. 100X
C. 5X
D. 1X

Break & Quiz

Q 2.3: Suppose we are given a dataset with *n*=10000 samples with 100-dimensional binary feature vectors. Our storage device has a capacity of 50000 bits. What's the lower compression ratio we can use?

A. **20X**
B. 100X
C. 5X
D. 1X

Break & Quiz

Q 2.3: Suppose we are given a dataset with n=10000 samples with 100-dimensional binary feature vectors. Our storage device has a capacity of 50000 bits. What's the lower compression ratio we can use?

A. **20X**
B. 100X
C. 5X
D. 1X

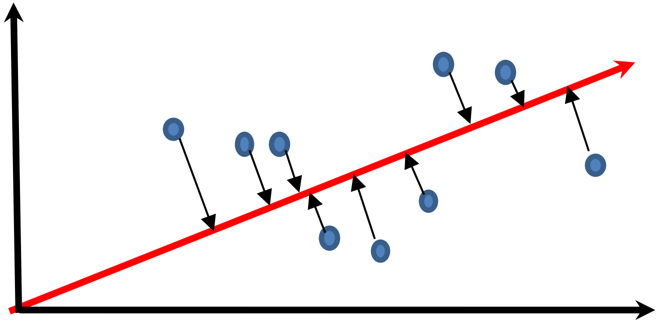
50,000 bits / 10,000 samples means compressed version must have 5 bits / sample.

Dataset has 100 bits / sample.

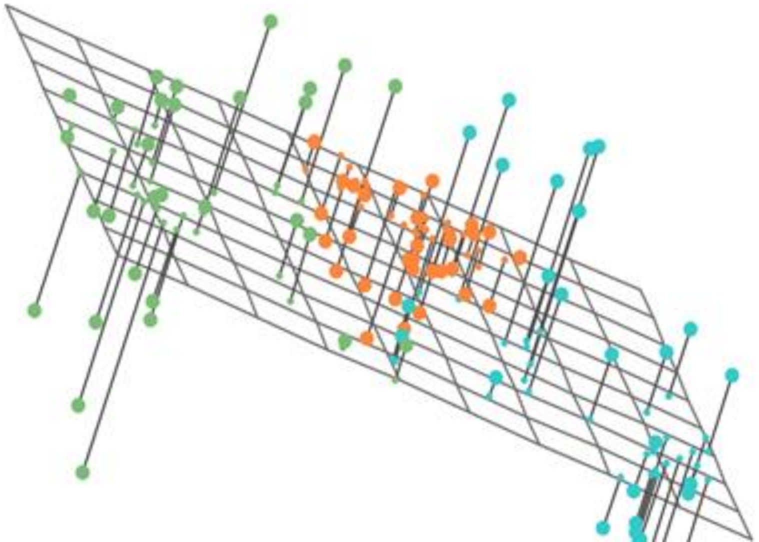
Must compress 20x smaller to fit on device.

Principal Components Analysis (**PCA**)

- A type of dimensionality reduction approach
    - For when data is **approximately lower dimensional**



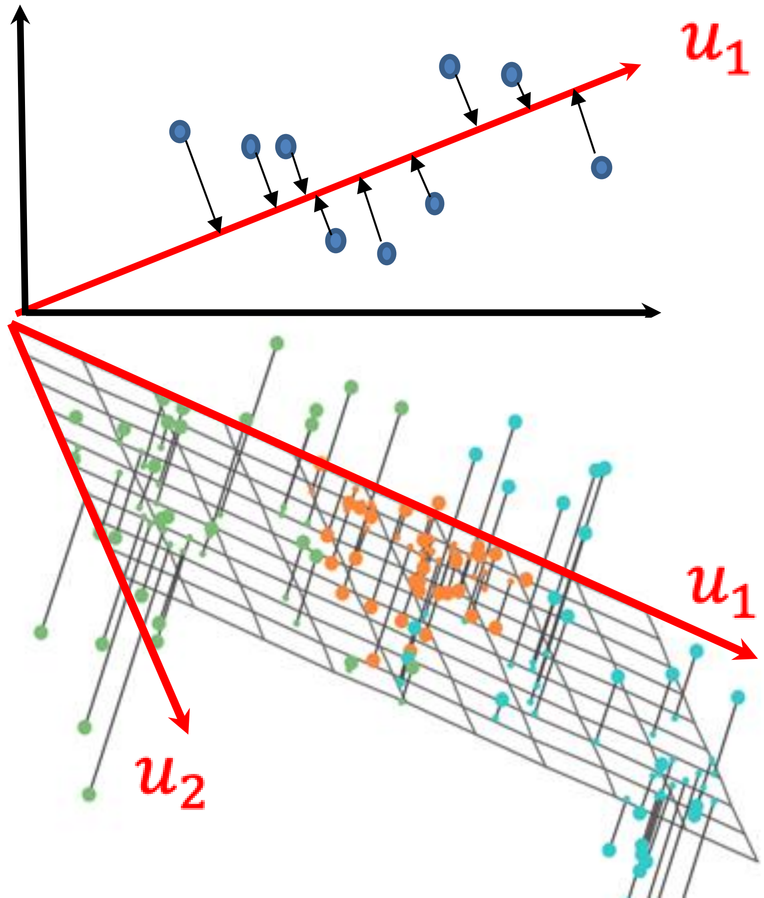
2D
↓
1D



3D
↓
2D

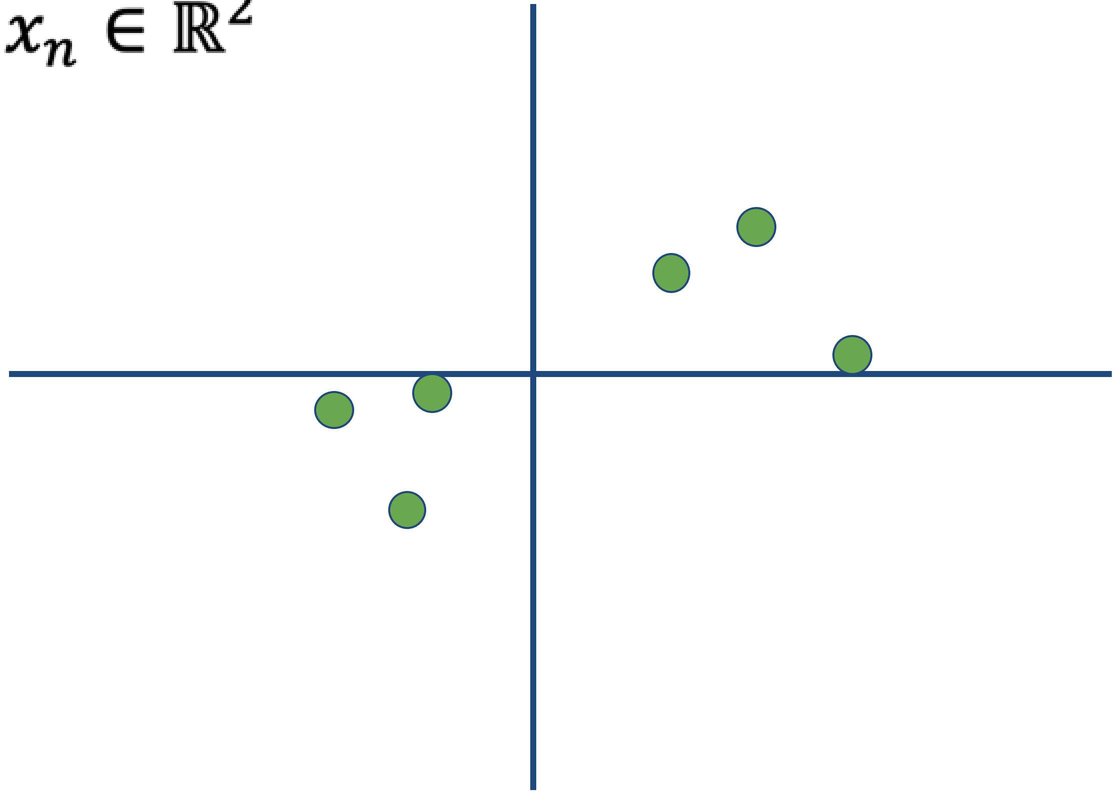
Principal Components Analysis (**PCA**)

- Find **axes** $u_1, u_2, \ldots, u_m \in \mathbb{R}^d$ of a subspace
  - Will project to this subspace
- Want to preserve data
  - minimize projection error
- These vectors are the **principal components**



Projection: An Example

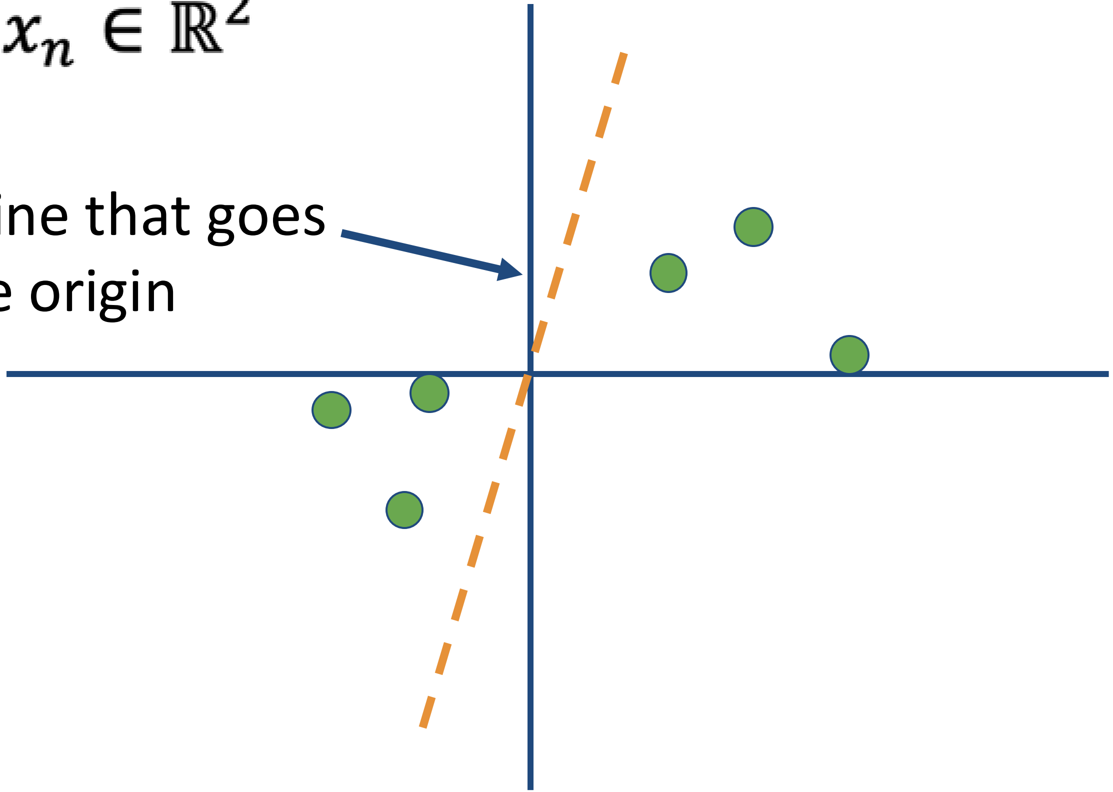
$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$



Projection: An Example

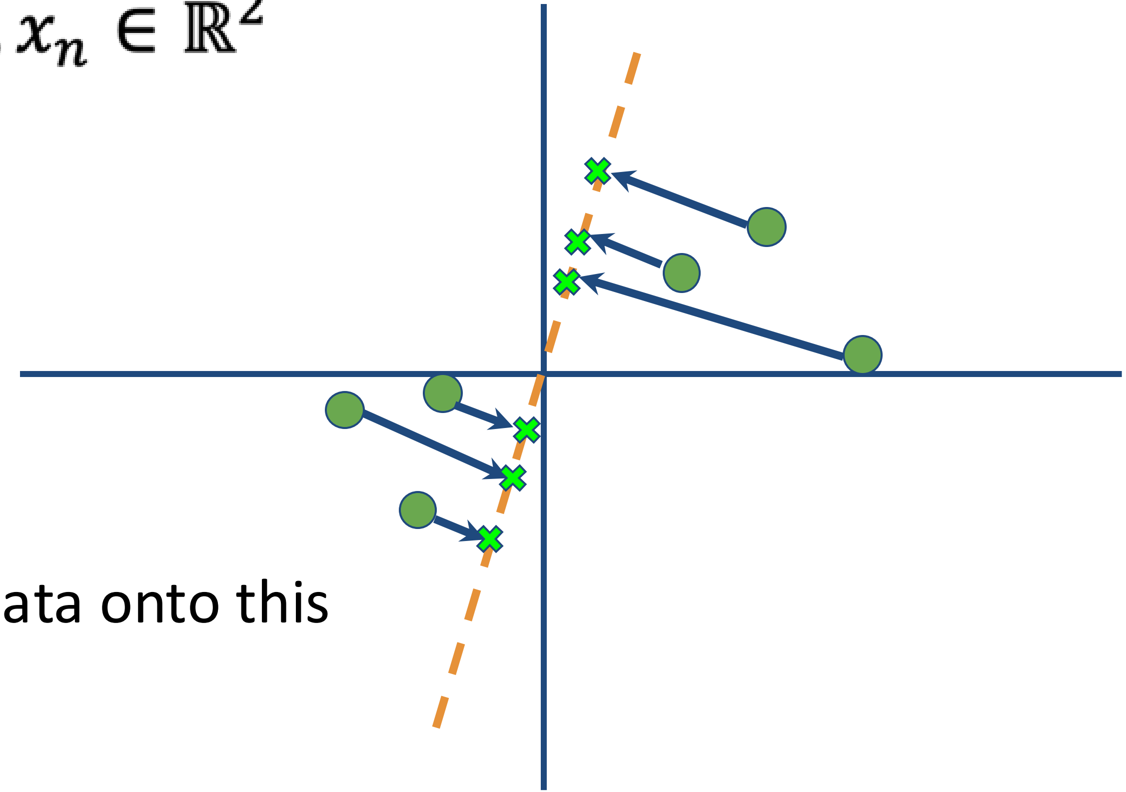
$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$

A random line that goes through the origin



Projection: An Example

$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$

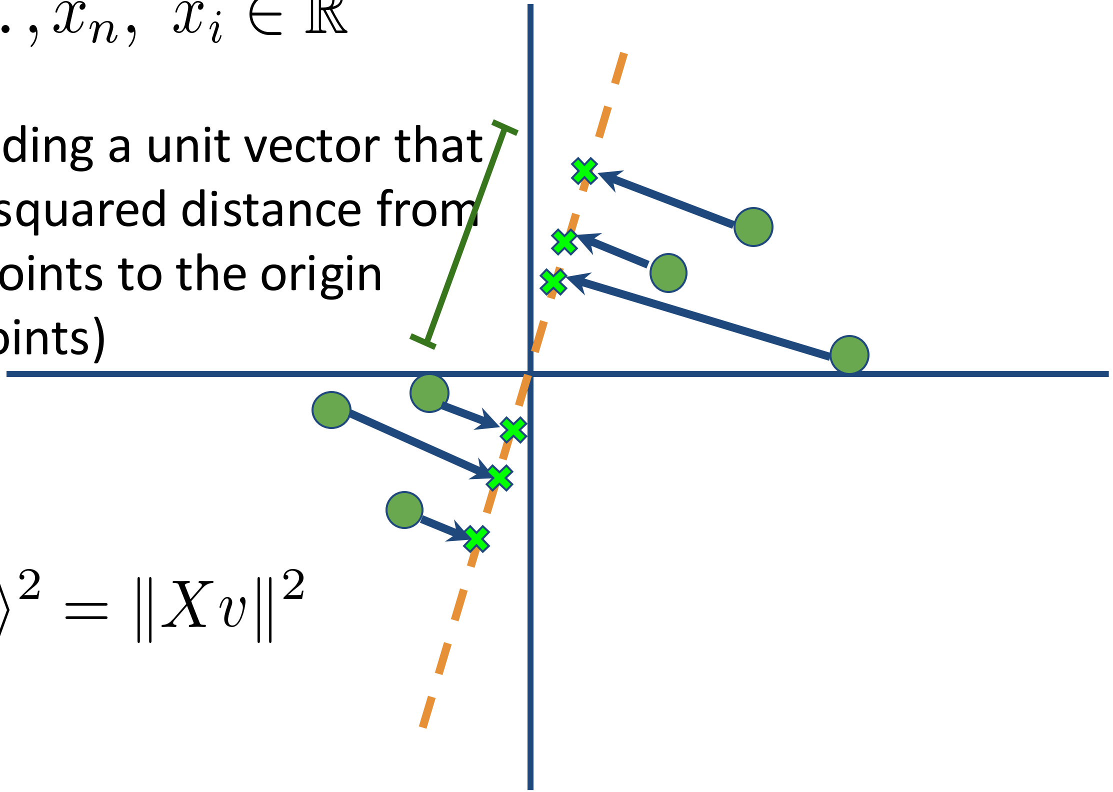


projects data onto this line

Projection: An Example

$x_1, x_2, \ldots, x_n, \;\; x_i \in \mathbb{R}^2$

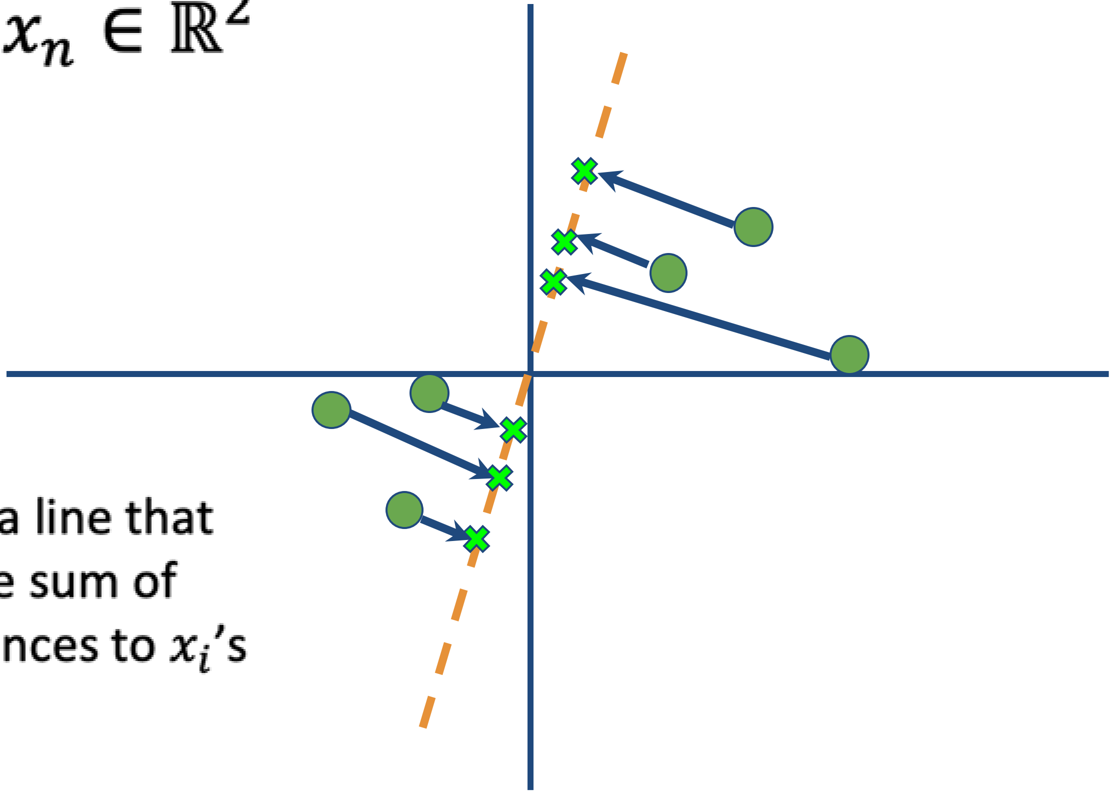
Goal of PCA: finding a unit vector that **maximizes** the squared distance from the projected points to the origin (sum over all points)



$$\sum_{i=1}^{n} \langle x_i, v \rangle^2 = \|Xv\|^2$$

Projection: An Example

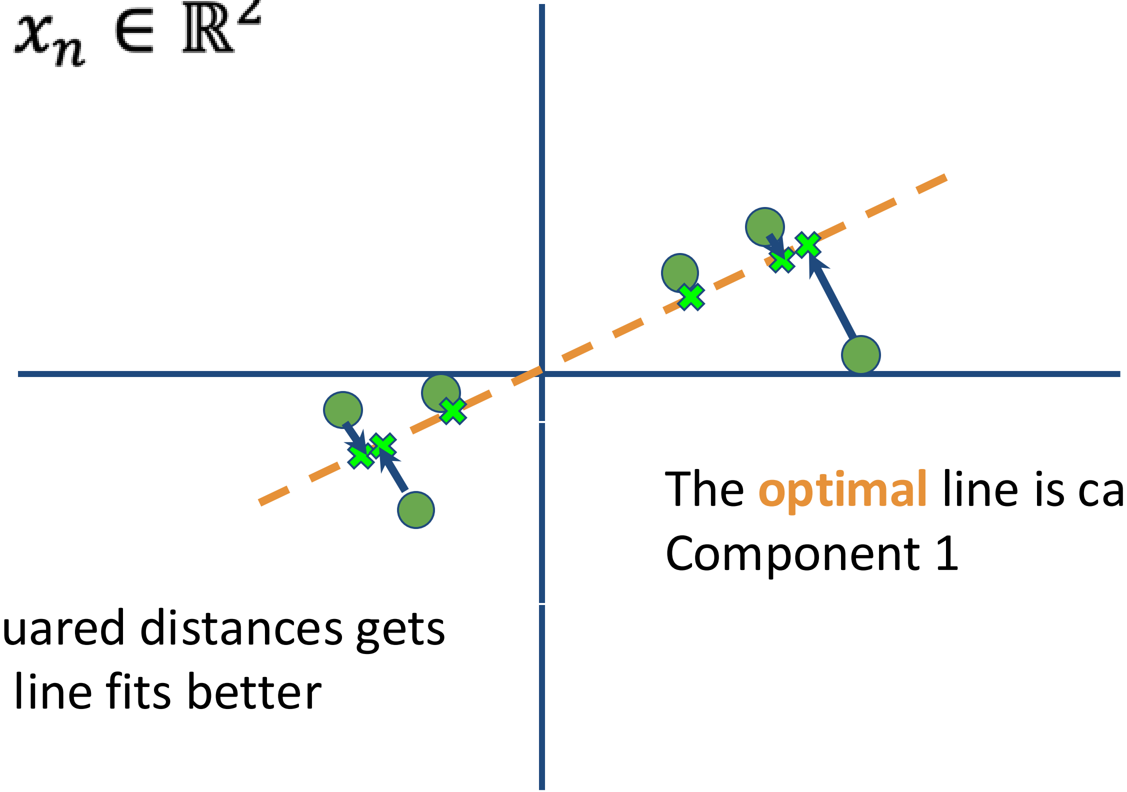
$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$



Goal: finding a line that **minimizes** the sum of squared distances to x_i's

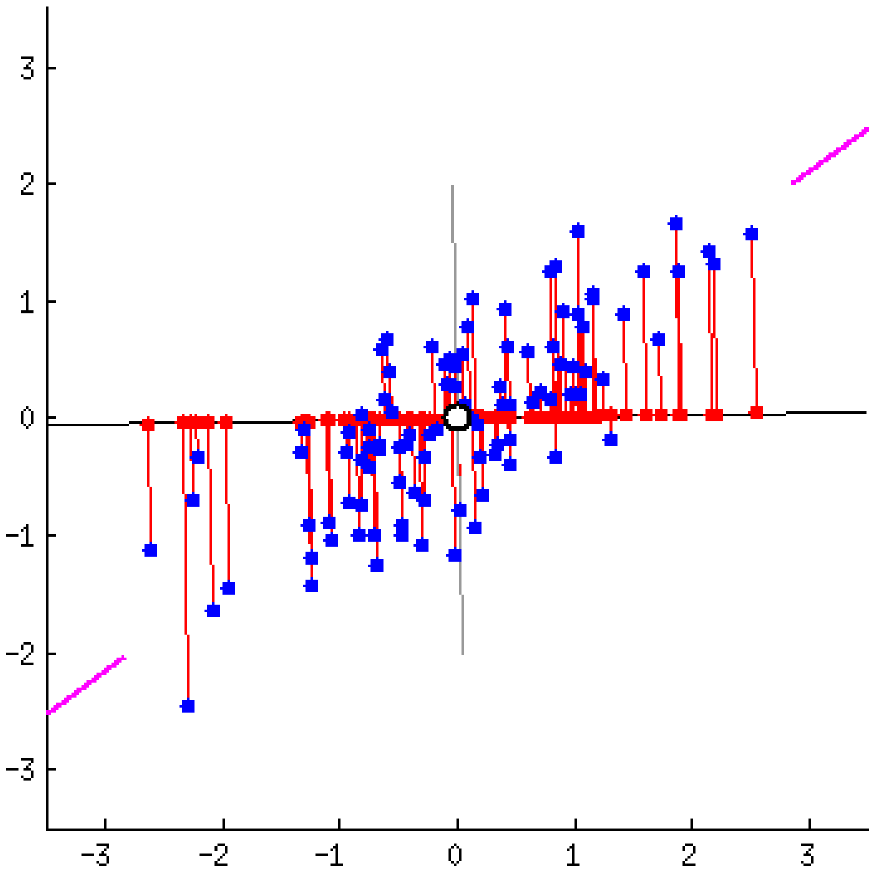
Projection: An Example

$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$



The **optimal** line is called Principal Component 1

The sum of squared distances gets smaller as the line fits better



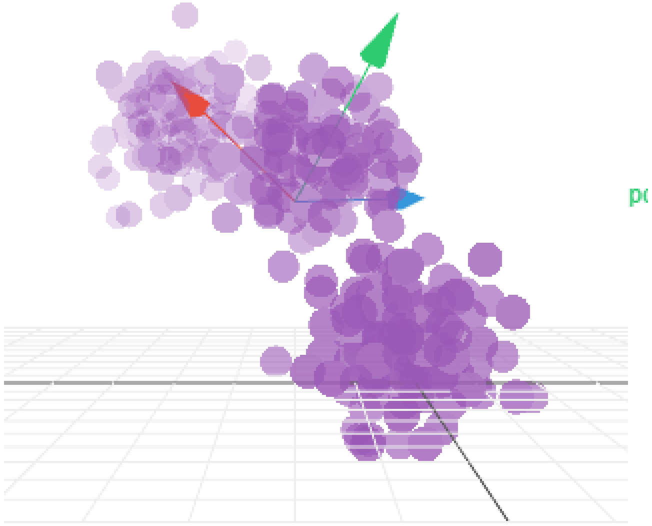
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

Summary: PCA Procedure

Inputs: data $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$

– **Center data so that $\frac{1}{n}\sum_{i=1}^{n} x_i = 0$**

Subtract the mean vector from each data



Victor Powell

Summary: PCA Procedure

First component,

$$v_1 = \arg\max_{\|v\|=1} \sum_{i=1}^{n} \langle v, x_i \rangle^2$$

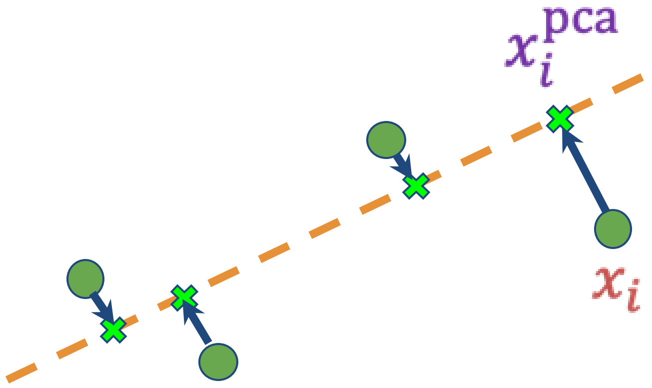
Same as getting

$$v_1 = \arg\max_{\|v\|=1} \|Xv\|^2$$

PCA Procedure

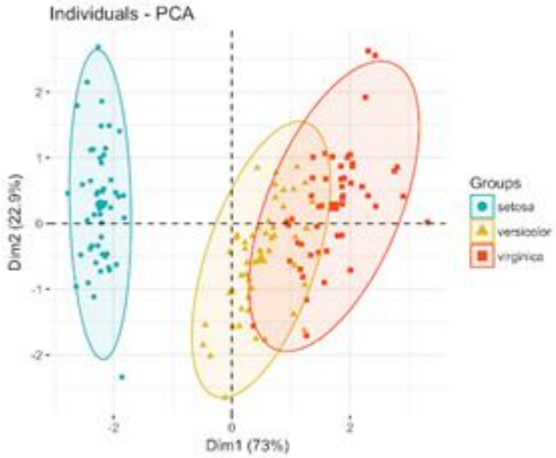
Output:

principal components $u_1, \ldots, u_m \in \mathbb{R}^d$



Many Variations

- PCA, Kernel PCA, ICA, CCA
  - Extract structure from high dimensional dataset
- Uses:
  - **Visualization**
  - Efficiency
  - Noise removal
  - Downstream machine learning use



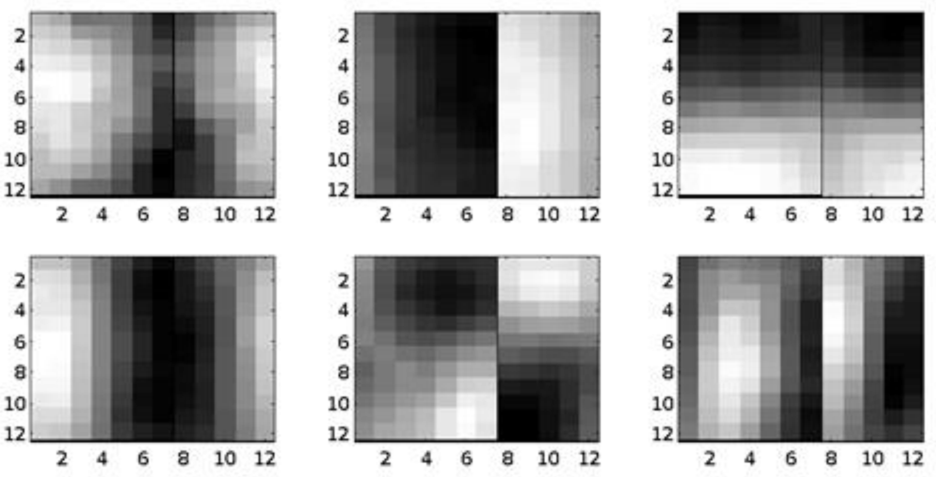
STHDA

Application: Image Compression

- Start with image; divide into 12x12 patches
  - That is, 144-D vector
  - **Original image:**

Application: Image Compression

- 6 principal components (as an image)



Application: Image Compression

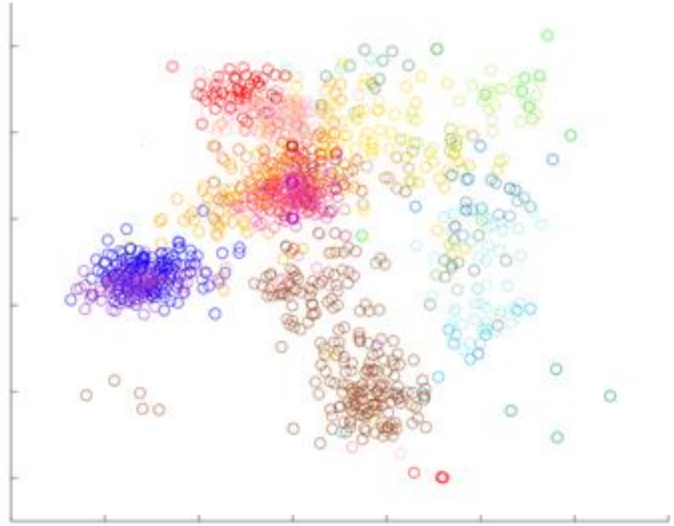
- Project to 6D

Compressed

Original

Application: Exploratory Data Analysis

- [**Novembre et al. ’08**]: Take top two singular vectors of people x SNP matrix (POPRES)



“Genes Mirror Geography in Europe”

Readings

- Vast literature on linear algebra.
- Local class: **Math 341**
- More on PCA (and other matrix methods in ML): **CS 532**

- **Suggested reading:**
  - Lecture notes on PCA by Roughgarden and Valiant
    https://web.stanford.edu/class/cs168/l/l7.pdf
  - 760 notes by Zhu https://pages.cs.wisc.edu/~jerryzhu/cs760/PCA.pdf

PCA Recursion (advanced material)

Once we have *k-1* components, next?

$$\hat{X}_k = X - \sum_{i=1}^{k-1} X v_i v_i^T$$

Deflation

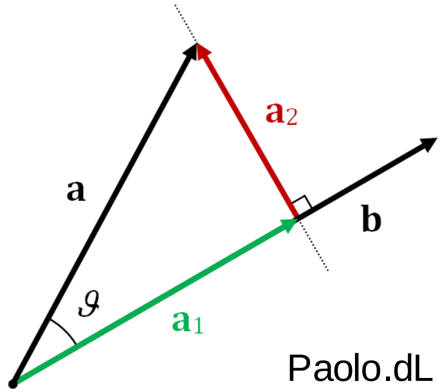
Then do the same thing

$$v_k = \arg\max_{\|v\|=1} \|\hat{X}_k w\|^2$$

To project *a* onto unit vector *b*,

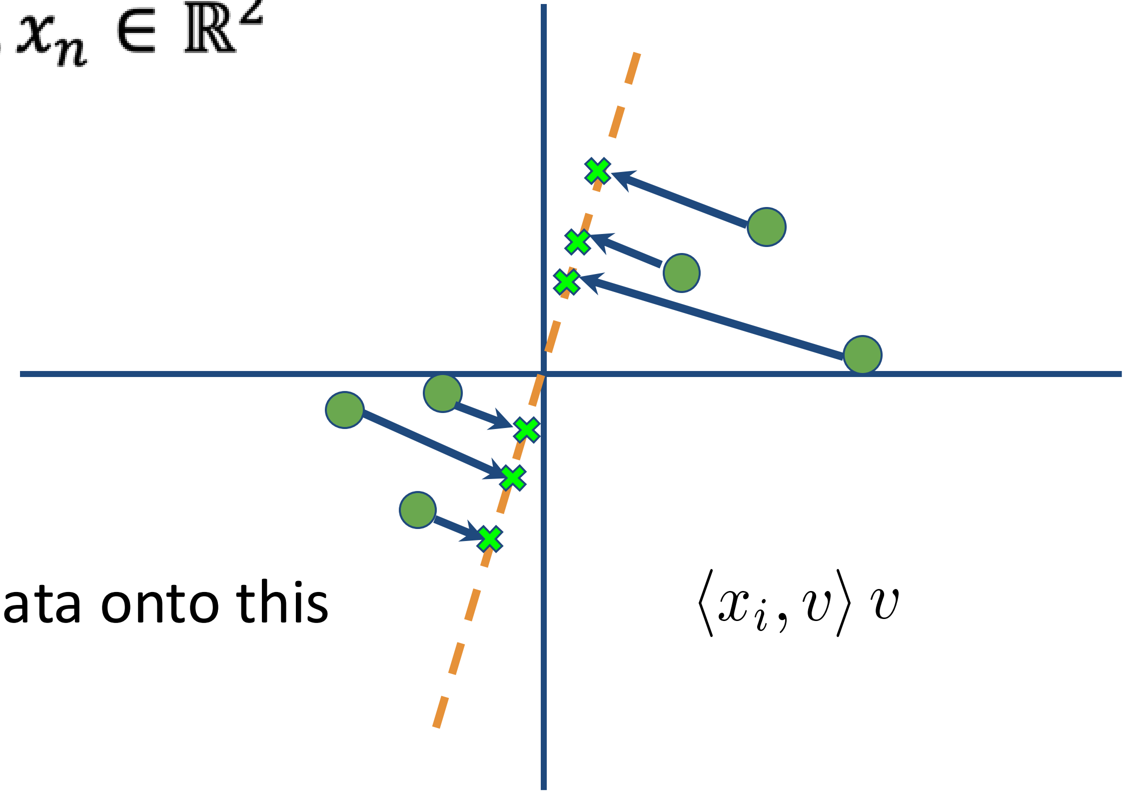
$$\langle a, b \rangle b \longleftarrow \text{Direction}$$

$\uparrow$ Length



Projection: An Example

$x_1, x_2, \ldots, x_n \in \mathbb{R}^2$



projects data onto this line

$\langle x_i, v \rangle \, v$