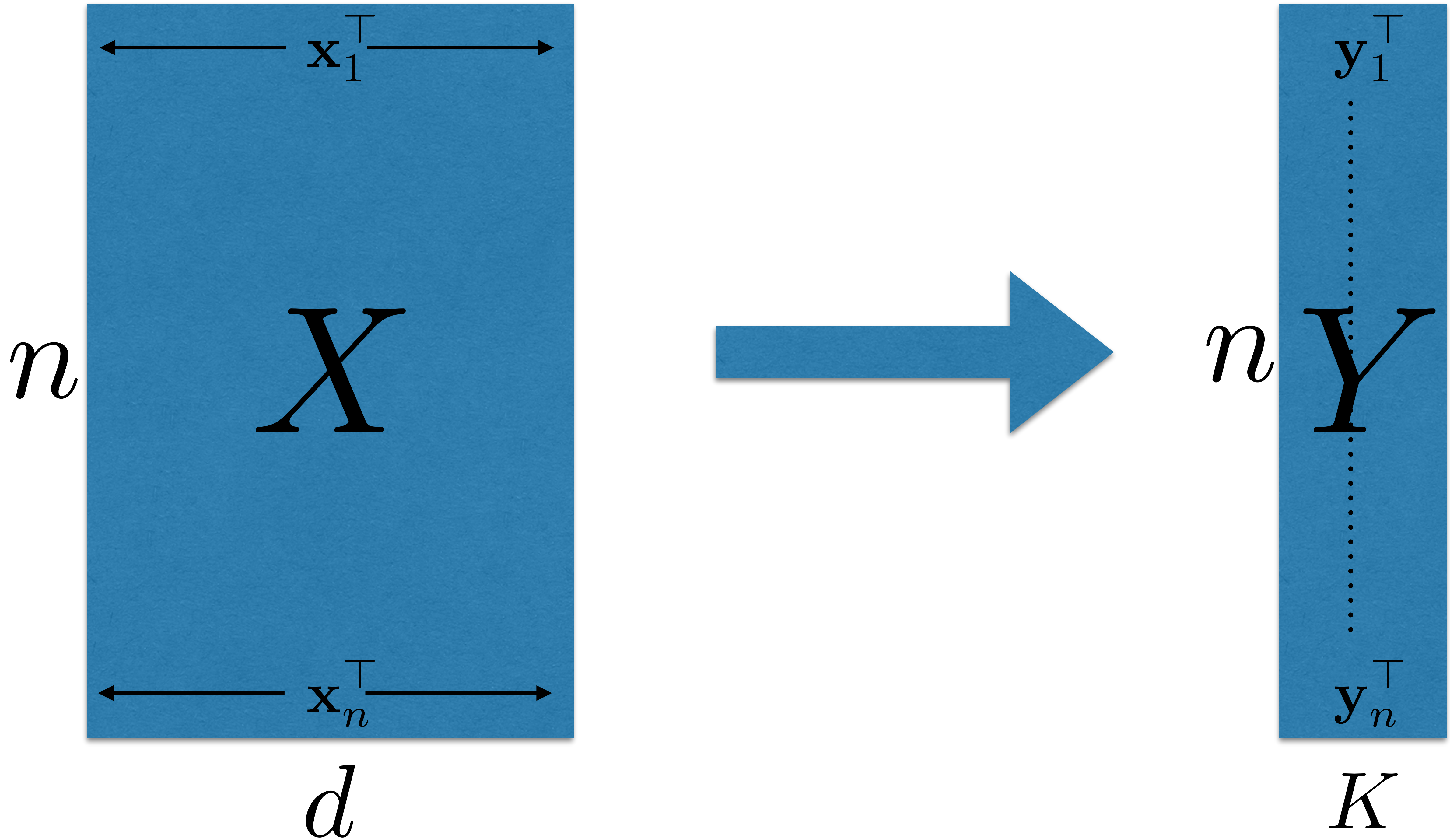
CS 3/5780

Lec 6:
Principle Component Analysi

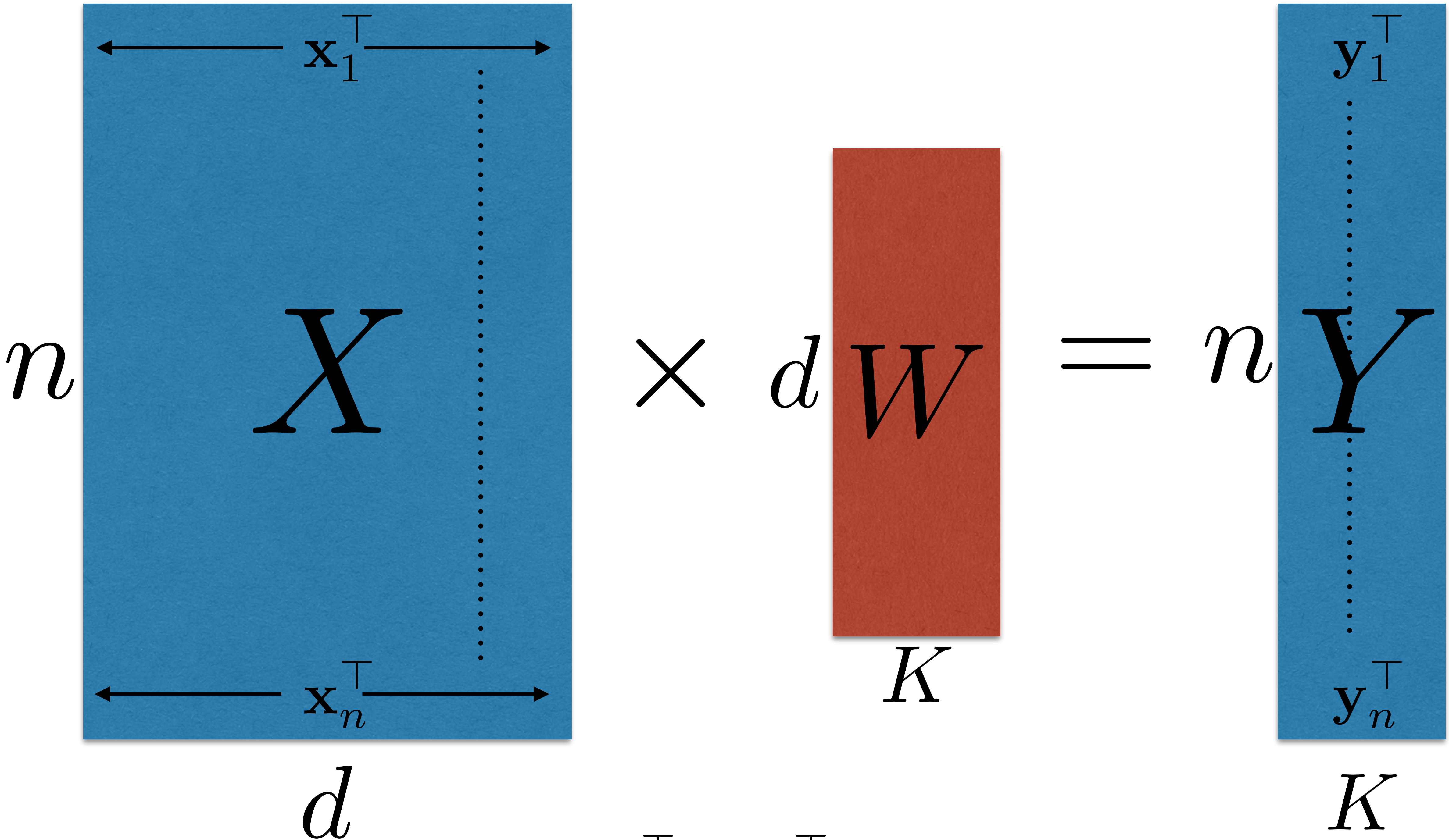
Dimensionality Reduction



Dimensionality Reduction

Given feature vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$, compress the data points into low dimensional representation $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^K$ where $K << d$

Dim Reduction: Linear Transformation



$$\mathbf{y}_i^\top = \mathbf{x}_i^\top W$$

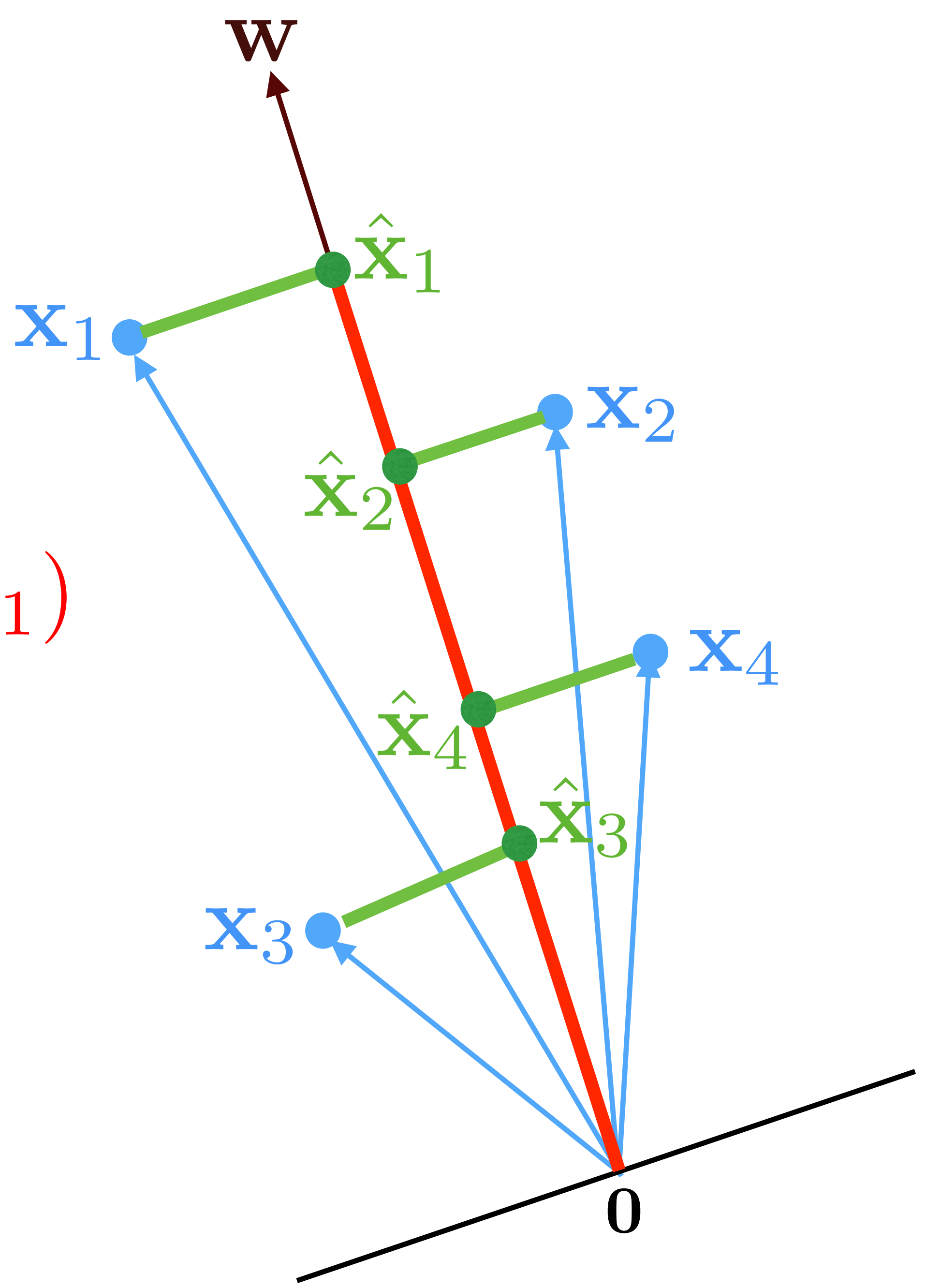
Prelude: Reducing to 1 Dim

- W is a d x 1 matrix (d dimensional vector)
- Each data point is compressed to a single number
- How do we pick this W?

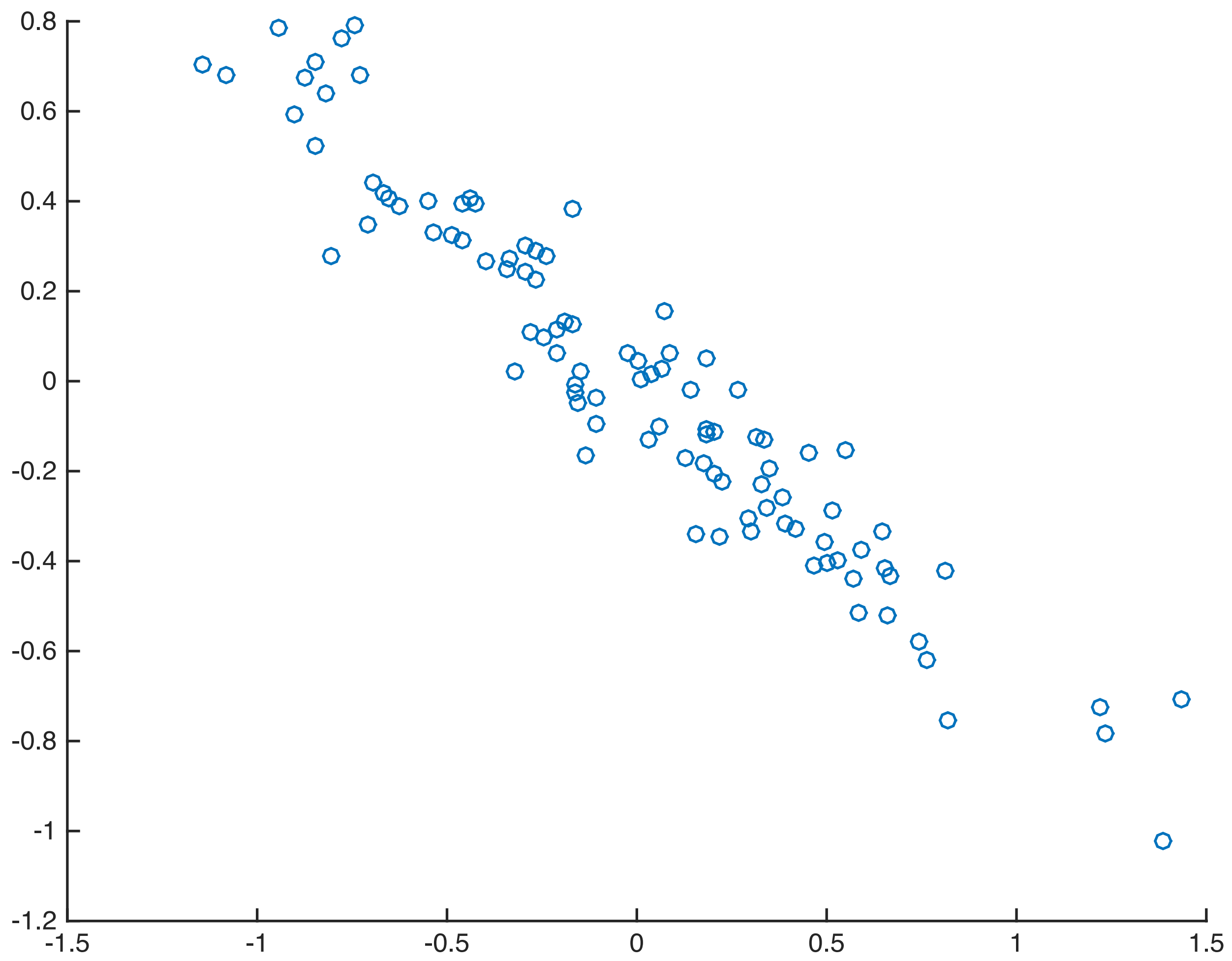
Dim Reduction: Linear Transformation

Prelude: reducing to 1 dimension

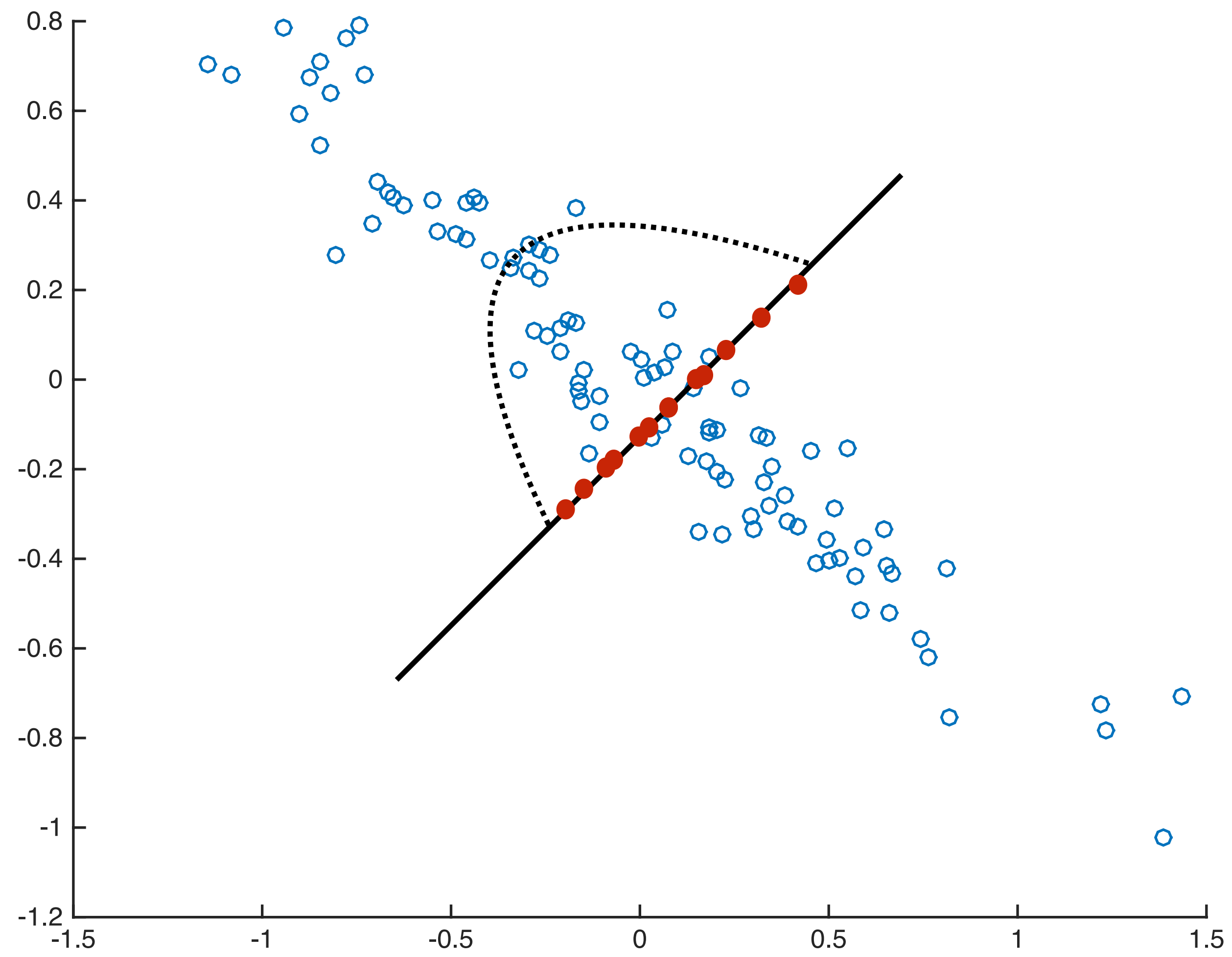
$$\mathbf{y}_1 = \mathbf{w}^\top \mathbf{x}_1 = \|\mathbf{x}_1\| \cos\left(\angle \mathbf{w}\mathbf{x}_1\right)$$



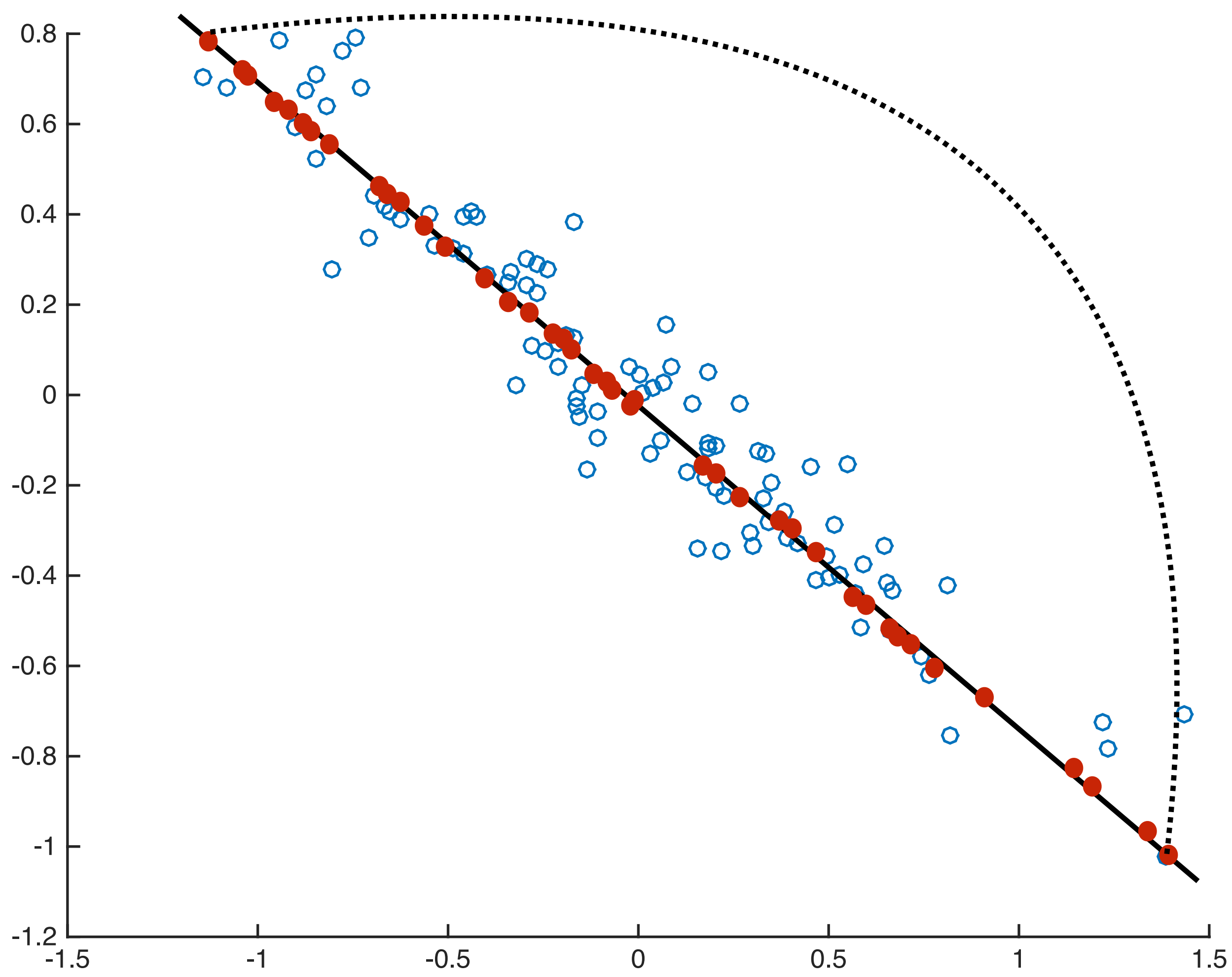
PCA: Variance Maximization



PCA: Variance Maximization



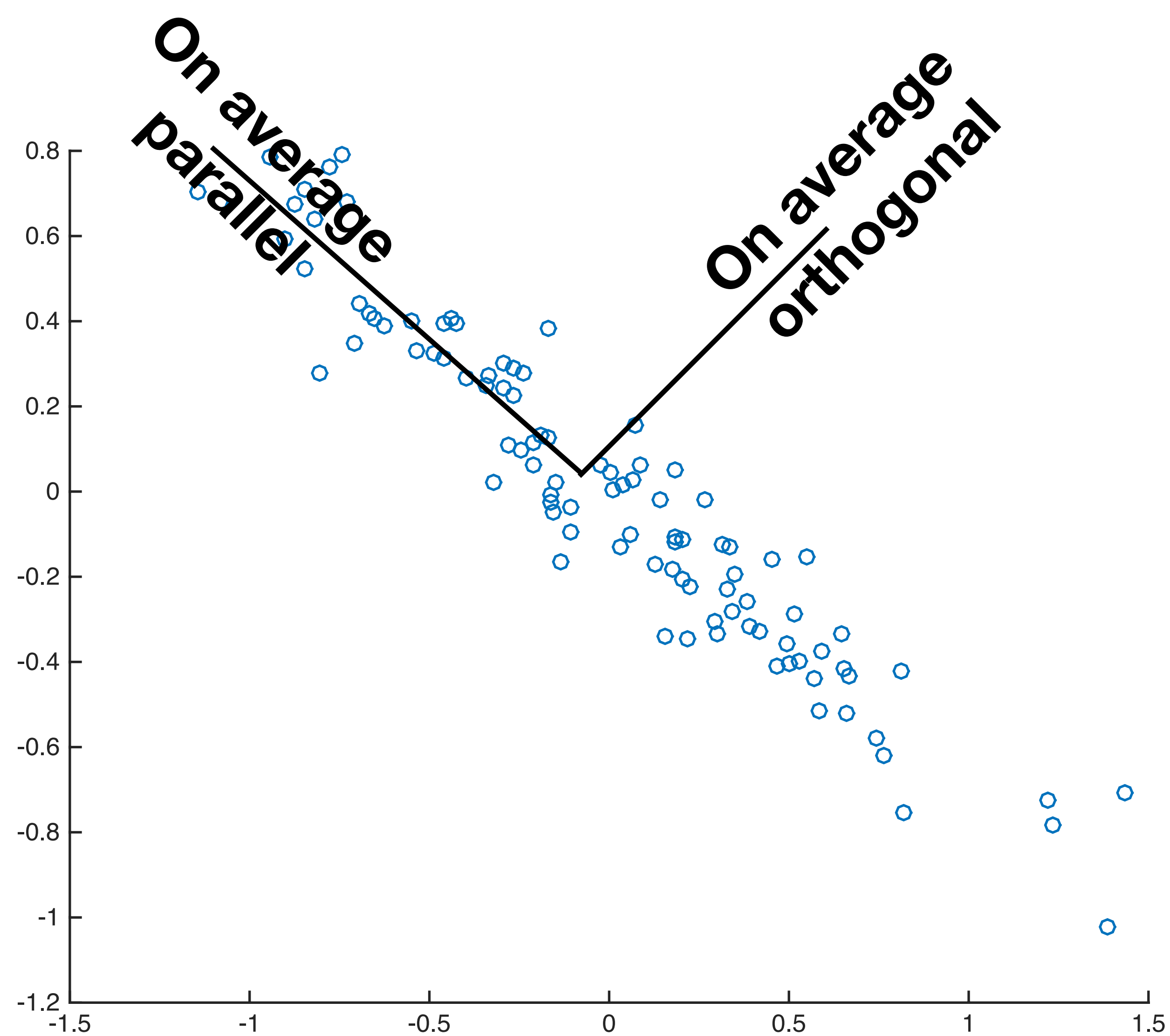
PCA: Variance Maximization



PCA: Variance Maximization

- Pick directions along which data varies the most

Which Direction?



$$\frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{w}^{\top}(\mathbf{x}_t-\mu)\right)^2=\frac{1}{n}\sum_{t=1}^{n}\|\mathbf{x}_t-\mu\|^2\text{cosine}(w,x_t-\mu)$$

PCA: Variance Maximization

- Pick directions along which data varies the most
- First principal component:

$$\mathbf{w}_1 = \arg\max_{\mathbf{w}:\|\mathbf{w}\|_2=1} \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{w}^\top \mathbf{x}_t - \frac{1}{n}\sum_{t=1}^{n}\mathbf{w}^\top \mathbf{x}_t\right)^2$$

$$= \arg\max_{\mathbf{w}:\|\mathbf{w}\|_2=1} \frac{1}{n}\sum_{t=1}^{n}\left(\mathbf{w}^\top (\mathbf{x}_t - \mu)\right)^2$$

PCA: Variance Maximization

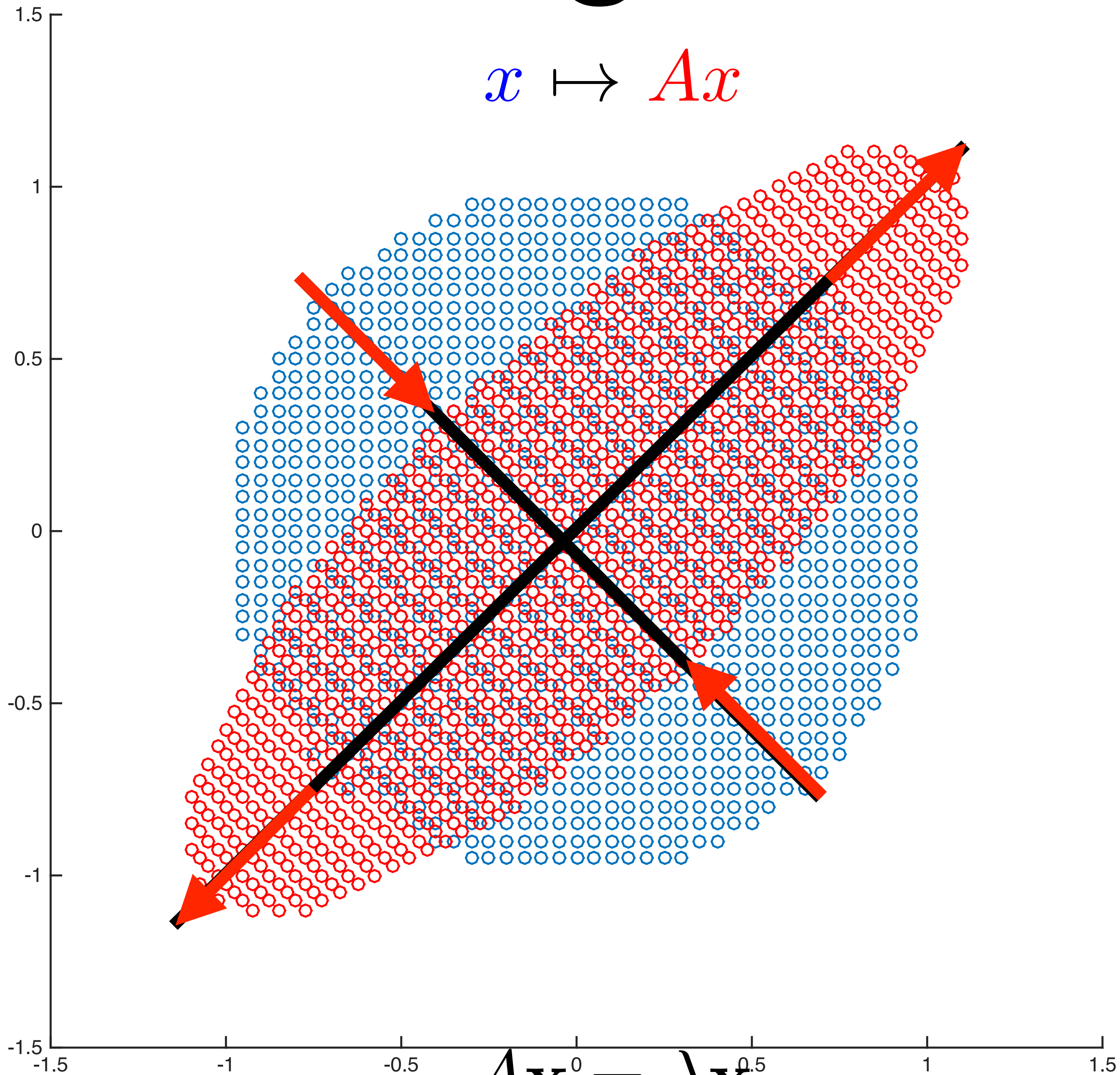
- Pick directions along which data varies the most
- First principal component:

$$\mathbf{w}_1 = \arg \max_{\mathbf{w}:\|\mathbf{w}\|_2=1} \mathbf{w}^\top \Sigma \mathbf{w}$$

Σ is the covariance matrix

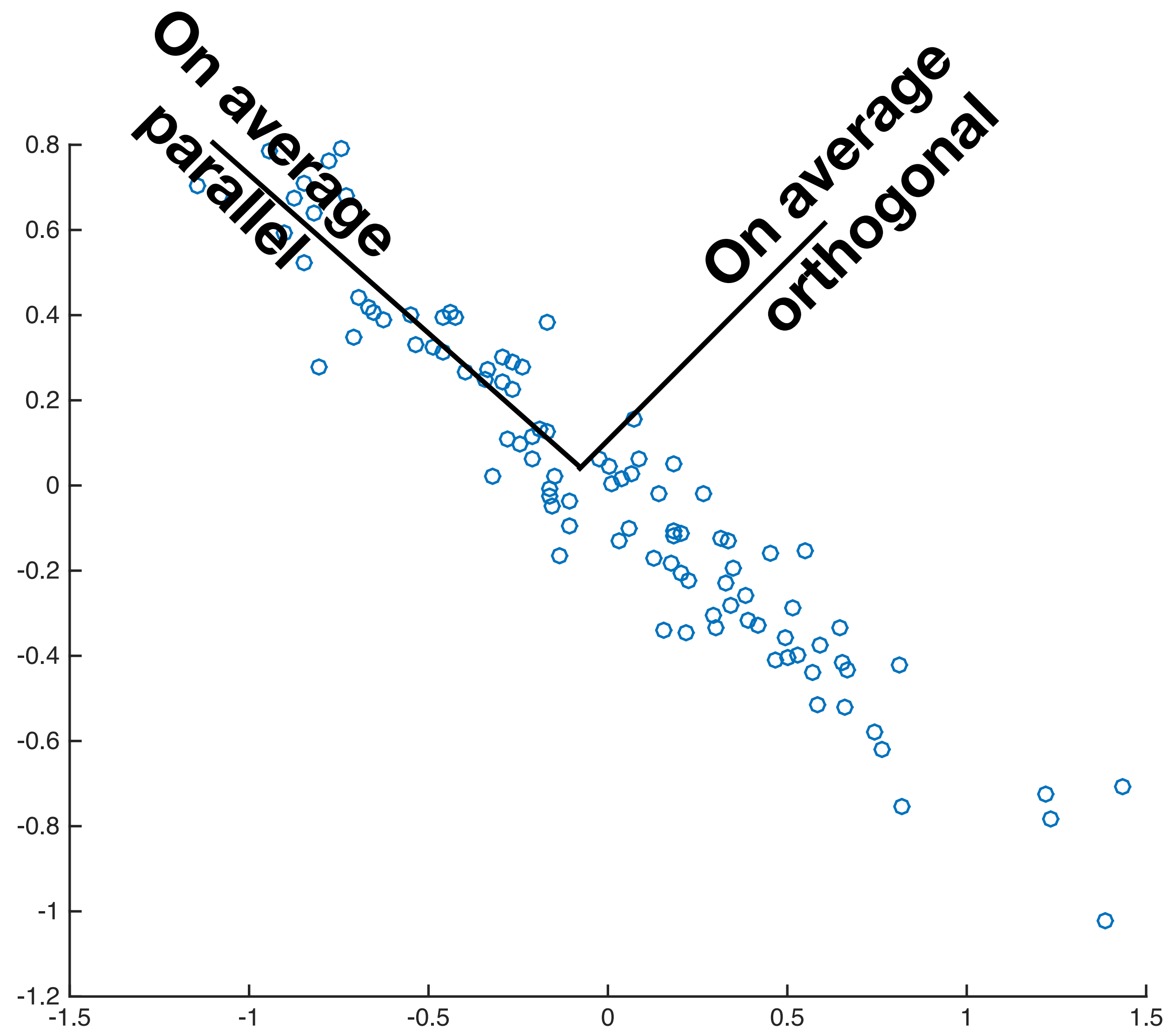
Solution: $\mathbf{w}_1 =$ Largest Eigenvector of Σ

What are Eigen Vectors?



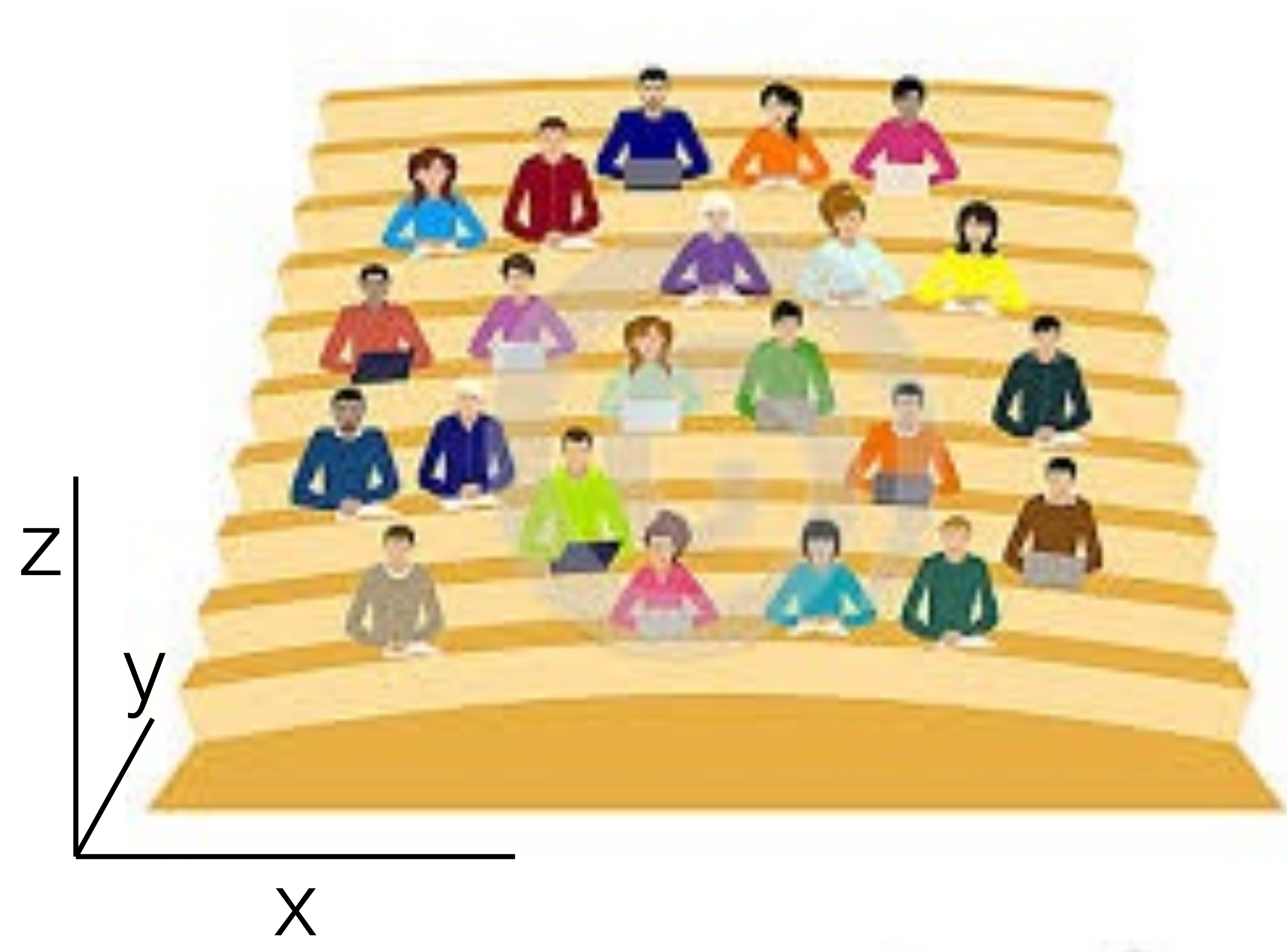
$$A\mathbf{x} = \lambda\mathbf{x}$$

Which Direction?



Top Eigenvector of covariance matrix

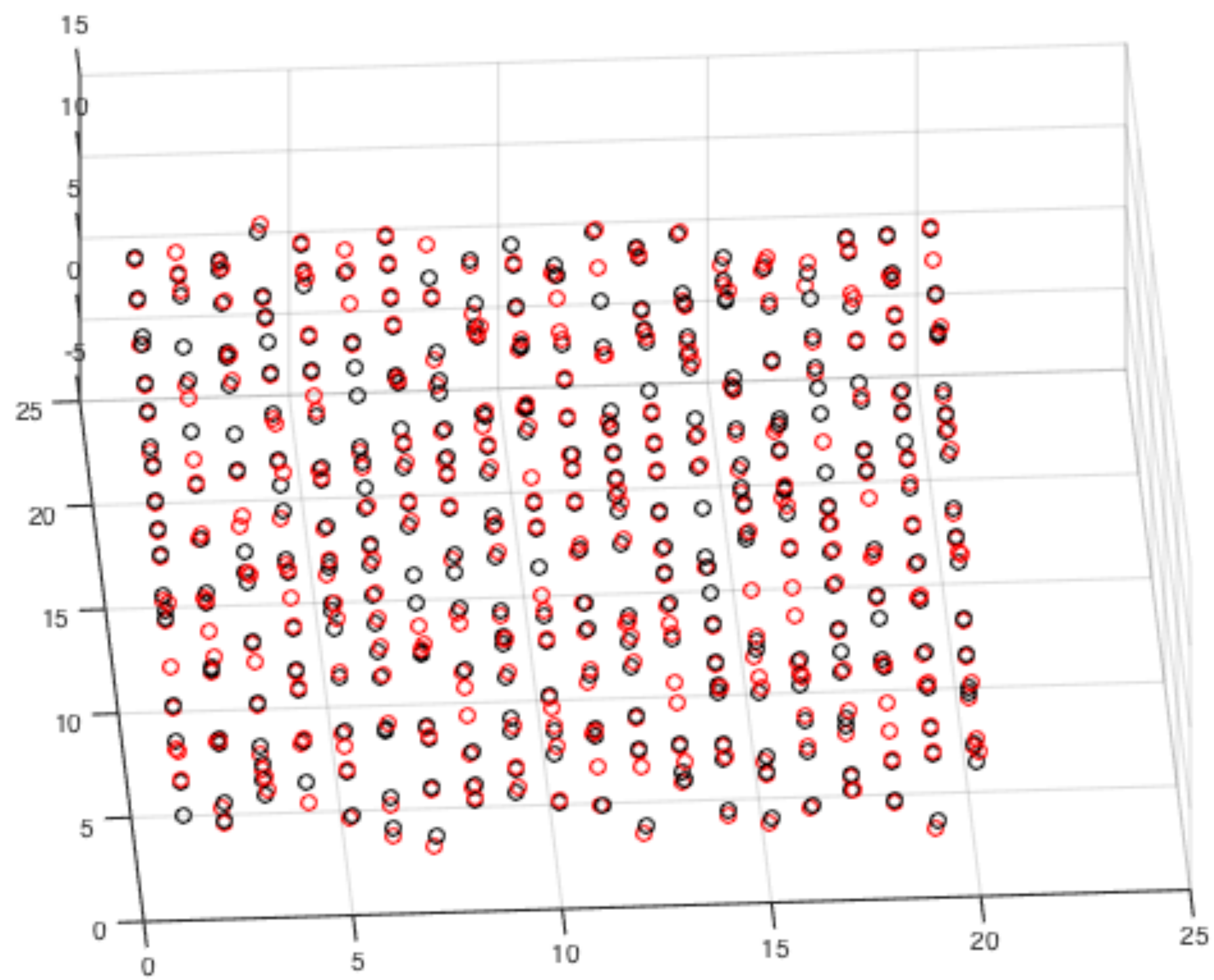
- What if we want more than one number for each data point?
- That is we want to reduce to K > 1 dimensions?



PCA: Variance Maximization

- How do we find the K components?

Ans: Maximize sum of spread in the K directions



PCA: Variance Maximization

- How do we find the K components?
- We are looking for orthogonal directions that maximize total spread in each direction

$$\sum_{k=1}^{d} \mathbf{w}_i[k]\mathbf{w}_j[k] = 0 \quad \& \quad \sum_{k=1}^{d} \mathbf{w}_i[k] = 1$$

Intuition: Remove top direction, now reduce dimension for remaining d-1 dimensions

1. $$\Sigma = \mathrm{cov}\left(X \right)$$

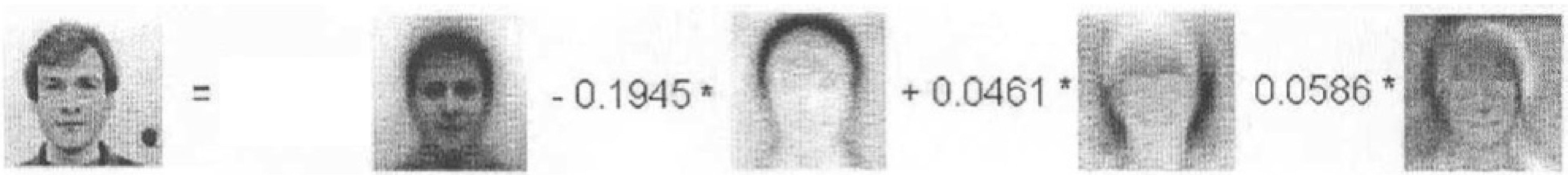
2. $$W = \mathrm{eigs}\left(\Sigma, K \right)$$

3. $$Y = X \times W$$

Principal Component Analysis (PCA)

Turk & Pentland'91

Eigen Face:



- Each x_t (each row of X) is a face image (vectorized version)
- Each y_t is the set of coefficients we multiply to the eigen face
- Each column of W is an Eigenface

ORTHONORMAL PROJECTIONS

- (Centered) Data-points as linear combination of some orthonormal basis, i.e.

$$\mathbf{x}_t = \mu + \sum_{j=1}^{d} \mathbf{y}_t[j]\mathbf{w}_j$$

where $\mathbf{w}_1, \ldots, \mathbf{w}_d \in \mathbb{R}^d$ are the orthonormal basis and $\mu = \frac{1}{n}\sum_{t=1}^{n} x_t$.

Both Views are Equivalent

- To Maximize variance pick the maximum K directions
- To Minimize reconstruction error drop the minimum d-K eigen directions

Principal Component Analysis

1. $$\Sigma = \mathrm{cov}\left(X\right)$$

2. $$W = \mathrm{eigs}\left(\Sigma, K\right)$$

3. $$Y = X - \mu \times W$$

Reconstruction

4. $$\hat{X} = Y \times W^\top + \mu$$