Lec 17: Dimensionality Reduction

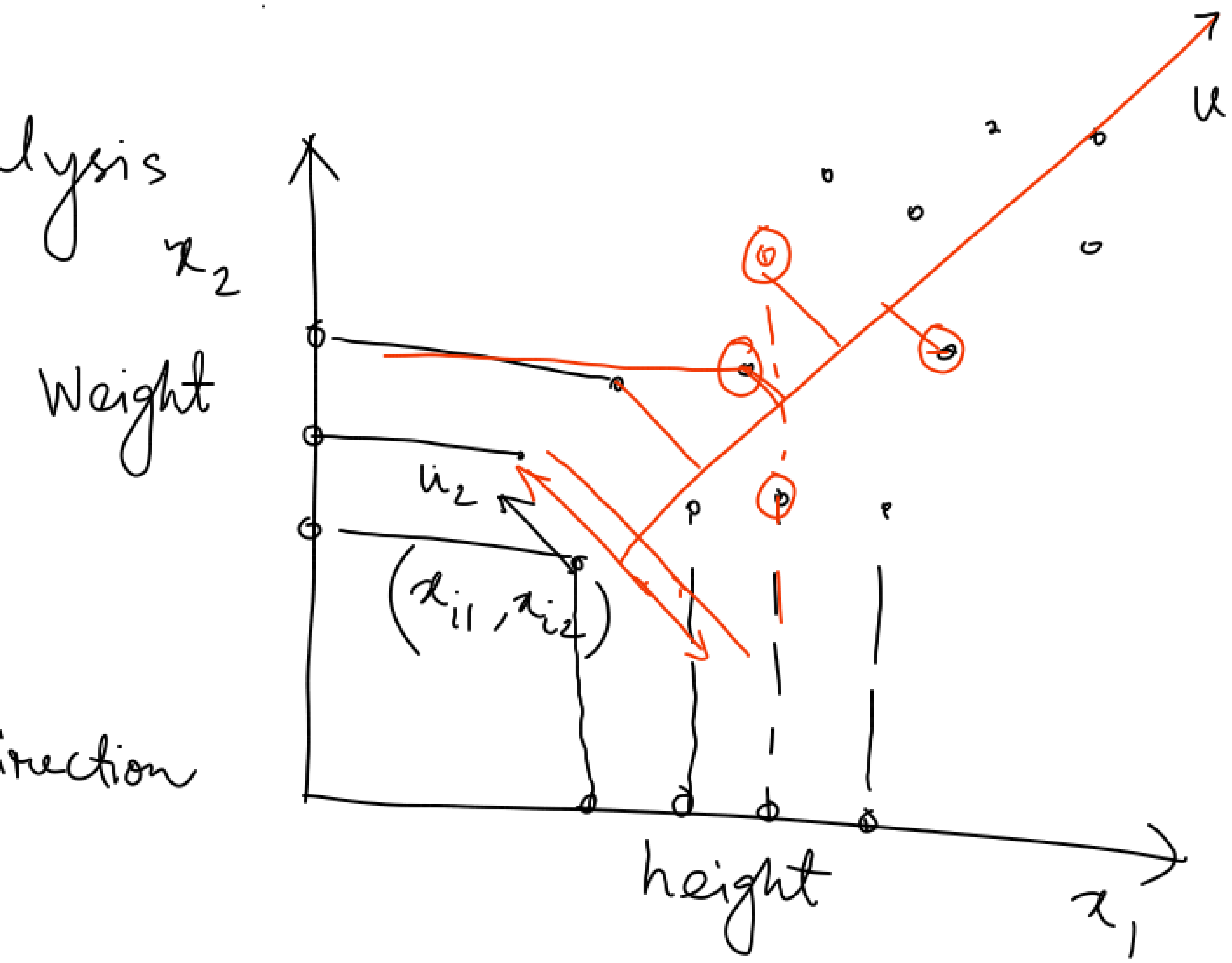
→ Q: Can we reduce the dimension without compromising the information?

Two methods

Unsupervised: Pricipal Component Analysis
$\{x_i\}$

Supervised: LDA →
$\{(x_i, y_i)\}$

PCA



u_1 is a better projection direction than u_2 for this data.

Uses:

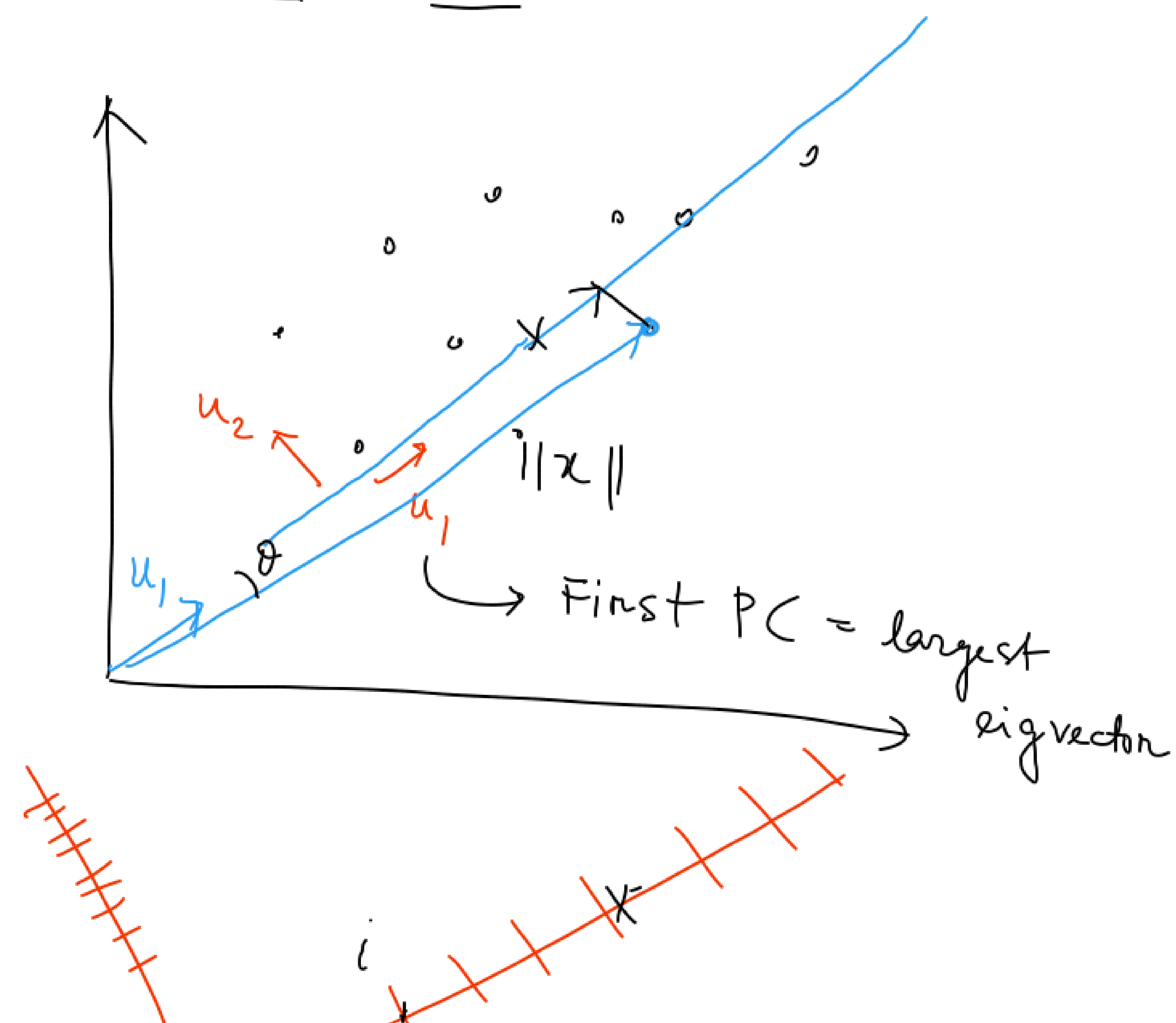
1. Data visualization
2. Feature extraction

<u>Setup</u>: $D = \{x_1, x_2, \dots, x_n\}$, $x_i \in \mathbb{R}^d$

Goal: to <u>map</u> these data on a dimension $\underline{m} < \underline{d}$

Q: which u_1 to pick?

→ objective: maximize the variance of the projected data.



Projection of x on u_1 direction

$$\underline{(x_i^T u_1)}\, u_1 = \text{proj}_{u_1}(x_i) \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

mean of the projected data on u_1 $= \frac{1}{n}\sum_{i=1}^{n} \text{proj}_{u_1}(x_i) = (u_1^T \bar{x})\, u_1$

Variance of the projections

$$\frac{1}{n}\sum_{i=1}^{n}\left(u_1^T x_i - u_1^T \bar{x}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(u_1^T \left(x_i - \bar{x}\right)\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_1^T \left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T u_1$$

$$= u_1^T \left[\underbrace{\frac{1}{n}\sum_i \left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T}_{S}\right] u_1$$

$$= u_1^T S u_1$$

Aside:

$$\begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} \qquad \begin{matrix} \mu_1 \\ \mu_2 \end{matrix}$$

$$X = \begin{bmatrix} x_1 - \mu & x_2 - \mu & \cdots & x_n - \mu \end{bmatrix}$$

$\underline{\frac{1}{n} X X^T}$ = Covariance matrix of the data points

$$= S$$

$$\begin{bmatrix} | & & | \\ x_1 - \mu & \cdots & x_n - \mu \\ | & & | \end{bmatrix} \begin{bmatrix} — (x_1 - \mu)^T — \\ \vdots \\ — (x_n - \mu)^T — \end{bmatrix}$$

Optimization

$$\max \quad u_1^T S u_1$$

$$\text{s.t.} \quad u_1^T u_1 = 1$$

not convex opt problem

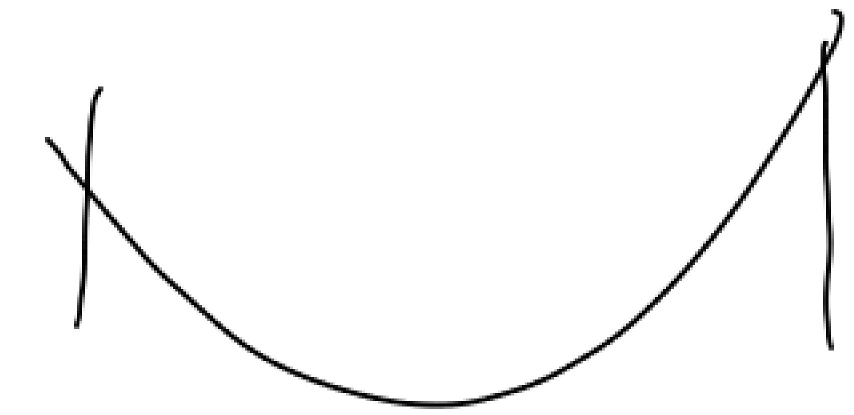
Fact: For a real symmetric matrix A, A has d orthonormal eigenvectors.
$d \times d$

$$\boxed{A u = \lambda u}$$

eigen value

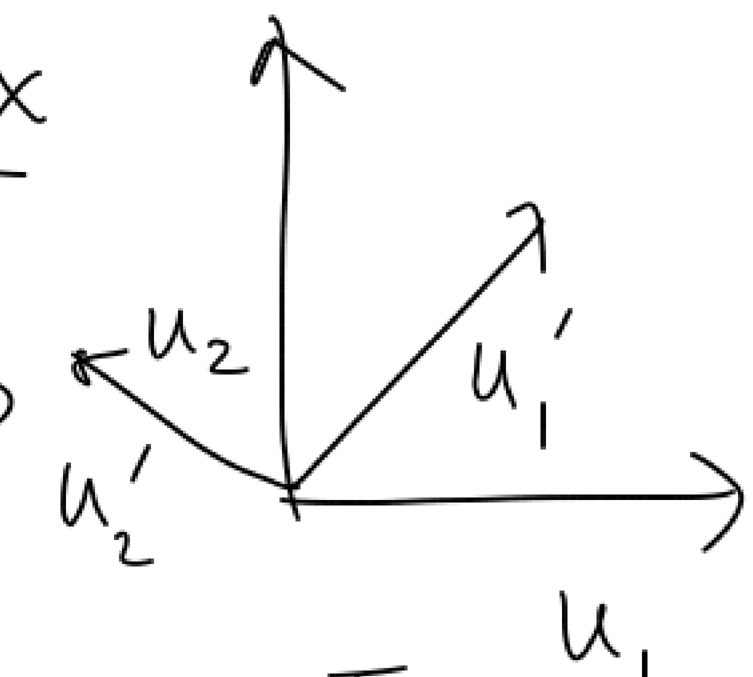
eigenvector

$\nabla^2 f(u_1)$ is PSD



$S \to$ real symmetric matrix

$$y^T (X X^T) y = \|X^T y\|^2 \geq 0$$



$$\mathcal{L}(\lambda, u_1), \quad \frac{\partial \mathcal{L}}{\partial u_1} = 0$$

$u_1^T u_2 = 0$

$u_1^T u_1 = u_2^T u_2 = 1$

S is Real Symmetric

$v_1, v_2, \ldots, v_d$ orthonormal eigenvectors

$\lambda_1, \lambda_2, \ldots, \lambda_d$ eigvalues $\quad v_i^T v_j = \mathbb{I}\{i=j\}$

$$Sv_i = \lambda_i v_i \qquad V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$$SV = \begin{bmatrix} Sv_1 & Sv_2 & \cdots & Sv_d \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_d v_d \end{bmatrix}$$

$$= V \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{bmatrix}}_{\Sigma} = V\Sigma \quad \cdots \quad (1)$$

unitary

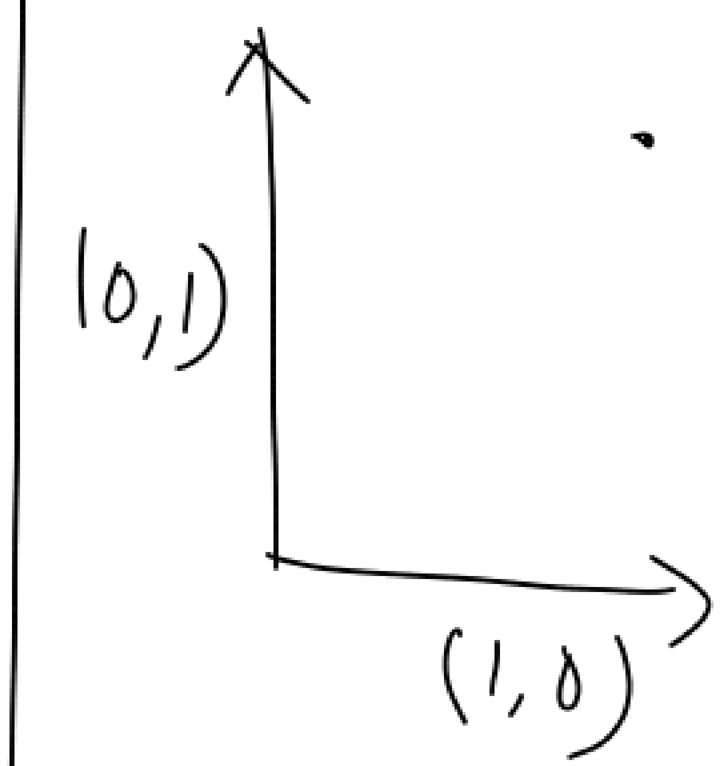
$$V^T V = \begin{bmatrix} -v_1^T- \\ -v_2^T- \\ \\ \end{bmatrix} \begin{bmatrix} | & | & \\ v_1 & v_2 & \\ | & | & \end{bmatrix} = I = VV^T$$

$$SVV^T = V\Sigma V^T$$

$$\boxed{S = V\Sigma V^T} \quad \text{---} \quad (2)$$

eigenvalue decomposition

$$u_1 = \sum_{j=1}^{d} \overset{\nearrow z_j}{x_j} \, v_j \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$



$$\boxed{u_1 = Vx} \quad (3)$$

$$\max \; u_1^T S u_1$$

$$\text{s.t. } u_1^T u_1 = 1 \longrightarrow x^T x = 1$$

$$x^T \underbrace{V^T V}_{I} \Sigma \underbrace{V^T V}_{I} x = x^T \Sigma x$$

not the data points

$$= \sum_{j=1}^{d} \lambda_j x_j^2$$

$$\max \sum_{j=1}^{d} \lambda_j x_j^2$$

$$\text{s.t. } \sum_{j=1}^{d} x_j^2 = 1$$

$\Rightarrow$ Solution $\Rightarrow$ largest λ_j is the output

$x_j = 1$, all others $= 0$

First PC $= u_1 = v_1$

How to generalize?

$$\max \; u_2^T S u_2$$

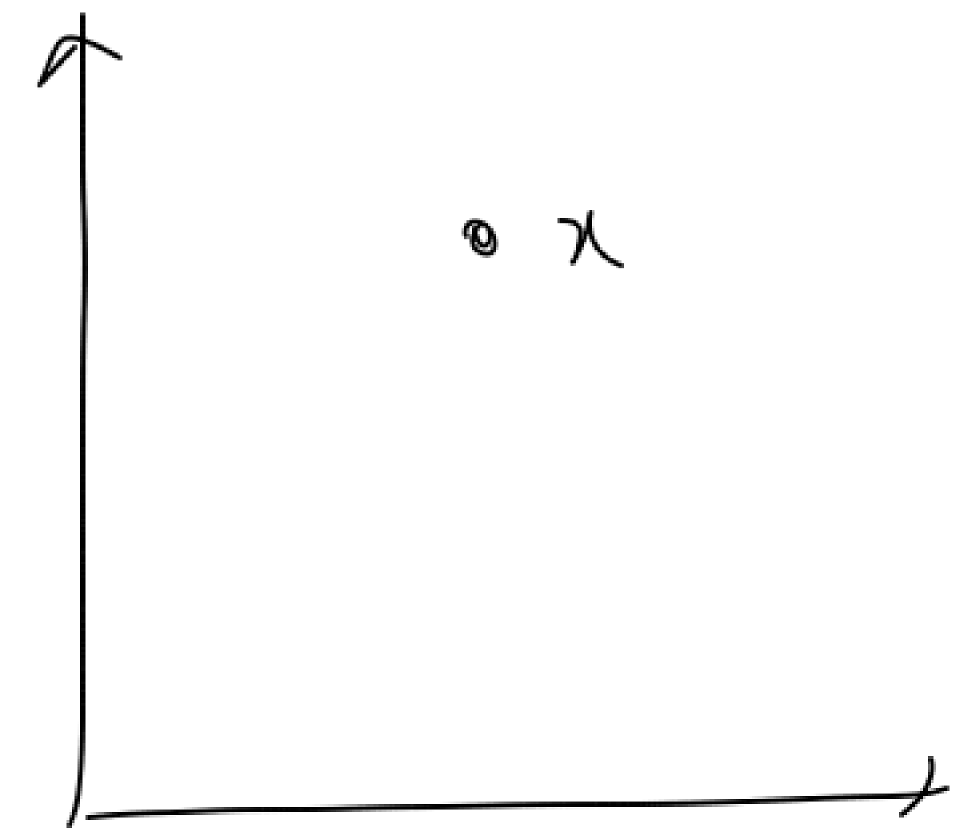
$$\text{s.t. } u_2^T u_2 = 1$$

$$u_2^T u_1 = 0$$

$\Rightarrow$

u_2 = second largest eigenvector

PCA algorithm



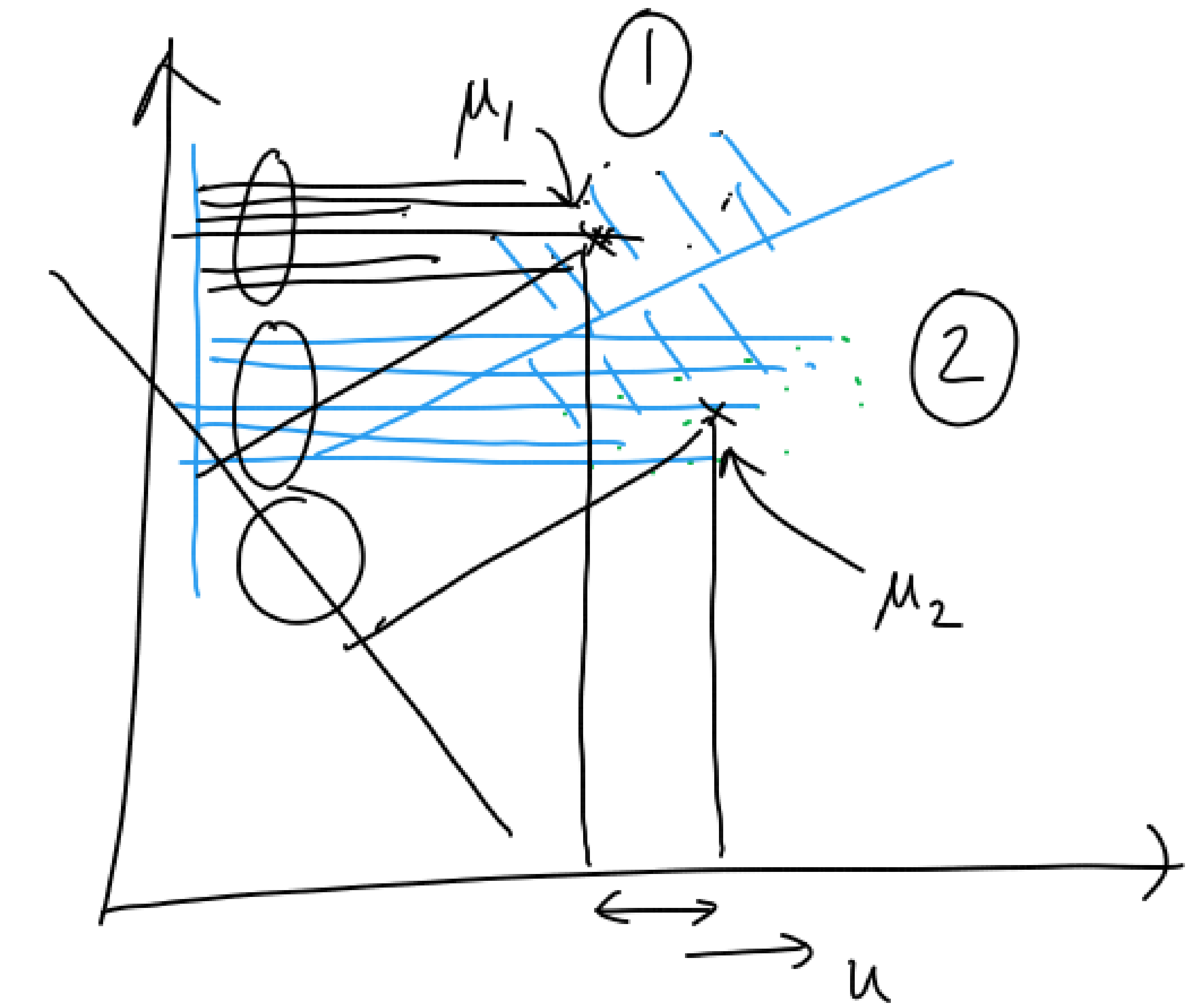
1. Compute mean of data points $\bar{x}$
2. Mean center the data:
   Compute $(x_i - \bar{x})$
3. Compute covariance matrix $S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$
4. Do eigenvalue decomposition of $S = V \Sigma V^T$
5. Pick m top eigenvectors $u_1, \ldots, u_m \to m$ PCs.
6. $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$ projection matrix

$$U^T x = \begin{bmatrix} u_1^T x \\ u_2^T x \\ \vdots \\ u_m^T x \end{bmatrix}$$

Supervised (x_i, y_i)

PCA fails to retain class information

We want a projection that separates the classes well.



Need : a measure

$$J(u) = \left| \mu_1^T u - \mu_2^T u \right|$$ is not a good measure.

Objective: 1 The different class means are well separated

2 The data in the same class are NOT very separated.