Open Session

General workflow

1) Given a dataset

$$D = \begin{bmatrix} 1 & 0 & 3 & 2 & -5 \\ 2 & -1 & -5 & 4 & 3 \end{bmatrix}$$

Five different data-points in $\mathbb{R}^2$.

- $n = 5, d = 2$

2) Centering the dataset

$$\boldsymbol{\mu} = \frac{1}{5}\sum_{i=1}^{5} \mathbf{x}_i = \frac{1}{5}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 3 \\ -5 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} -5 \\ 3 \end{bmatrix}\right\} = \frac{1}{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 0.8 & -0.2 & 2.8 & 1.8 & -5.2 \\ 1.4 & -1.6 & -5.6 & 3.4 & 2.4 \end{bmatrix}$$

3) Covariance matrix

$$\mathbf{C} = \frac{1}{5} \cdot \sum_{i=1}^{5} \mathbf{x}_i \mathbf{x}_i^T = \begin{bmatrix} 0.8 \\ 1.4 \end{bmatrix}\begin{bmatrix} 0.8 & 1.4 \end{bmatrix} = \begin{bmatrix} 0.8 \times 0.8 & 0.8 \times 1.4 \\ 1.4 \times 0.8 & 1.4 \times 1.4 \end{bmatrix}$$

$$+ \ldots$$

$$= \begin{bmatrix} 7.76 & -4.12 \\ -4.12 & 10.64 \end{bmatrix}$$

Checks:

- Symmetric

- $d \times d$

4) Find the principal components → eigenvectors of $\mathbf{C}$

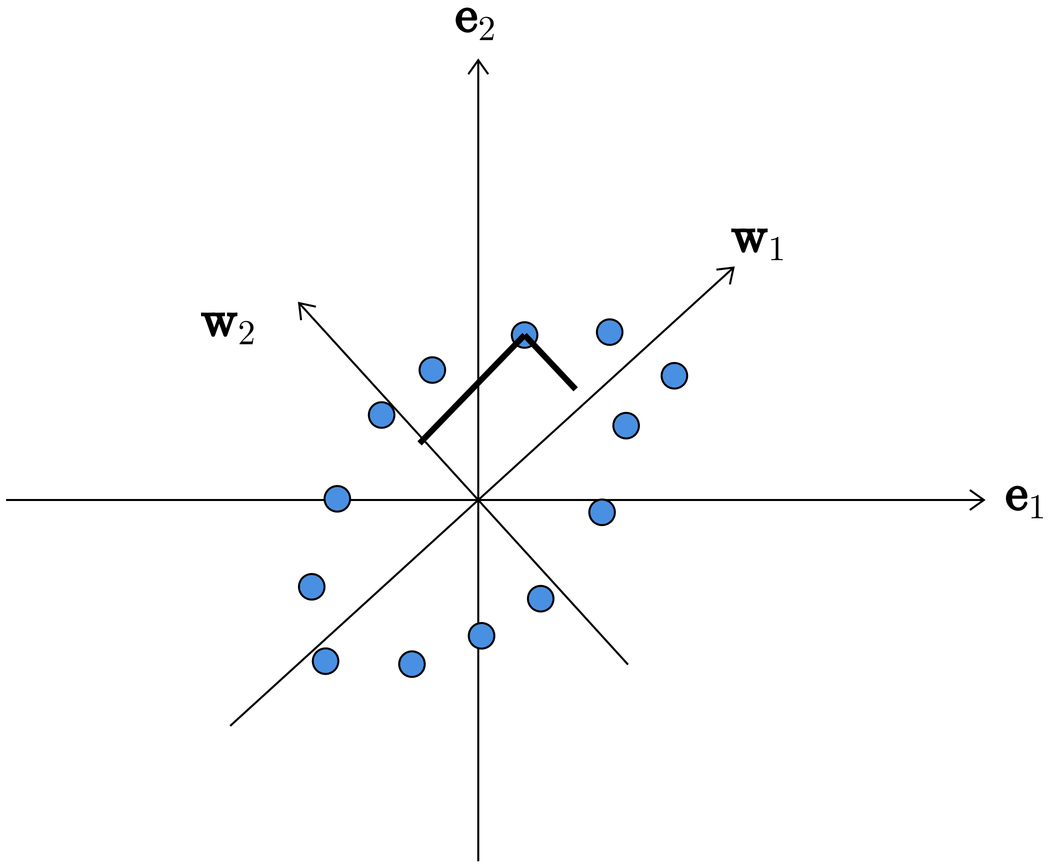
$$\lambda_1 = 13.56,\ \lambda_2 = 4.84$$

$$\mathbf{w}_1 = \begin{bmatrix} -0.58 \\ 0.82 \end{bmatrix},\ \mathbf{w}_2 = \begin{bmatrix} -0.82 \\ -0.58 \end{bmatrix}$$

Checks:

- Orthogonality
- Unit norm
- Eigenvector-eigenvalue

5) Select the value of k



$$\mathbf{X} = \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix} \qquad \mathbf{C} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$$

$$\mathbf{C} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T \qquad \mathbf{Y} = \mathbf{Q}^T\mathbf{X}$$

$$\mathbf{Q} = \begin{bmatrix} | & | \\ \mathbf{w}_1 & \mathbf{w}_2 \\ | & | \end{bmatrix}$$

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

$$\begin{aligned} \mathbf{C}' &= \frac{1}{n}\mathbf{Y}\mathbf{Y}^T \\ &= \frac{1}{n}\left(\mathbf{Q}^T\mathbf{X}\right)\left(\mathbf{X}^T\mathbf{Q}\right) \\ &= \mathbf{Q}^T\mathbf{C}\mathbf{Q} \\ &= \mathbf{D} \\ &= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \end{aligned}$$

$\mathbf{C}'$: Covariance matrix
in the new basis
is diagonal

$$\begin{aligned} \mathbf{Y} &= \mathbf{Q}^T\mathbf{X} \\ &= \begin{bmatrix} — & \mathbf{w}_1^T & — \\ — & \mathbf{w}_2^T & — \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{w}_1^T\mathbf{x}_1 & \cdots & \mathbf{w}_1^T\mathbf{x}_n \\ \mathbf{w}_2^T\mathbf{x}_1 & \cdots & \mathbf{w}_2^T\mathbf{x}_n \end{bmatrix} \end{aligned}$$

$$\mathbf{x}_i = \left(\mathbf{x}_i^T\mathbf{w}_1\right)\mathbf{w}_1 + \left(\mathbf{x}_i^T\mathbf{w}_2\right)\mathbf{w}_2$$