TD PCA

Exercice 5.1

$$g = X^T \frac{1}{n} I_n \mathbf{1} = \frac{1}{n} X^T \mathbf{1} = \begin{bmatrix} 3 \\ 20 \end{bmatrix}$$

$$\bar{X} = X - \mathbf{1} g^T = \begin{bmatrix} -3 & 3 \\ -1 & -1 \\ 0 & 1 \\ 1 & 3 \\ 3 & 0 \end{bmatrix}$$

$D_1 = \mathbb{V}(X_1) = 5$

$D_2 = \mathbb{V}(X_2) = 5$

donc

$$Y = \frac{1}{\sqrt{5}} \begin{bmatrix} -3 & 3 \\ -1 & -1 \\ 0 & 1 \\ 1 & 3 \\ 3 & 0 \end{bmatrix}$$

$$C = \frac{1}{5} Y^T Y = \frac{1}{25} \begin{bmatrix} 20 & 13 \\ 13 & 20 \end{bmatrix}$$

$\lambda_1 = 33/25$

$v_1 = (1, 1)$

$\lambda_2 = 7/25$

$v_2 = (-1, 1)$

The best line that approximates the dataset is the line of equation $y = x + 17$,