Principal Component Analysis

6.1 Centrage

$$\tilde{X} = \left(\overrightarrow{X^{(1)}} \quad \overrightarrow{X^{(2)}} \right) = \begin{bmatrix} \overline{X^{(1)}} & \overline{X^{(2)}} \\ \overline{X^{(1)}} & \overline{X^{(2)}} \end{bmatrix}$$

$$3 \qquad 20$$

$$Y = \begin{bmatrix} \tilde{X}^{(1)} & \tilde{X}^{(2)} \\ \sigma_{\tilde{X}^{(1)}} & \sigma_{\tilde{X}^{(2)}} \end{bmatrix} \qquad \sigma_{\tilde{X}^{(1)}} = \frac{1}{N-1} \sum_{1}^{N} \tilde{X}_i^{(1)^2}$$

$$\overline{X} = \begin{pmatrix} -3 & -3 \\ -1 & -1 \\ 0 & 1 \\ 1 & 3 \\ 3 & 0 \end{pmatrix} \qquad Y = \frac{1}{\sqrt{5}} \times \overline{X}$$

$$Cov(Y) = \mathbb{E}[Y^T Y] = \frac{1}{N-1} Y^T Y = \begin{bmatrix} \sigma^2_{Y1} & \rho\sigma_{Y1}\sigma_{Y2} \\ \rho\sigma_{Y2}\sigma_{Y1} & \sigma^2_{Y2} \end{bmatrix} = \begin{bmatrix} 1 & 13/20 \\ 13/20 & 1 \end{bmatrix}$$

Axes Principaux : $\vec{V}^{(1)}, \vec{V}^{(2)}$ sont donnés par les vecteurs propres unitaires de $Cov(Y)$

$$\det(Cov(Y) - \lambda Id) = 0$$

$$\det \begin{bmatrix} 1-\lambda & 13/20 \\ 13/20 & 1-\lambda \end{bmatrix} = 0 \Leftrightarrow (1-\lambda)^2 - (13/20)^2 = 0 \qquad \lambda_1 = 33/20 = \sigma_1^2$$

$$\lambda_2 = 7/20 = \sigma_2^2$$

$$r_1 = \sigma_1^2 / (\sigma_1^2 + \sigma_2^2) \qquad r_2 = \sigma_2^2 / (\sigma_1^2 + \sigma_2^2)$$

$$\lambda_1 = 33/20 \begin{bmatrix} -13/20 & 13/20 \\ 13/20 & -13/20 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \vec{V}' = \begin{bmatrix} V_1 \\ V_1 \end{bmatrix}$$

$$\|\vec{V}\| = \sqrt{2 V_1^2} = \sqrt{2}\,|V_1| \qquad \vec{V}^{(1)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 7/20 \begin{bmatrix} 13/20 & 13/20 \\ 13/20 & 13/20 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \vec{V}' = \begin{bmatrix} V_1 \\ -V_1 \end{bmatrix} \qquad \vec{V}^{(2)} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$C = YV$

Composantes principales $\left[\vec{V}^{(1)} \quad \vec{V}^{(2)} \right]$

$$C = \frac{1}{\sqrt{10}} \begin{bmatrix} -6 & 0 \\ 2 & 0 \\ 1 & -1 \\ 4 & -2 \\ 3 & -3 \end{bmatrix}$$

Approximation de rang 1 :

$Y = USV^T$

$$C_1 = YV^{(1)} = US^T V^{(1)} = US \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} = U \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} = U_1 \sigma_1 V_1^T$$

$$Y_{[1]} = C_{[1]} \cdot V^{(1)T} = \frac{1}{\sqrt{20}} \begin{bmatrix} -6 & -6 \\ -2 & -2 \\ 1 & 1 \\ 4 & 4 \\ 3 & 3 \end{bmatrix}$$

Meilleur droite d'approximation :

$$\vec{W}^T \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$\vec{V}^{(2)T} \begin{bmatrix} y_2 \\ y_1 \end{bmatrix} = 0 \rightarrow Y_2 = Y_1$$

$\hookrightarrow$ Orthogonal au vecteur $\vec{V}^{(1)}$