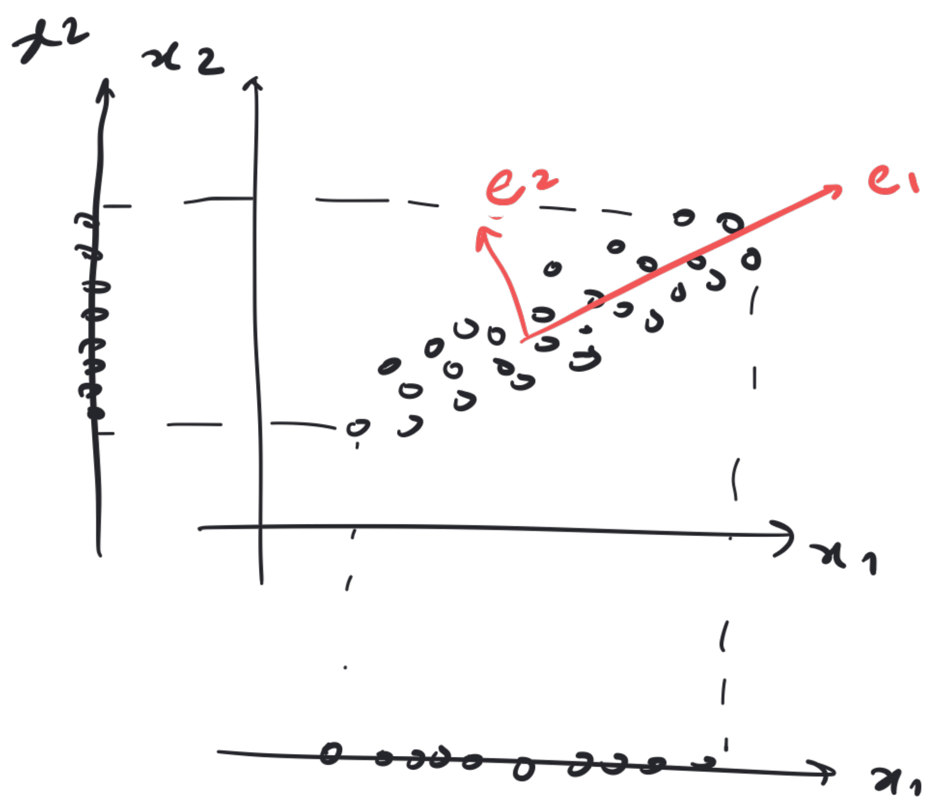
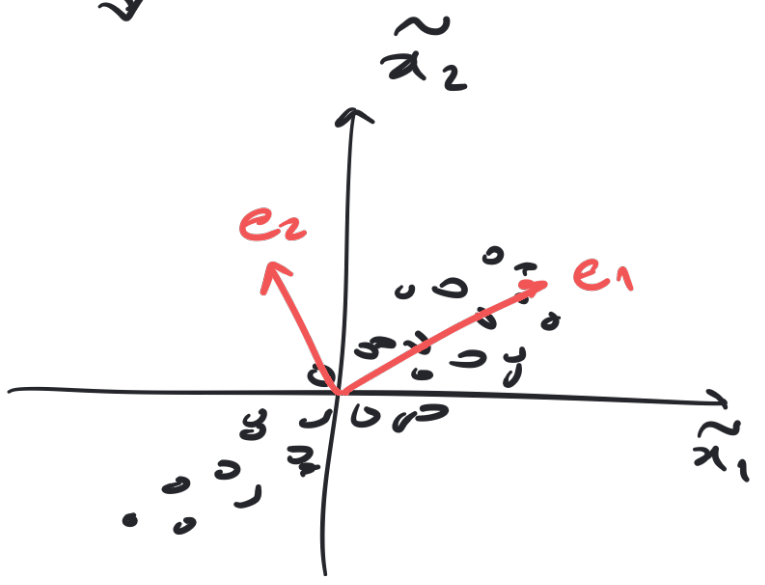
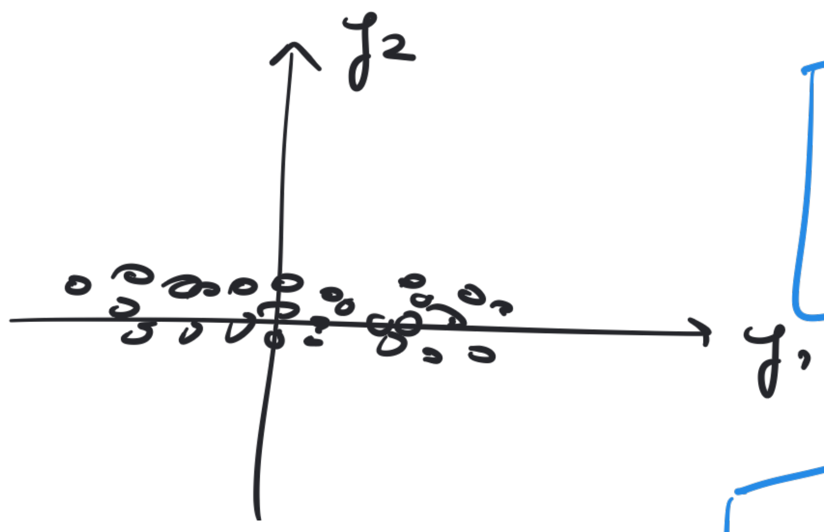
DEMEAN DATA $\tilde{x} = x - \mu_x$





NOTE:
DEMEANING or STANDARDIZATION ARE METHODS for preprocessing while using PCA.



$$\boxed{\underset{2\times N}{y} = \underset{2\times 2}{A} \cdot \underset{2\times N}{\tilde{X}}}$$

Apply a change of basis (rotation)

$$\boxed{\underset{1\times N}{y} = \underset{1\times 2}{A_2} \, \underset{2\times N}{\tilde{X}}}$$

Dimensionality reduction

Keeps direction that explains the most variance

Linear transformation: $y = A.\tilde{X}$

where $\tilde{X} = X - \mu_x$

and $\tilde{X}$ is a $D \times N$ matrix

A is a $M \times D$ matrix $(M \leq D)$

$$A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_N^T \end{bmatrix}$$ where a_i is a $D \times 1$

Goal: Find A such that

① $\mathrm{COV}(y_i, y_j) = 0, \ \forall y_i \neq y_j$

② $\mathrm{COV}(y_i, y_i) = \mathrm{VAR}(y_i)$ is maximum, $\forall y_i$

Covariance of $Y_{M\times N}$: K_Y

$$\boxed{K_{Y_{M\times M}} = E[(Y-\mu_Y)(Y-\mu_Y)^T]}$$

$$= E[YY^T]$$

$\mu_Y = 0$

$$= E[A.\tilde{X}.(A.\tilde{X})^T]$$

$$= E[A.\tilde{X}.\tilde{X}^T.A^T]$$

$$= A.\underbrace{E[\tilde{X}\tilde{X}^T]}.A^T]$$

$E[\cdot]$ is a linear operator

$K_{\tilde{X}}$ = covariance of $\tilde{X} = X - \mu_X$

$$\boxed{= A.K_{\tilde{X}}.A^T}$$

$$K_Y = A \cdot K_{\tilde{X}} \cdot A^T$$

$$= \underbrace{\begin{bmatrix} a_1^T \\ \hline a_2^T \end{bmatrix}}_{2\times2} \cdot \underbrace{K_{\tilde{X}}}_{2\times2} \cdot \underbrace{\left[\, a_1 \,\middle|\, a_2 \,\right]}_{2\times2}$$

$$= \begin{bmatrix} a_1^T \cdot K_{\tilde{X}} \\ a_2^T \cdot K_{\tilde{X}} \end{bmatrix} \cdot \left[\, a_1 \,\middle|\, a_2 \,\right]$$

$$= \underbrace{\begin{bmatrix} a_1^T \cdot K_{\tilde{X}} \cdot a_1 & \overbrace{a_1^T K_{\tilde{X}} \cdot a_2}^{=0} \\ \underbrace{a_2^T \cdot K_{\tilde{X}} \cdot a_1}_{=0} & a_2^T \cdot K_{\tilde{X}} \cdot a_2 \end{bmatrix}}_{2\times2} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

we want to find a_1 and a_2 such that ① $a_1^T \cdot K_{\tilde{X}} \cdot a_2 = 0 = a_2^T \cdot K_{\tilde{X}} \cdot a_1$

② $a_1^T \cdot K_{\tilde{X}} \cdot a_1 > a_2^T \cdot K_{\tilde{X}} \cdot a_2$

If we are projecting onto a 1-D space,

$$\underset{1 \times N}{Y} = \underset{1 \times D}{a_1^T} \cdot \underset{D \times N}{\tilde{X}}$$

$$\underset{\vec{a_1}}{\arg\max} \; a_1^T \cdot K_x \, a_1$$

such that $\|a_1\|^2 = 1$

$\Leftrightarrow a_1^T \cdot a_1 = 1$

because we only care about the direction of $\vec{a_1}$ not its magnitude.

Lagrangian function:

$$\boxed{\mathcal{L}(a_1, \lambda) = a_1^T \cdot K_{\tilde{X}} \cdot a_1 + \lambda\,(1 - a_1^T a_1)}$$

$$\left(\text{or } \mathcal{L}(a_1, \lambda) = -\,a_1^T \cdot K_{\tilde{X}} \cdot a_1 - \lambda(1 - a_1^T a_1)\right)$$

minimization formulation

$$\frac{\partial \mathcal{L}}{\partial a_1} = 0 \iff \cancel{2} \cdot K_{\tilde{X}} \cdot a_1 - \cancel{2}\,\lambda \cdot a_1 = 0$$

$$\iff \boxed{K_{\tilde{X}} \cdot a_1 = \lambda \cdot a_1}$$

this is the generalized eigenvector eq.

$\therefore$ a_1 is an eigenvector of $K_{\tilde{X}}$ with eigenvalue λ

Since, in this example, $K_{\tilde{X}}$ has 2 eigenvectors, we pick the one with largest eigenvalue.

PSEUDO-code:

① Subtract the mean: $\tilde{x} = x - \mu_x$

② Compute the covariance of $\tilde{x}$, $K_{\tilde{x}}$.

③ Compute the eigenvectors and eigenvalues of $K_{\tilde{x}}$.

and store the eigenvectors in decreasing order of their eigenvalues

$$U = \left[E_1 \mid E_2 \mid \dots \mid E_D \right]_{D \times D}$$

MODAL MATRIX

$$\Lambda = \left[\lambda_1, \lambda_2, \dots, \lambda_D \right], \quad \lambda_1 > \lambda_2 > \dots > \lambda_D$$

④ Apply rotation or dimensionality reduction

$$Y = A \cdot \tilde{X}, \quad \text{where} \quad A = U^T$$

4.1) For dimensionality reduction $A = \left[E_1 \mid E_2 \mid \dots \, E_M \right]^T$

Explained Variance Ratio

M - dimensional space explains

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_M}{\sum_{i=1}^{D} \lambda_i} \%$$ of variance in the data