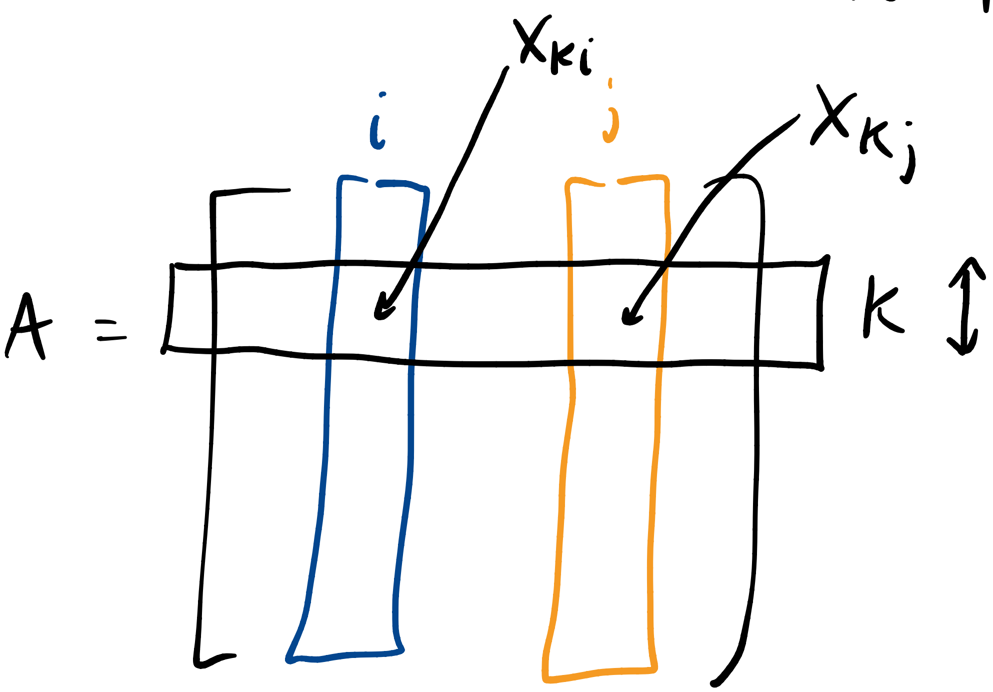
Lecture 18 | Covariance matrix computation

$$\Sigma = \begin{bmatrix} \overset{(i,j)}{\square} & \\ & \end{bmatrix}$$

Scalar equation

$$\Sigma(i,j) = \sum_{k=1}^{N} \frac{(x_{ki} - \mu_i)(x_{kj} - \mu_j)}{N-1}$$



__diagonal entry__: Assume $i = j$

$$\Sigma(i,i) = \sum_{k=1}^{N} \frac{(x_{ki} - \mu_i)(x_{ki} - \mu_i)}{N-1}$$

$$= \frac{1}{N-1} \sum_{k=1}^{N} (X_{ki}-\mu_i)(X_{ki}-\mu_i)$$

$$\boxed{= \frac{1}{N-1} \sum_{k=1}^{N} (X_{ki}-\mu_i)^2}$$

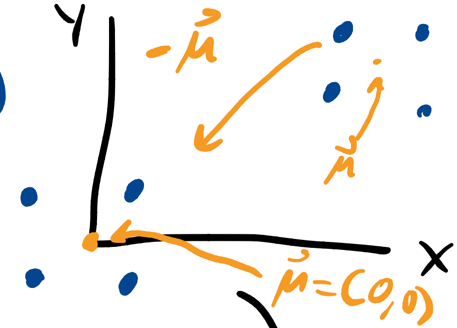
Variance of X_i

$$= \sigma_i^2$$

COV matrix :

$$\begin{array}{cc}
\begin{bmatrix} \overbrace{\mathrm{cov}(x,x)}^{\sigma_x^2} & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \underbrace{\mathrm{cov}(y,y)}_{\sigma_y^2} \end{bmatrix} & \begin{matrix} X \\ Y \end{matrix} \\
\begin{matrix} X \qquad\qquad & Y \end{matrix} &
\end{array}$$

Vectorized equation



$$\underbrace{\Sigma}_{(m,m)} = \frac{1}{N-1}\sum_{k=1}^{N} \underbrace{\underbrace{(\overbrace{\vec{x}_k}^{(1,m)} - \overbrace{\vec{\mu}_k}^{(1,m)})}_{(1,m)}{}^{T}}_{(m,1)} \underbrace{(\vec{x}_k - \vec{\mu}_k)}_{(1,m)}$$

$$A = \begin{bmatrix} \quad \end{bmatrix} \quad \vec{x}_k$$

$$A \quad \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

$$- \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix} \qquad - \begin{bmatrix} 2, & 5, & 8 \end{bmatrix}$$

$$= \begin{bmatrix} \quad \end{bmatrix} \qquad \underbrace{\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix}}_{\text{Centered } A = A_c}$$

Subtracting $\vec{\mu}$ from A is called <u>centering</u> the data — makes mean of each var $= 0$.

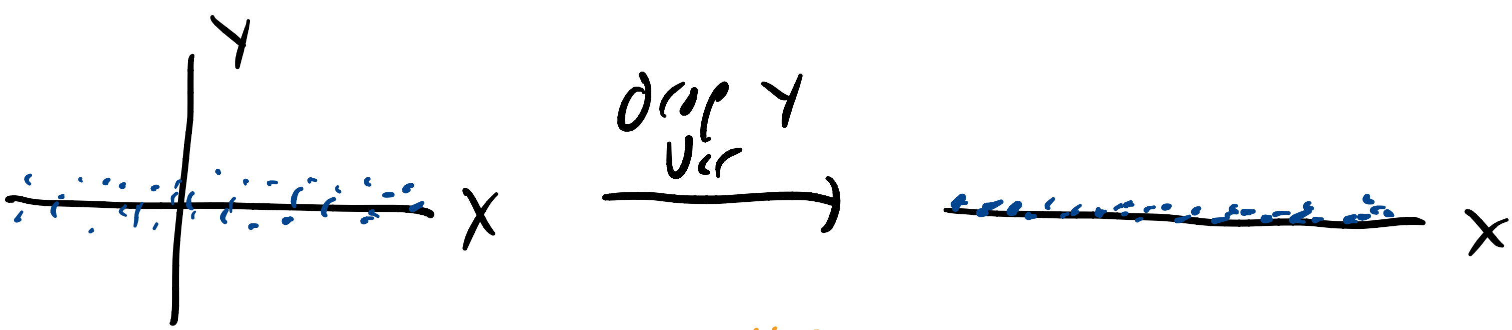
Matrix equation:

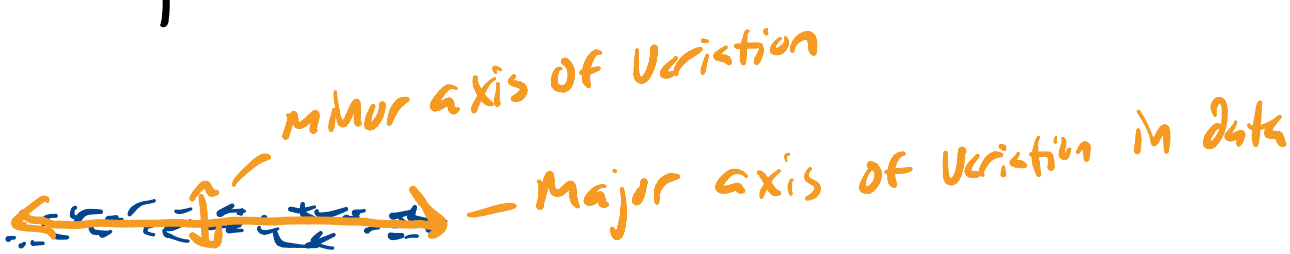
$$\boxed{\Sigma = \frac{1}{N-1} A_c^T @ A_c}$$

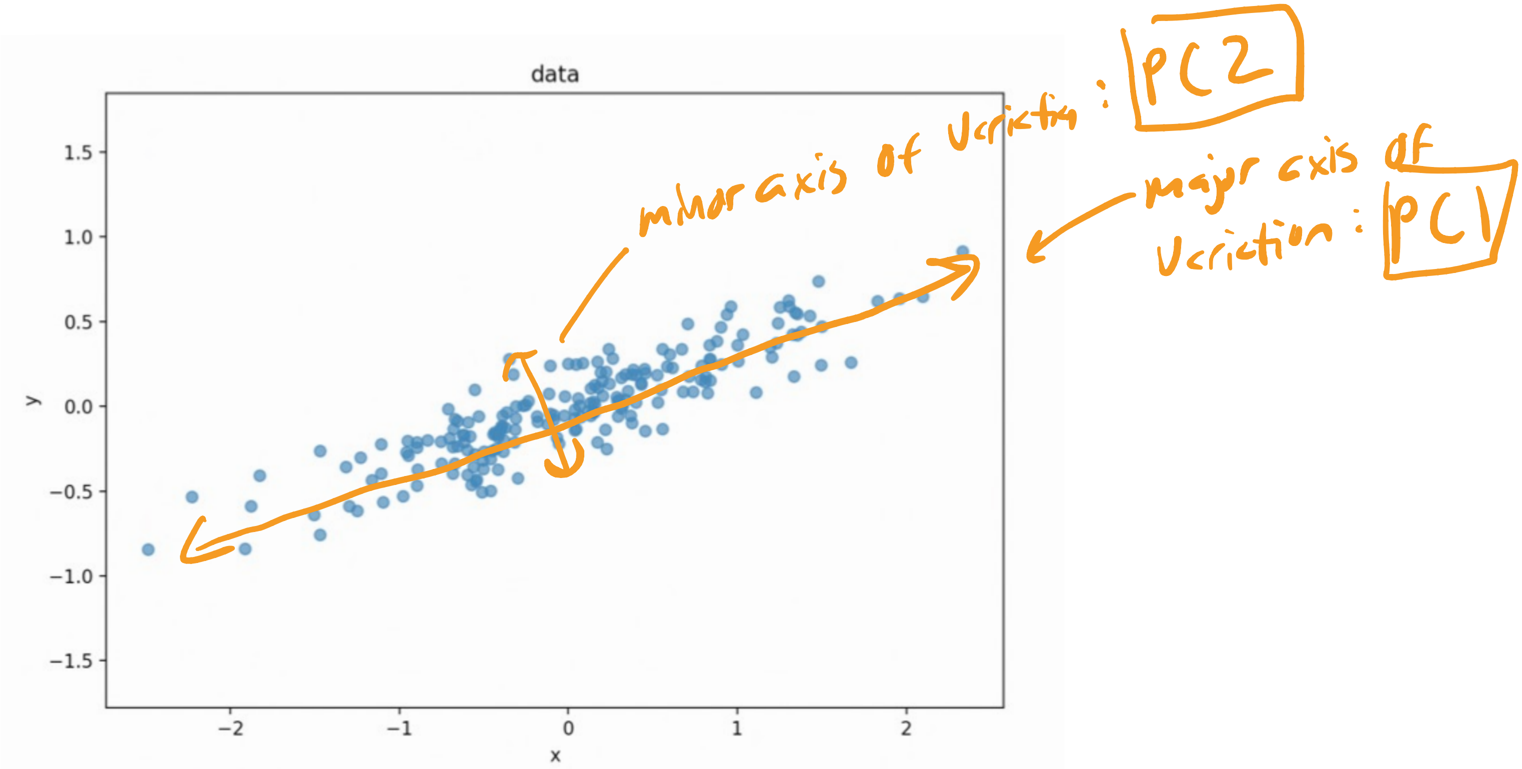
centered version of A

$$A - \vec{\mu} = A_c$$

<u>Principal component analysis (PCA)</u>







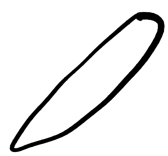
major axis != x variable dim
minor axis != y variable dim

principal components: Axes of variation of the data
(PC)

Number them: lower numbers mean more variation in that direction/axis

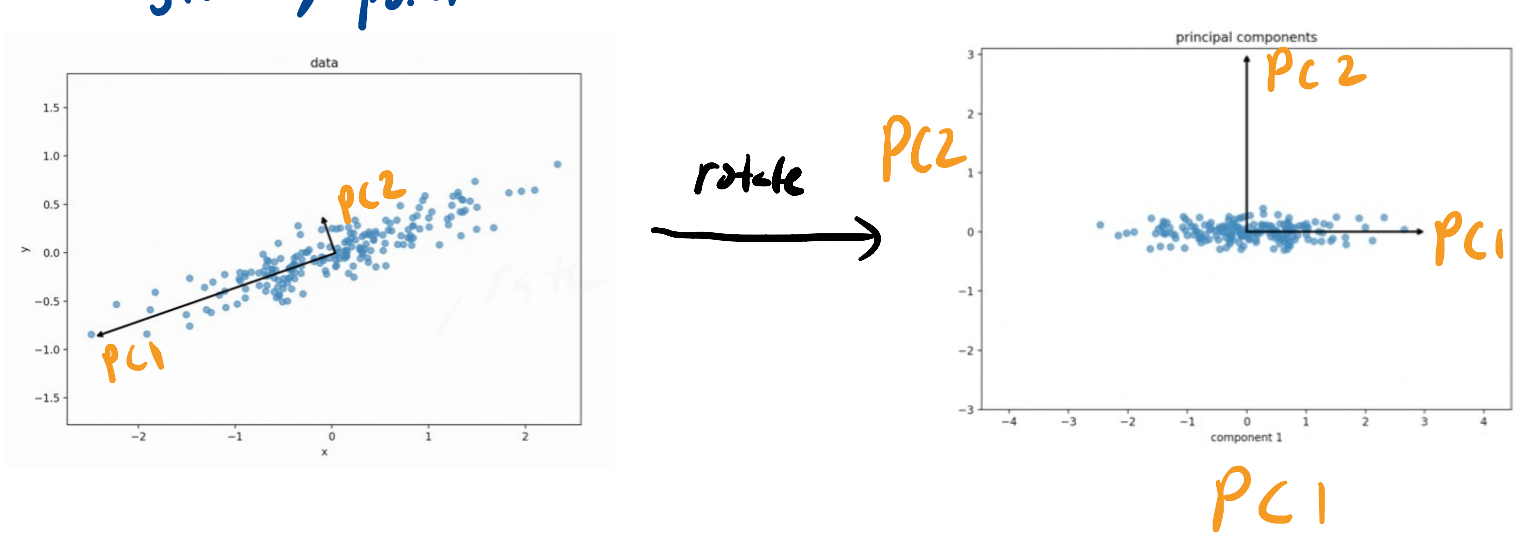
Goal of PCA

Lower dimensionality of data (e.g. 10,000D → 30D), do this by dropping <u>intrinsic</u> axes of variation in data (PCs)
⇒ <u>NOT</u> data variables

e.g.  example drop PC2.

★ Rotate coordinates so that PC dimensions become coordinate axes than raw data vars

Starting point:



To do this: Apply a translation and a rotation matrix:

$$\overbrace{\hat{A}_c}^{\text{rotated data centered at 0}} = \Big(\overbrace{P.T}^{\text{rotation matrix}} @ \overbrace{A_c}^{\text{centered data}}\Big).T$$

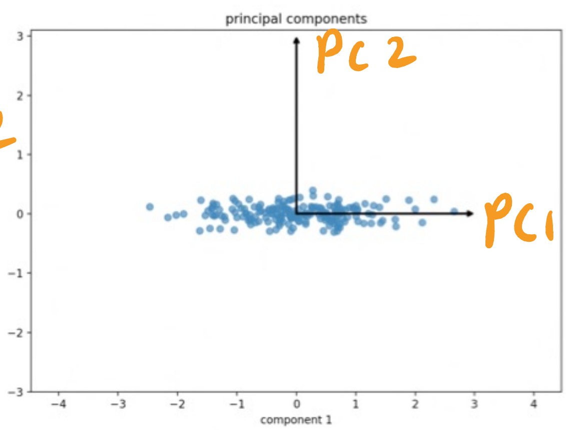
centered data $A_c = A \underbrace{- \vec{\mu}}_{\text{translation}}$

same

$$\hat{A}_c = A_c @ P$$

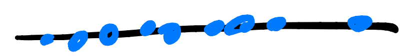
$\hat{A}_c$

PC2



PC1

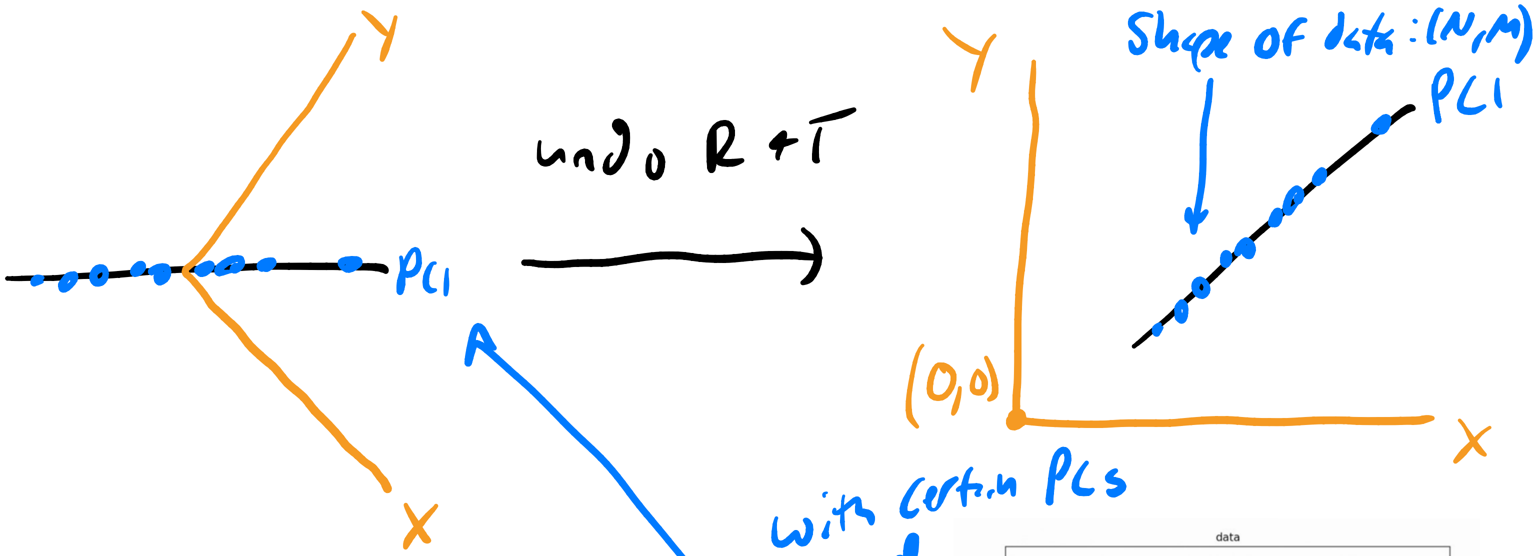
$$\hat{A}_c \xrightarrow{\overbrace{\hat{A}_c[:,0]}^{\text{drop PC2}}}$$



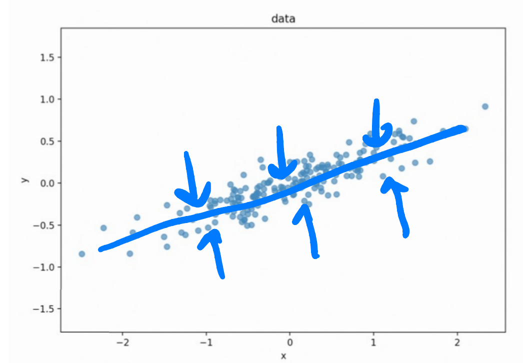
lowered dimensionality of data

Reconstruct data after dropping PCs!

Undo rotation and translation:



with certain PCs dropped



rotation matrix: cols from it dropped cooresponding to PCs

rotated data in PCA space: coordinate axes = PCs

$$A_{reconstruct} = \left(\hat{P} @ \hat{A}_c.T\right).T + \underbrace{\vec{\mu}}_{\text{un-center data}}$$

equal

$$= \hat{A}_c @ P.T + \vec{\mu}$$

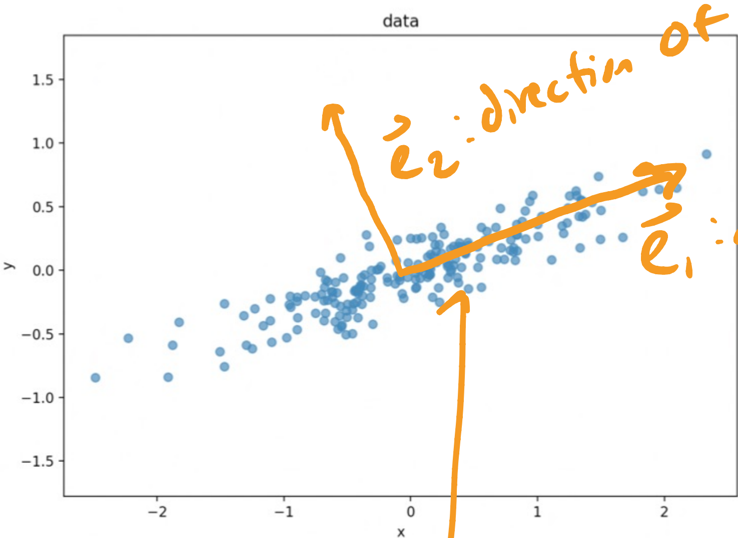
What is rotation matrix P?

P is matrix of eigenvectors of the $\underbrace{\text{covariance matrix}}_{\text{needs } A_c = A - \vec{\mu}}$ Σ of the data A.

Eigenvectors $P = [\overbrace{\vec{e}_1, \vec{e}_2, \vec{e}_3, \dots, \vec{e}_M}^{\text{Shape: } (M,)}]$

$\vec{e}_2$: direction of PC2

$P = [\vec{e}_1, \vec{e}_2]$



$\vec{e}_1$: direction of PC1

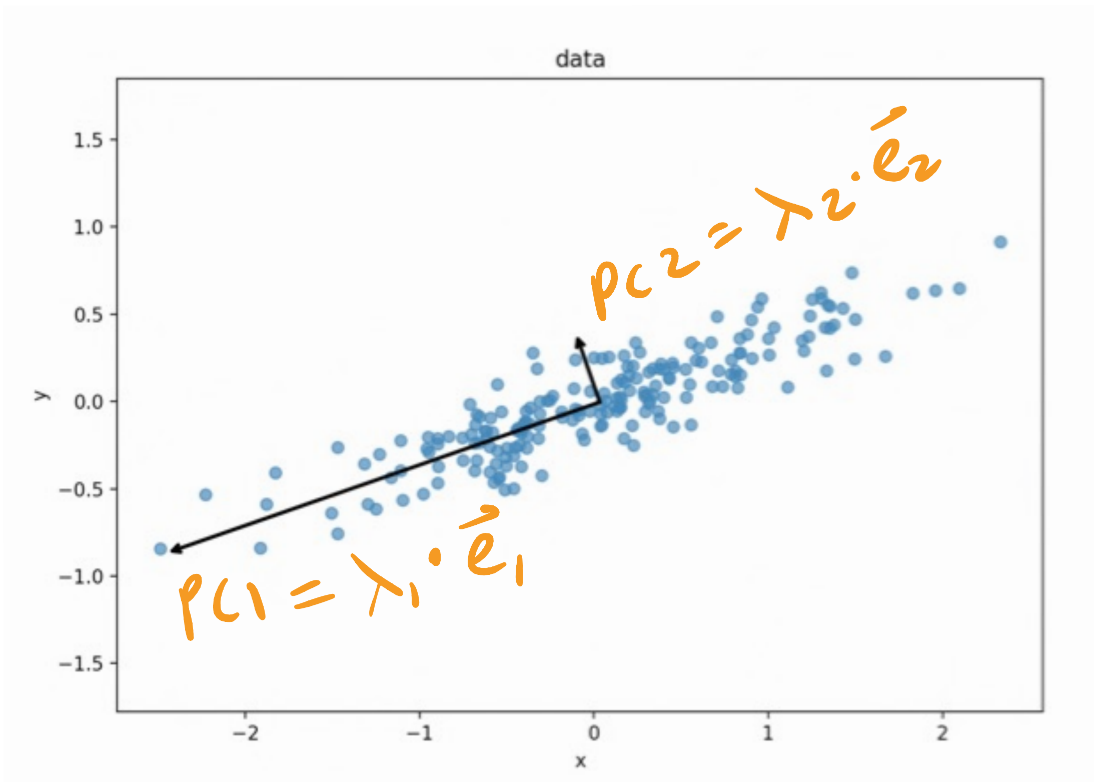
all unit lengths, not scaled.

Eigen value : Same # as eigen vectors

$\Rightarrow$ tell us <u>amount</u> of variation in each PC direction.

Scaler

$$e\_vals = [\underbrace{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_M}_{\text{Shape: } (M,)}]$$



Here:

$\lambda_1 > \lambda_2$