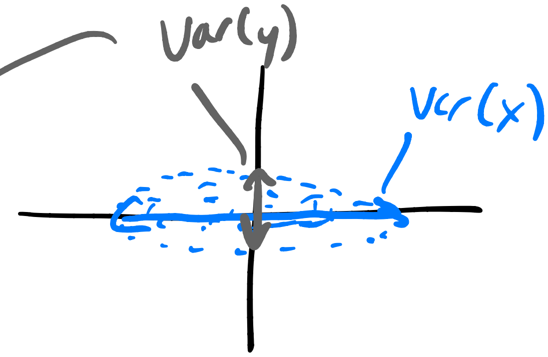
Lecture 17

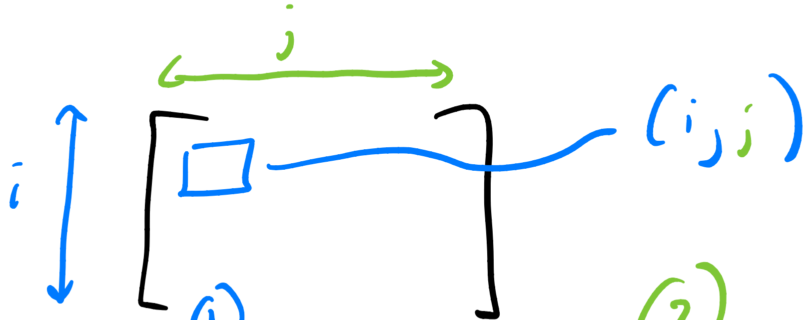
Var(x)

Var(y)

$$\Sigma = \begin{array}{cc} \begin{matrix} x & \;\; y \end{matrix} & \\ \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} & \begin{matrix} x \\ y \end{matrix} \end{array}$$



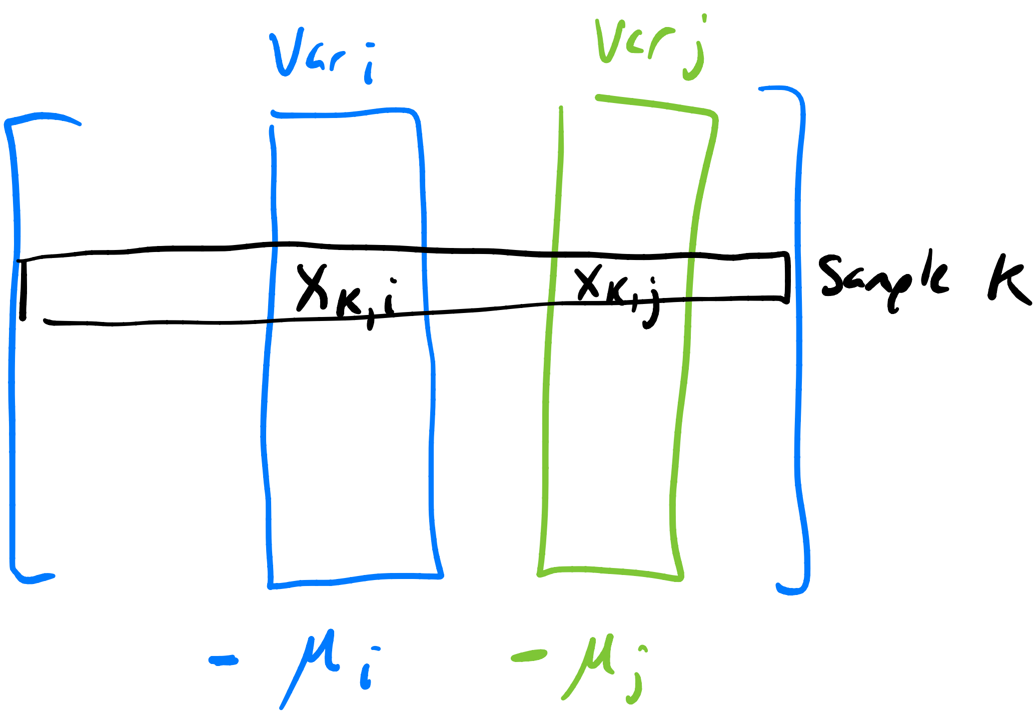
Scalar equation



Sum up over all samples

$$\Sigma(i,j) = \sum_{k=1}^{N} \frac{\overbrace{(x_{k_i} - \mu_i)}^{1} \cdot \overbrace{(x_{k_j} - \mu_j)}^{2}}{N-1}$$

$A =$



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

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

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

$$= \frac{1}{N-1} \sum_{i=1}^{N} (x_i-\mu)^2$$

ordinary variance

$i=0,\ j=1 : \Sigma(i,j)$

$= \Sigma(0,1)$

2D

$$\Sigma = \begin{bmatrix} Var(x) & cov(x,y) \\ cov(y,x) & Var(y) \end{bmatrix}$$

$=$

Vectorized equation:

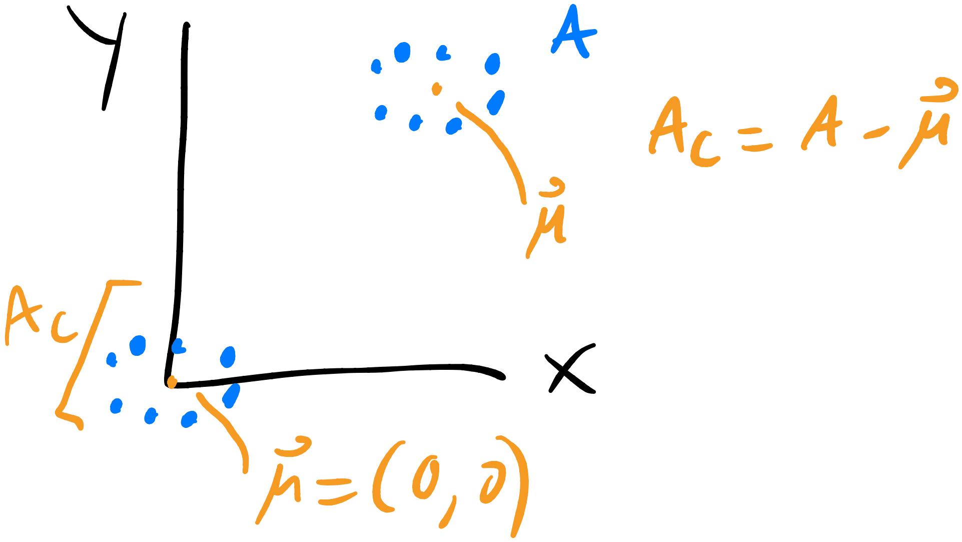
$$\underbrace{\Sigma}_{\text{result shape: }(m,m)} = \frac{1}{N-1} \sum_{k=1}^{N} \underbrace{\underbrace{\left[\underbrace{\left(\vec{x}_k - \overset{\text{means of all cols}}{\vec{\mu}}\right)}_{\text{shape: }(1,m)}.T \right]}_{\text{shape: }(m,1)} \cdot \underbrace{\left(\vec{x}_k - \vec{\mu}\right)}_{(1,m)}}_{}$$

$$A = \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} \vdots & \vdots & \vdots \end{bmatrix} \end{array} \qquad \begin{array}{c} \begin{array}{ccc} \text{var1} & \text{var2} & \text{var3} \end{array} \\ \begin{bmatrix} 1 & 4 & 7 \\ \boxed{2 \quad 5 \quad 8} & & \\ 3 & 6 & 9 \end{bmatrix} \end{array} \begin{array}{l} \text{sample } k \\ \vec{x}_k \end{array}$$

$$- [\mu_1, \mu_2, \mu_3] \qquad - [2, 5, 8]$$

$$A_c = \begin{bmatrix} \vdots & \vdots & \vdots \end{bmatrix} \qquad \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix}$$

$\text{shape}(A_c) = \text{shape}(A)$



Taking data A, subtracting column means $\vec{\mu}$, $\vec{\mu}$
gives us A_c — Centered data ← mean is $\vec{0} = (0, 0 \dots)$

if we have A_c [precomputed]

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

where $A_c = A - \vec{\mu}$ ← column means

Example: Calculate Σ, but 1st A_c

$$A = \begin{array}{ccc} \text{var1} & \text{var2} & \text{var3} \\ \end{array}$$
$$A = \begin{bmatrix} 1 & 2 & -1 \\ 10 & 2 & 1 \\ 3 & 4 & 5 \\ 3 & 2 & 1 \end{bmatrix}$$

$$- [4.25, 2.5, 1.5]$$

$$\vec{\mu} = \left(\frac{1}{4}(1+10+3+3), \frac{1}{4}(2+2+4+2), \frac{1}{4}(-1+1+5+1) \right)$$

$$\vec{\mu} = (4.25, 2.5, 1.5)$$

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

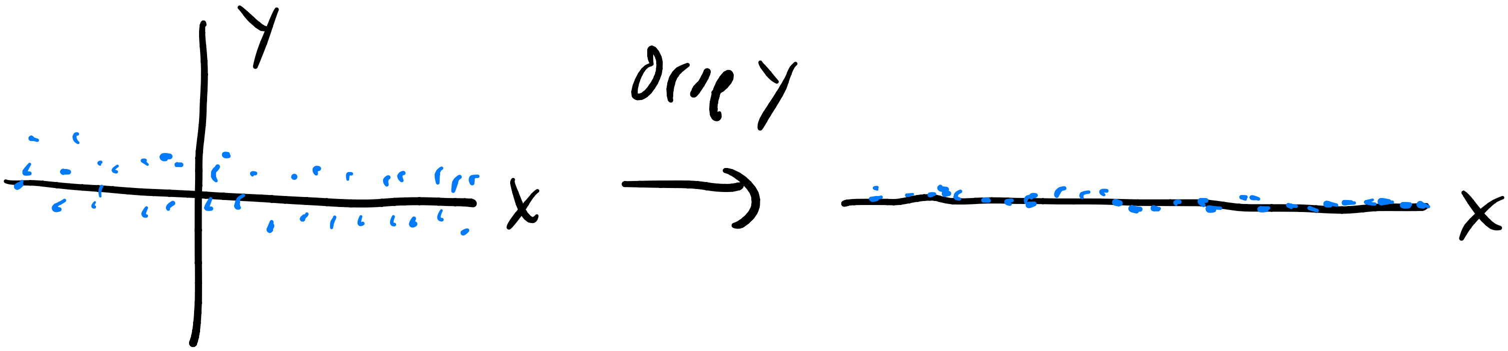
$$A_c = \begin{bmatrix} -3.25 & -1/2 & -2.5 \\ 5.75 & -1/2 & -1/2 \\ -1.25 & 1.5 & 3.5 \\ -1.25 & -1/2 & -1/2 \end{bmatrix}$$

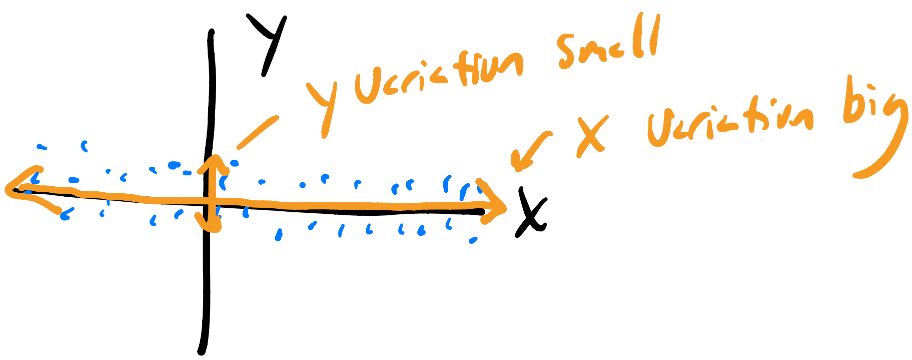
$$\overbrace{\Sigma}^{(3,3)} = \frac{1}{N-1} \overbrace{A_c.T}^{(3,4)} @ \overbrace{A_c}^{(4,3)}$$

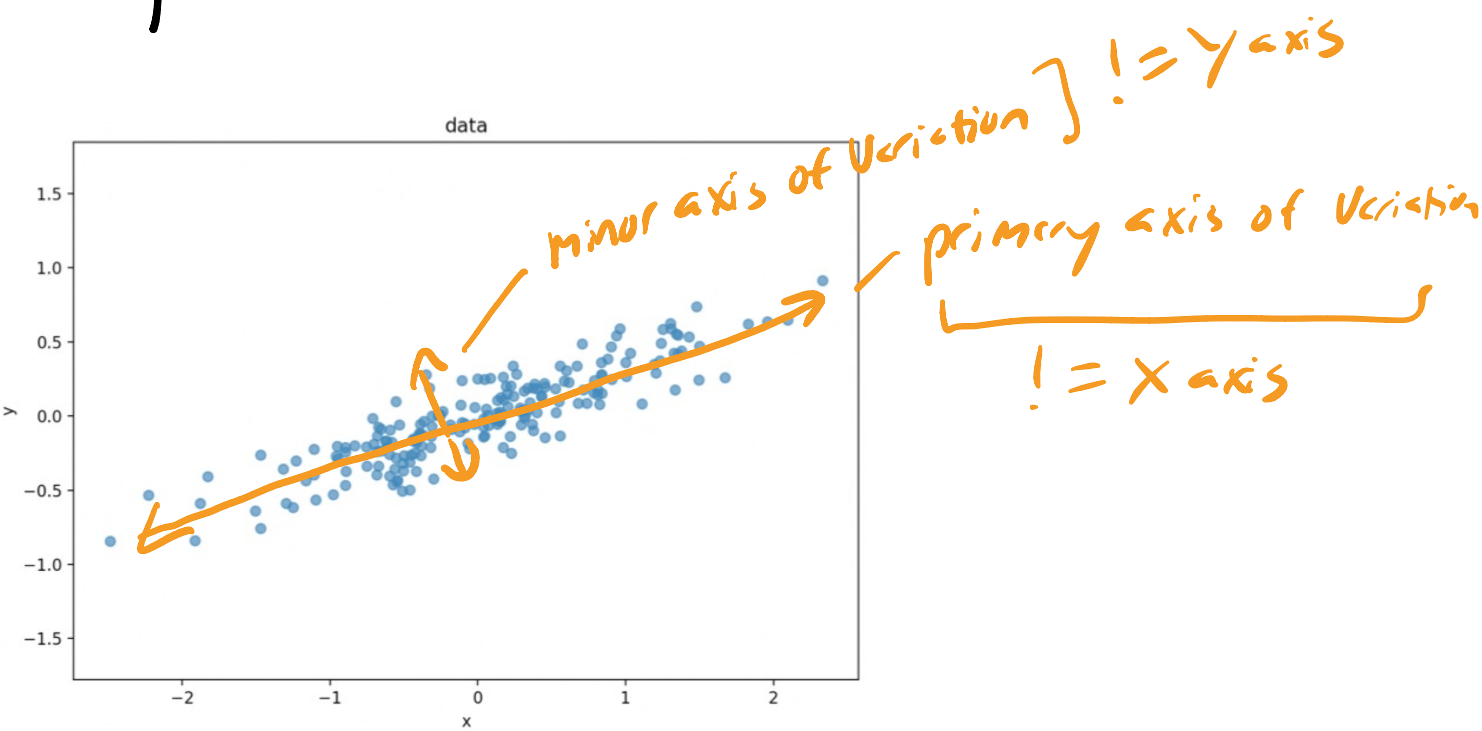
$$\Sigma = \frac{1}{3}\underbrace{\begin{bmatrix} -3.25 & 5.75 & -1.25 & -1.25 \\ -1/2 & -1/2 & 1.5 & -1/2 \\ -2.5 & -1/2 & 3.5 & -1/2 \end{bmatrix}}_{A_c.T}\underbrace{\begin{bmatrix} -3.25 & -1/2 & -2.5 \\ 5.75 & -1/2 & -1/2 \\ -1.25 & 1.5 & 3.5 \\ -1.25 & -1/2 & -1/2 \end{bmatrix}}_{A_c}$$

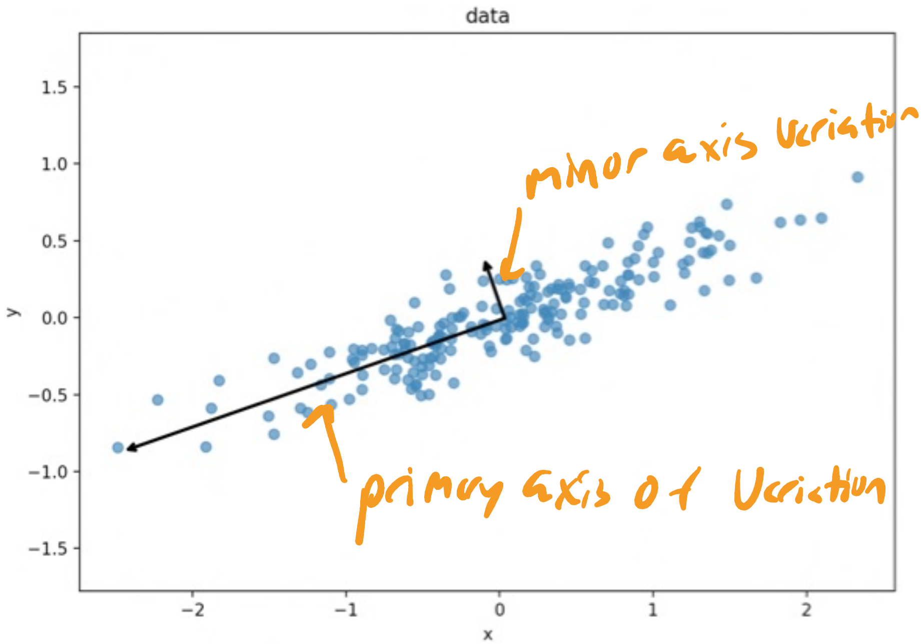
$$= \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 15.8 & -0.83 & 0.5 \\ -0.83 & 1.0 & 2.3 \\ 0.5 & 2.3 & 6.3 \end{bmatrix} \end{array} \begin{array}{c} \\ x \\ y \\ z \end{array}$$

- [ ] Principal component analysis (PCA)

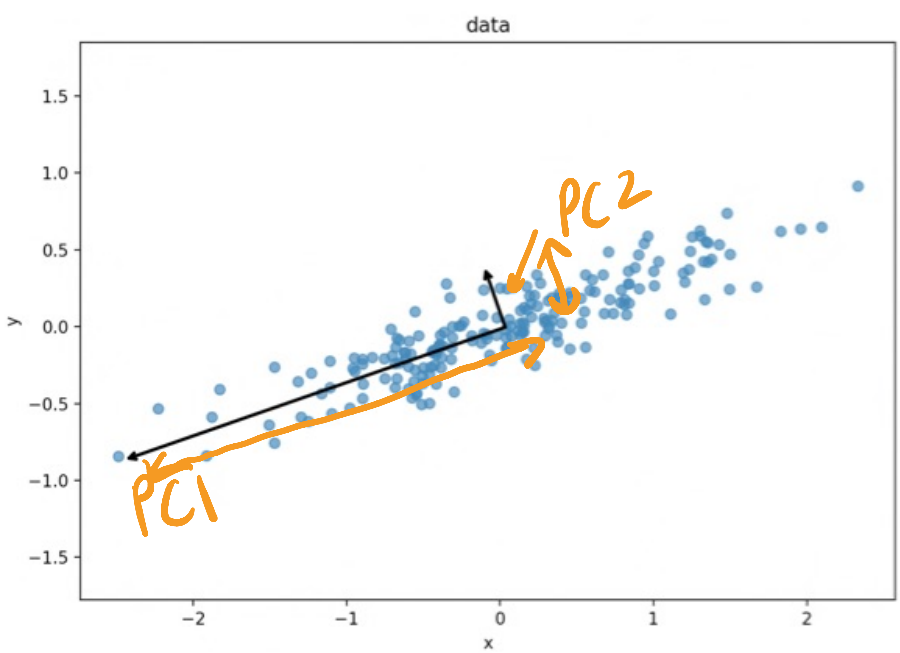








Principal components (PCs): intrinsic axes of variation in data

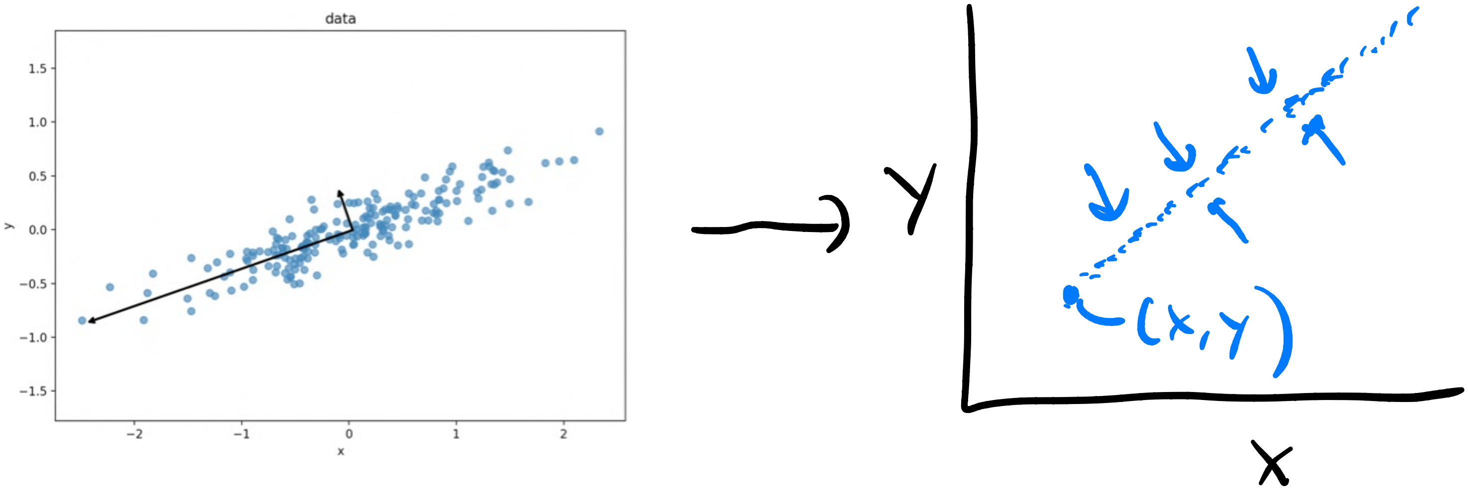


Smaller the PC number bigger the axis of variation.

$\boxed{\text{Goal of PCA}}$

We want to drop PCs intrinsic axes of variation that are insignificant / small

$\Rightarrow$ Do <u>not</u> want to drop data variables


data
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2
−1
0
1
2
x
y
Y
(x, y)
X