PRINCIPAL COMPONENT ANALYSIS (PCA) → Numerical data

Data matrix $A_{N_{obser} \times N_{features}} = \begin{array}{c|c} \begin{pmatrix} 0 & \\ 0 & \\ 0 & \\ 0 & \\ 0 & \end{pmatrix} & \begin{array}{c} obs\,1 \\ \vdots \\ obs\,N \end{array} \\ f_1 \cdots\cdots f_M & \end{array}$

M-dimensional data

$\Big\downarrow$ PCA

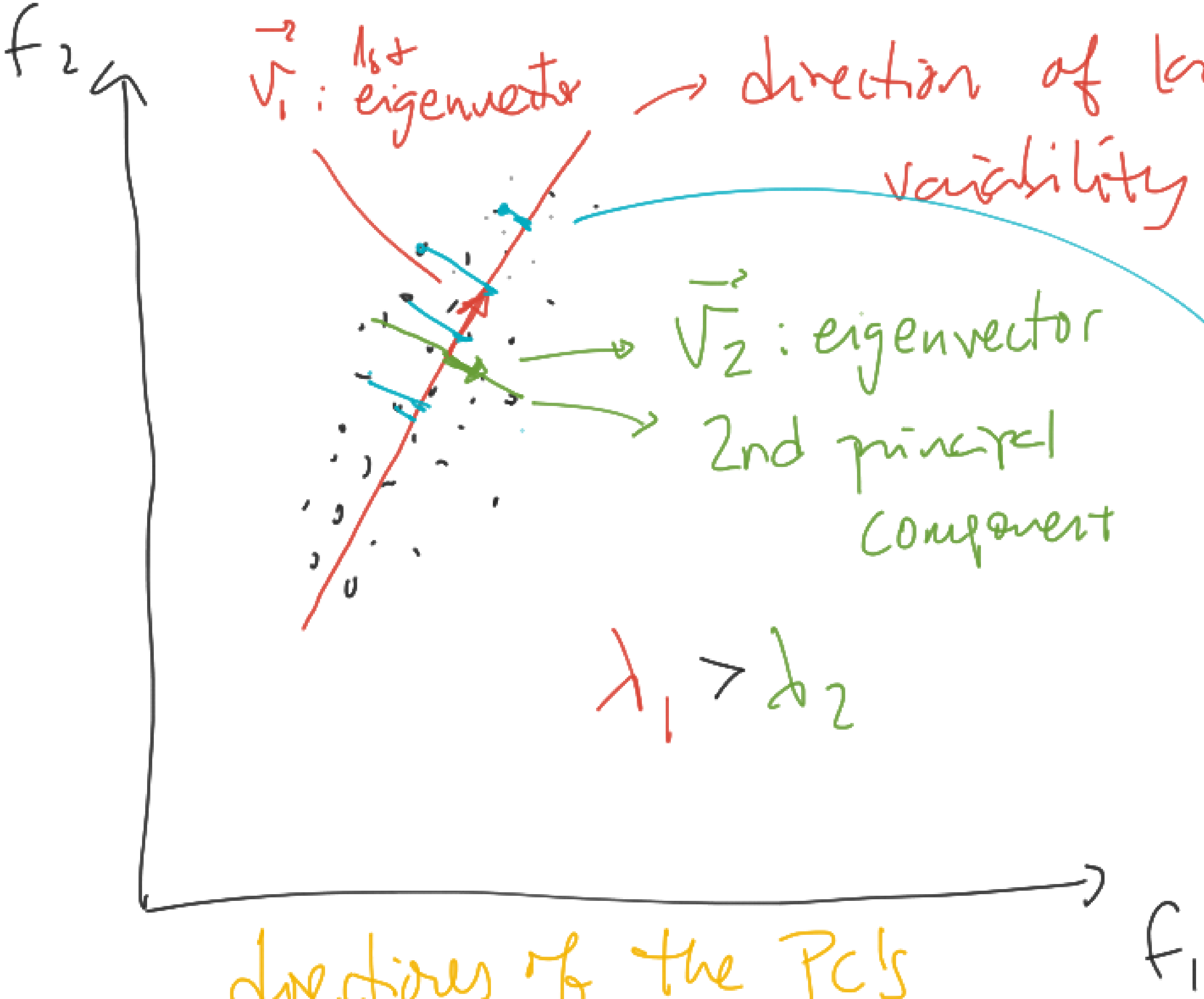
d-dimensional projection $(d \ll M)$

Ex: $M = 2$

(f_1, f_2)

$\downarrow$ PCA

$d = 1$



$\vec{v}_1$: 1st eigenvector → direction of largest variability → 1st principal component

$\vec{v}_2$: eigenvector → 2nd principal component

projection of data into 1st PC

$\lambda_1 > \lambda_2$

directions of the PC's

diagonalize the data covariance matrix C

$\vec{v}_1, \lambda_1, \vec{v}_2, \lambda_2$

variability explained by each PC

$M = 18 \xrightarrow{\text{PCA}} \bar{v}_i, \lambda_i, \; i = 1 \ldots 18$

How do I know how many to retain? d

rule: Keep as many PC's as needed to explain $\geq 95\%$ total variability data

61%   10%

$$\frac{\lambda_1}{\sum_i \lambda_i} \quad \frac{\lambda_2}{\sum \lambda_i} \quad \frac{\lambda_3}{\sum \lambda_i} \quad \ldots \quad \frac{\lambda_{18}}{\sum \lambda_i}$$

fraction of variability explained by PC1

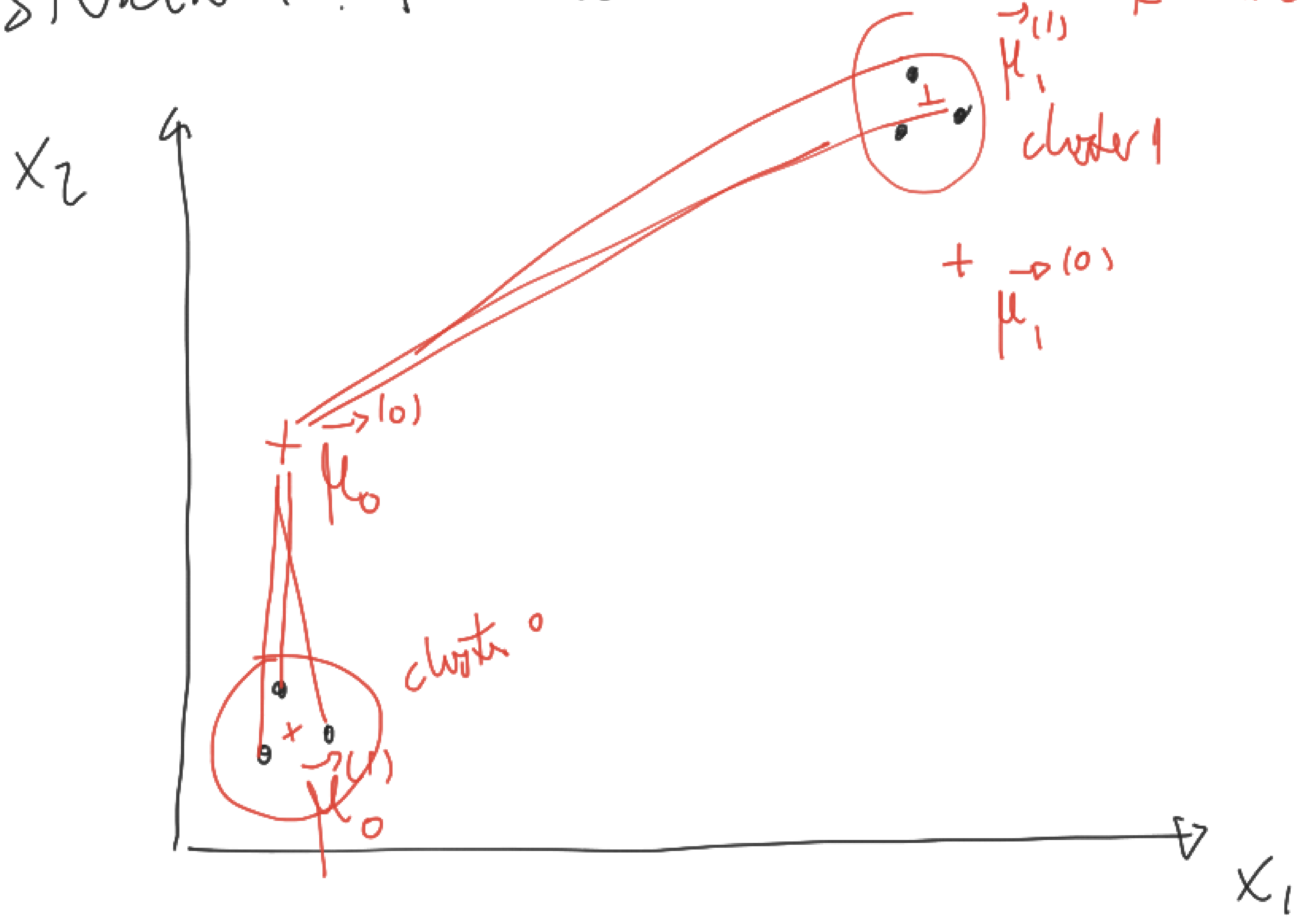
cumsum $= 71\%$

cumsum $\geq 95\%$

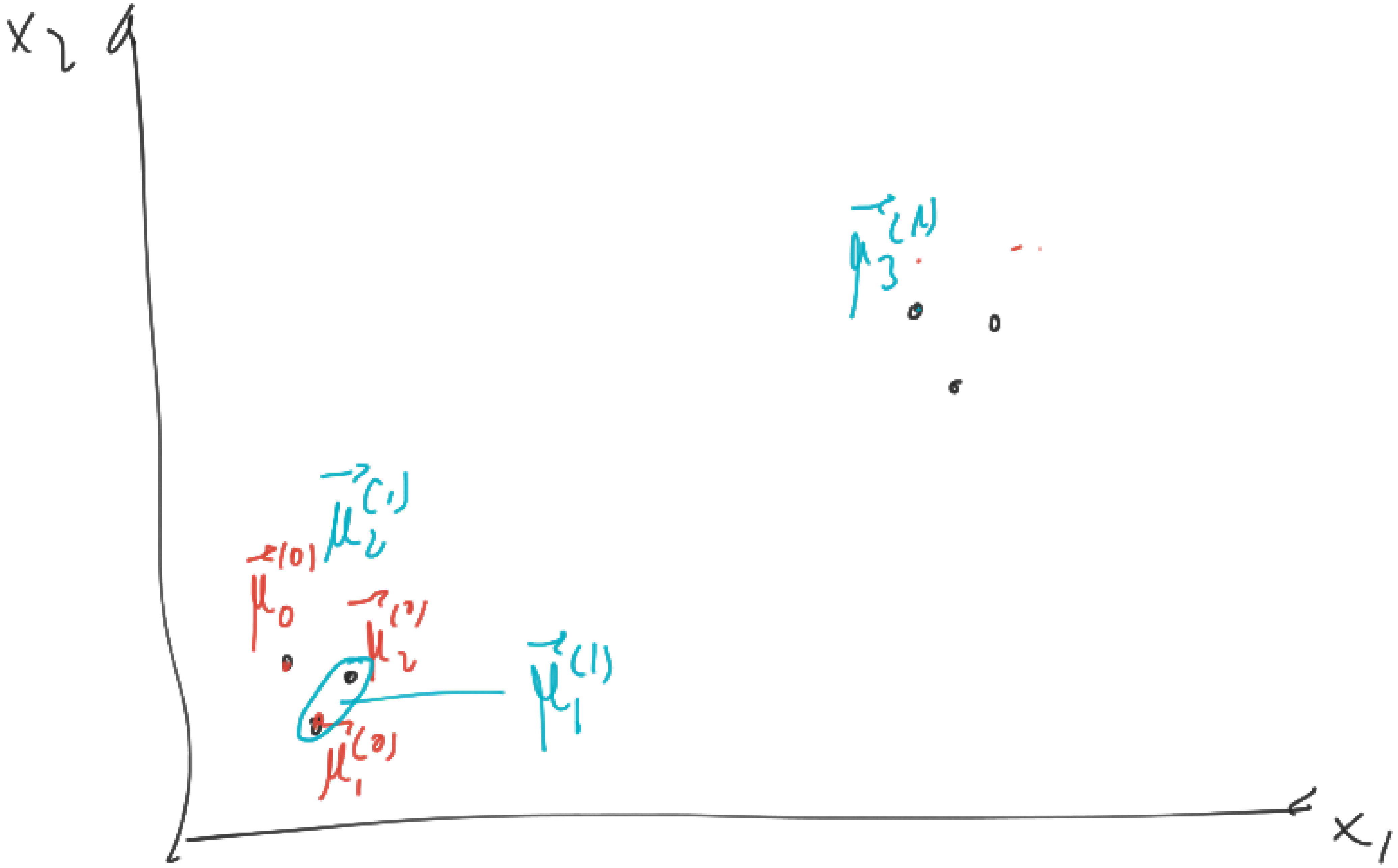
d

CLUSTERING : k-means

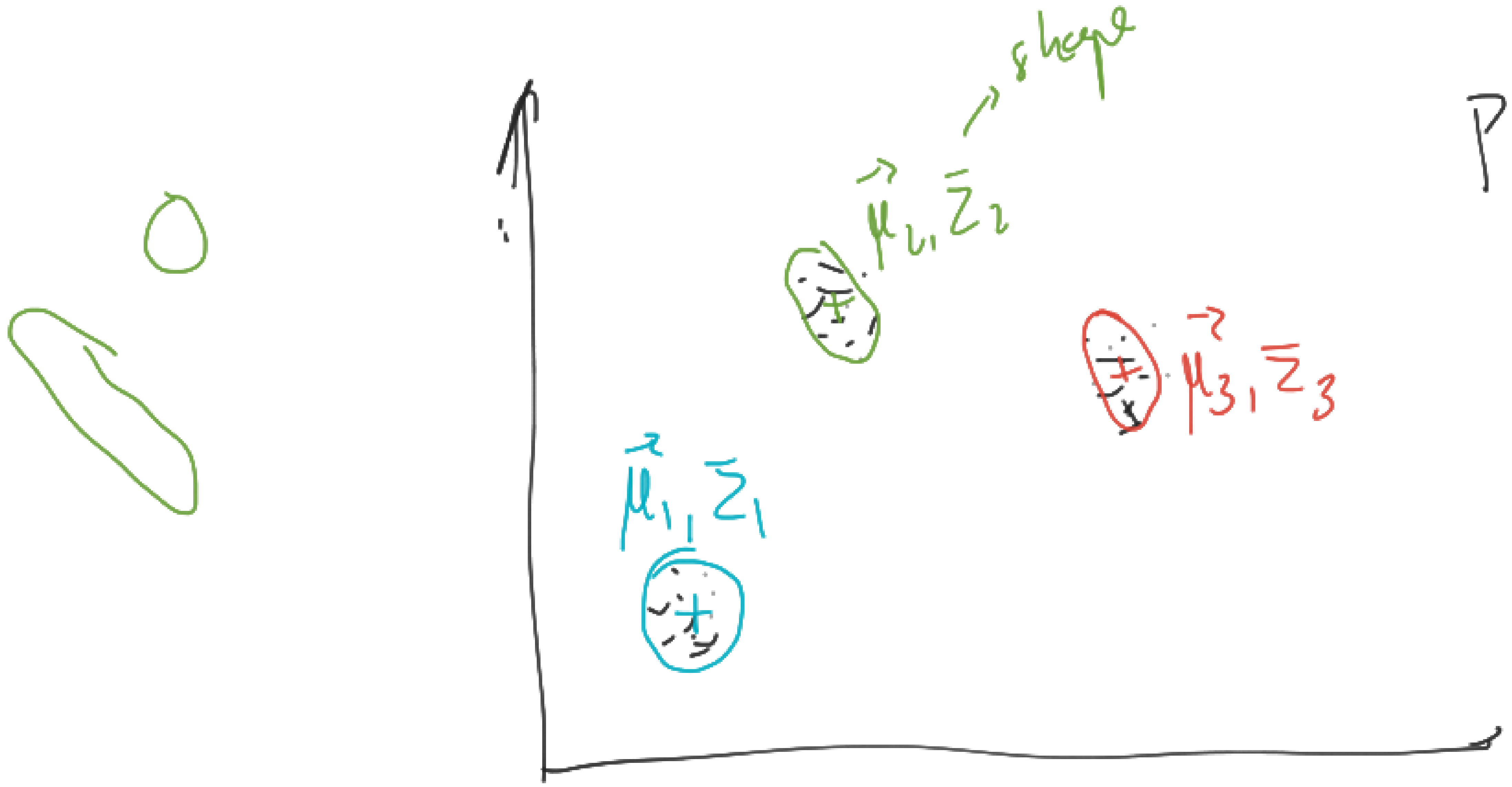
k : #clusters $\rightarrow 2$



Agglomerative clustering

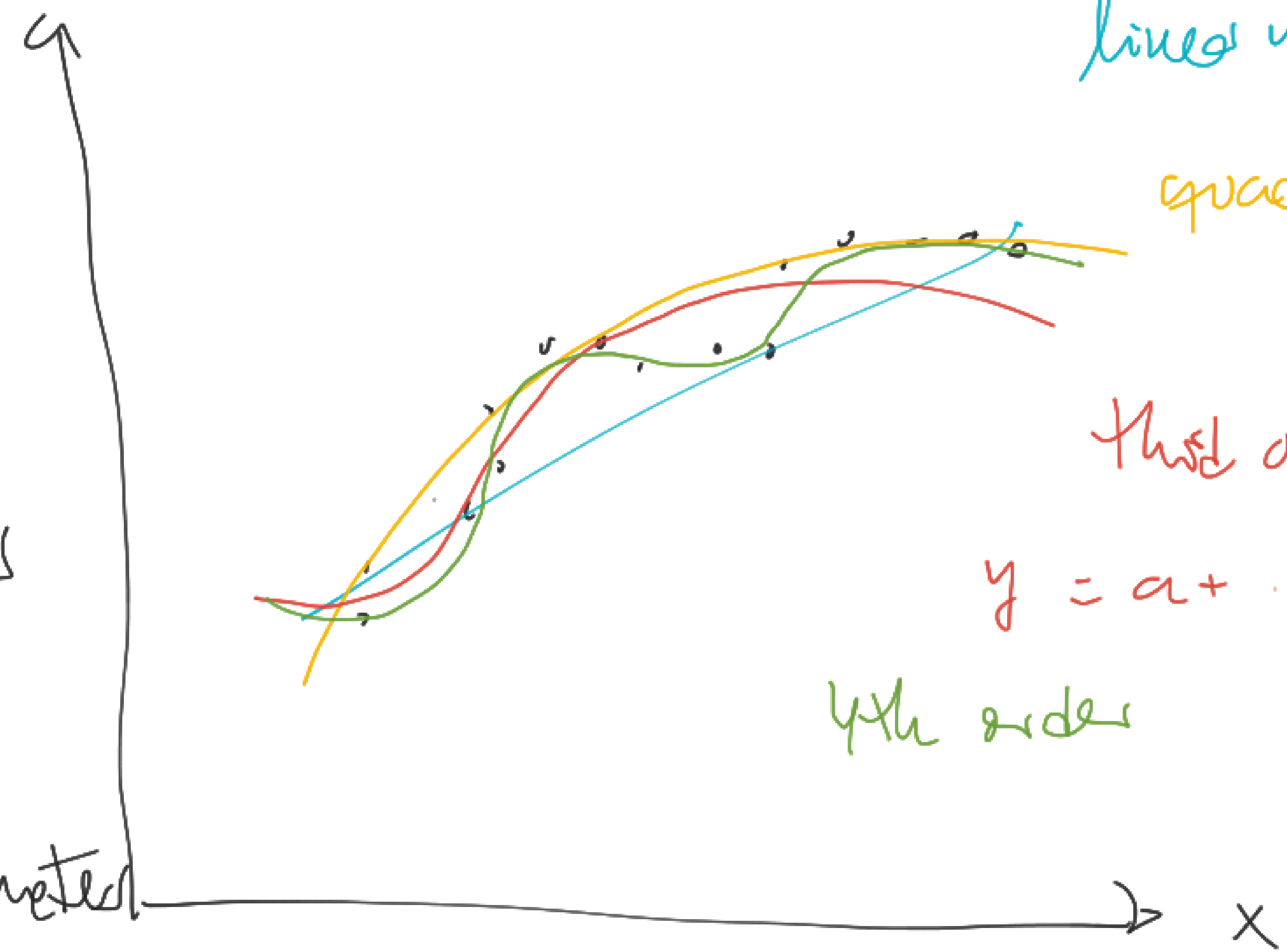


GMM – model statistics



$$p(\vec{x}) = \sum_{i=1}^{G} \pi_i \cdot N(\vec{\mu}_i, \Sigma_i)$$

Model selection → Parsimony criterion : ERROR vs COMPLEXITY



linear model $y = a + bx$

quadratic

$y = a + bx + cx^2$

third order model

$y = a + \ldots$

4th order

ERROR → sum of residuals (MSE)

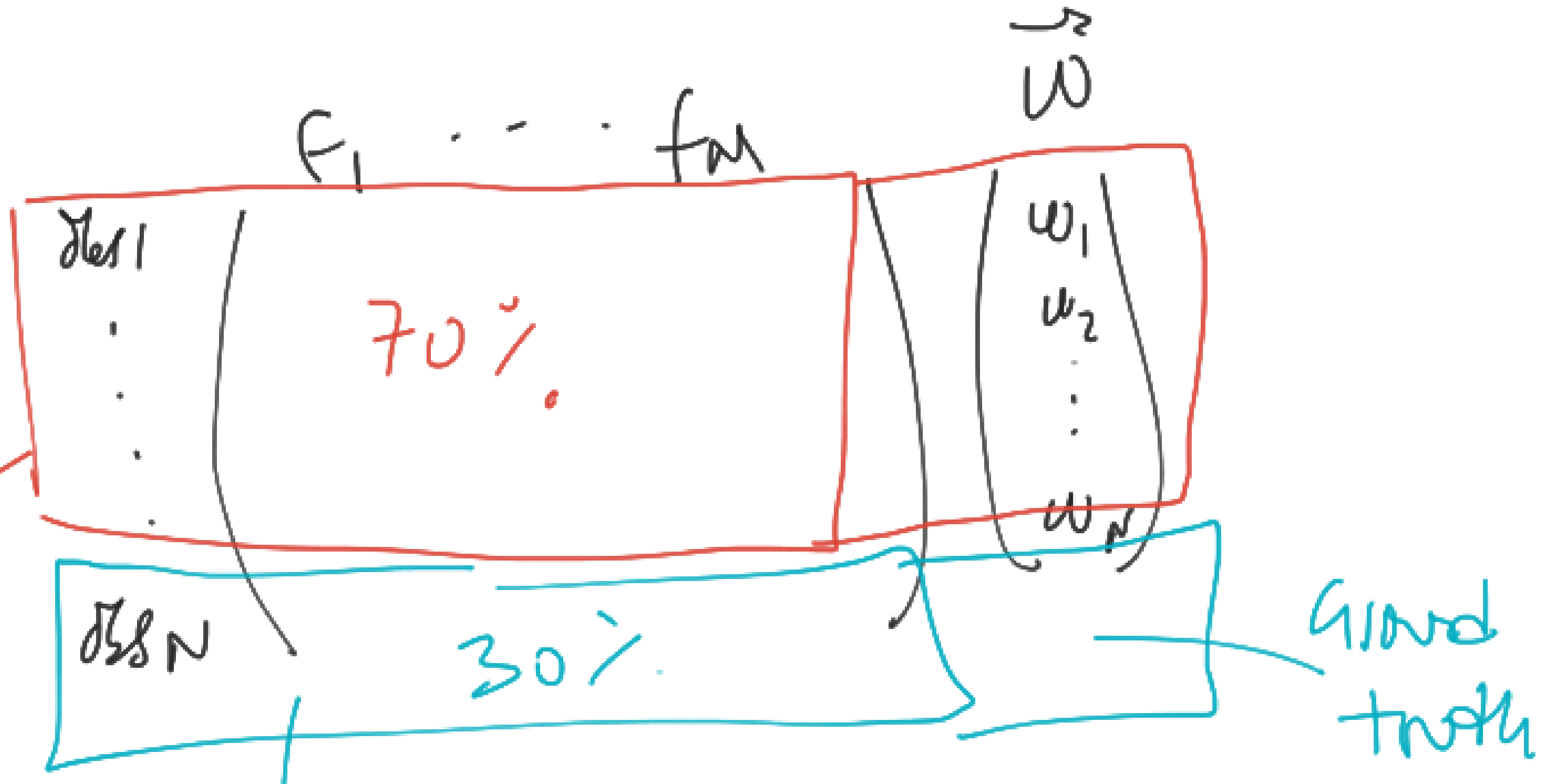
COMPLEXITY → # parameter

MODEL	MSE	#parameters	PARSIMONY INDEX
1st	10	2	70
2nd	1.6	3	60 → Best!
3rd	0.2	4	63
4th	0.01	5	68

SUPERVISED CLASSIFICATION

data $A_{N_{obs} \times M_{features}}$

class labels $\vec{w}$ $N_{obs} \times 1$



TEST MODEL

TRAIN MODEL

TRAIN A LDA classifier: training data

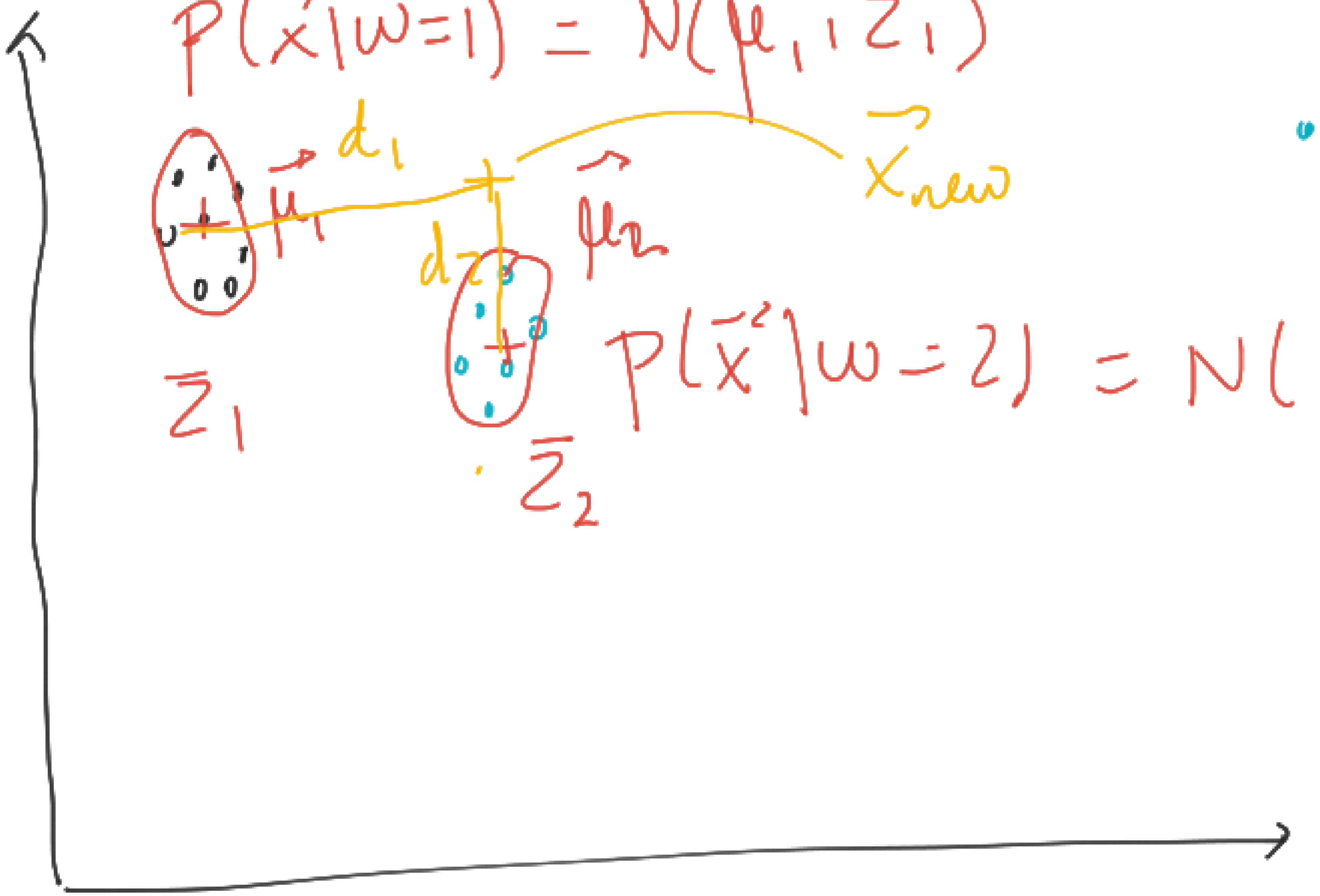
↓

sample estimates of $\vec{\mu}_1, \vec{\mu}_2$, Σ_1, Σ_1 from the training samples.

- class $\omega = 1$
- class $\omega = 2$

$$P(\vec{x} \mid \omega = 1) = N(\vec{\mu}_1, \Sigma_1)$$

$$P(\vec{x} \mid \omega = 2) = N(\vec{\mu}_2, \Sigma_2)$$



KNN