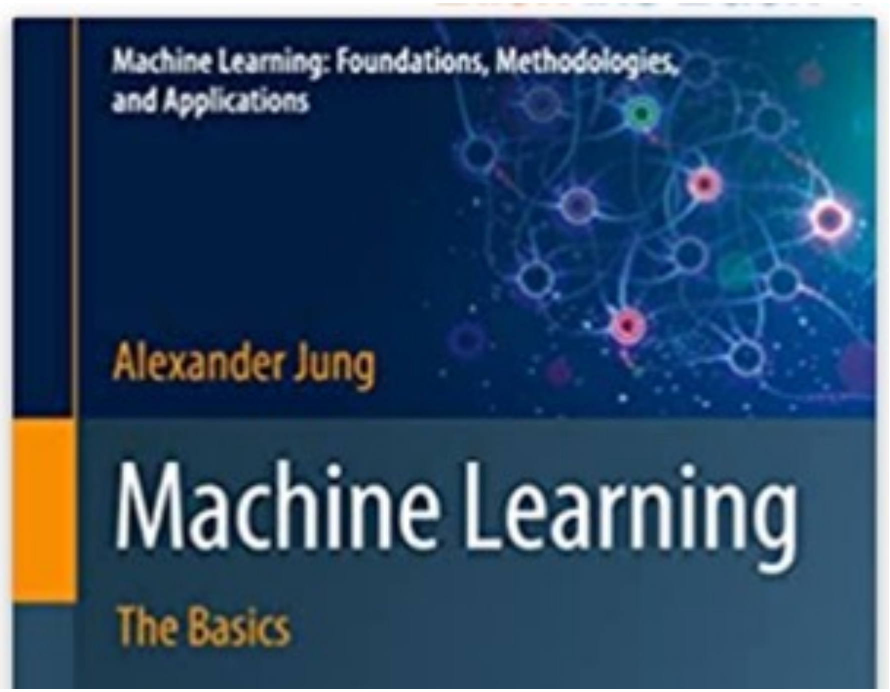
Soft Clustering

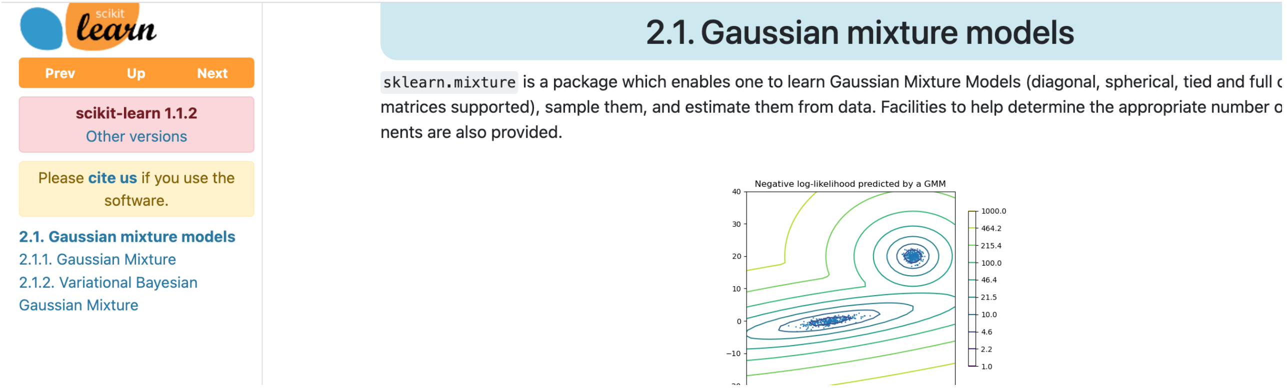
Alex(ander) Jung
Assistant Professor for Machine Learning
Department of Computer Science
Aalto University

Reading.

Sec. 8.2. of https://mlbook.cs.aalto.fi



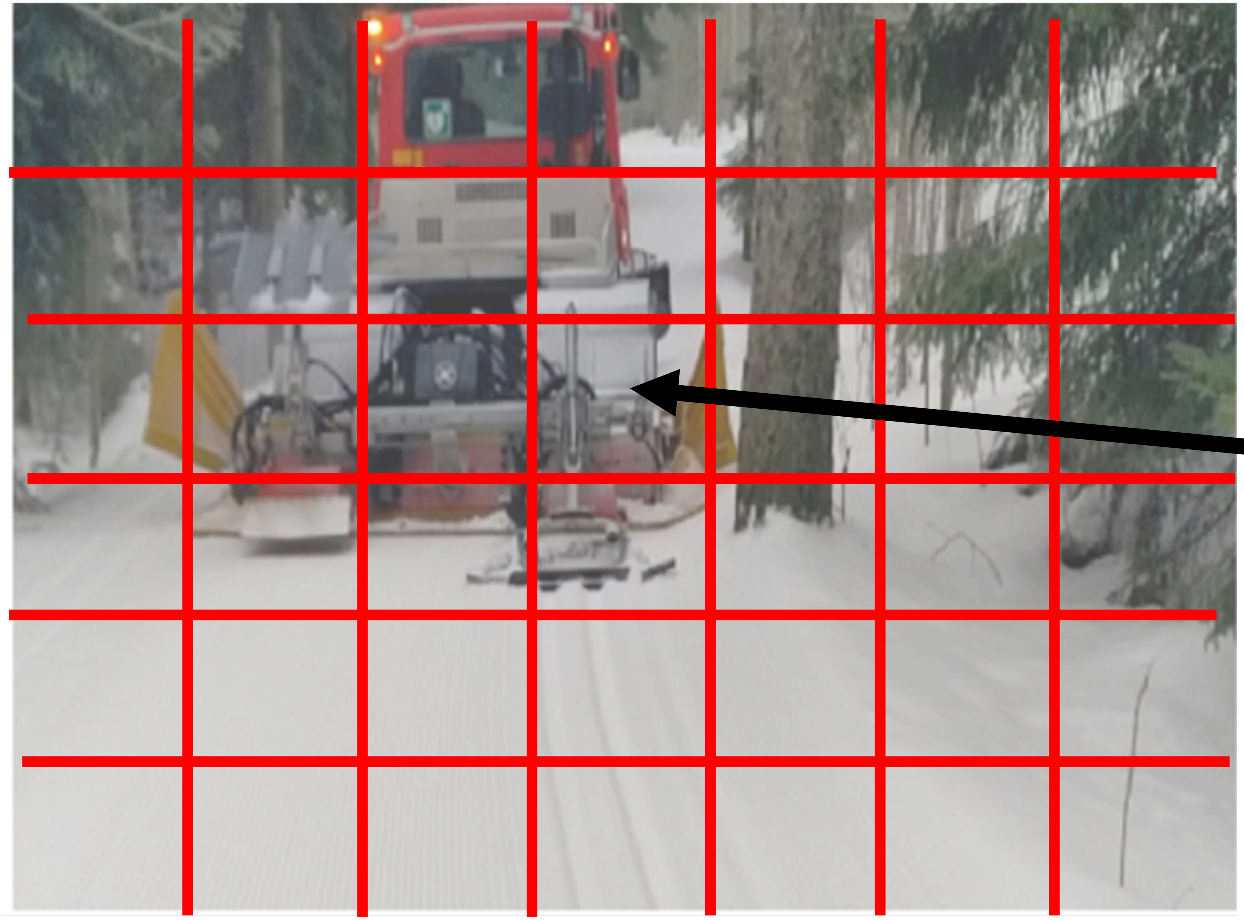
https://scikit-learn.org/stable/modules/mixture.html#gmm



Learning Goals

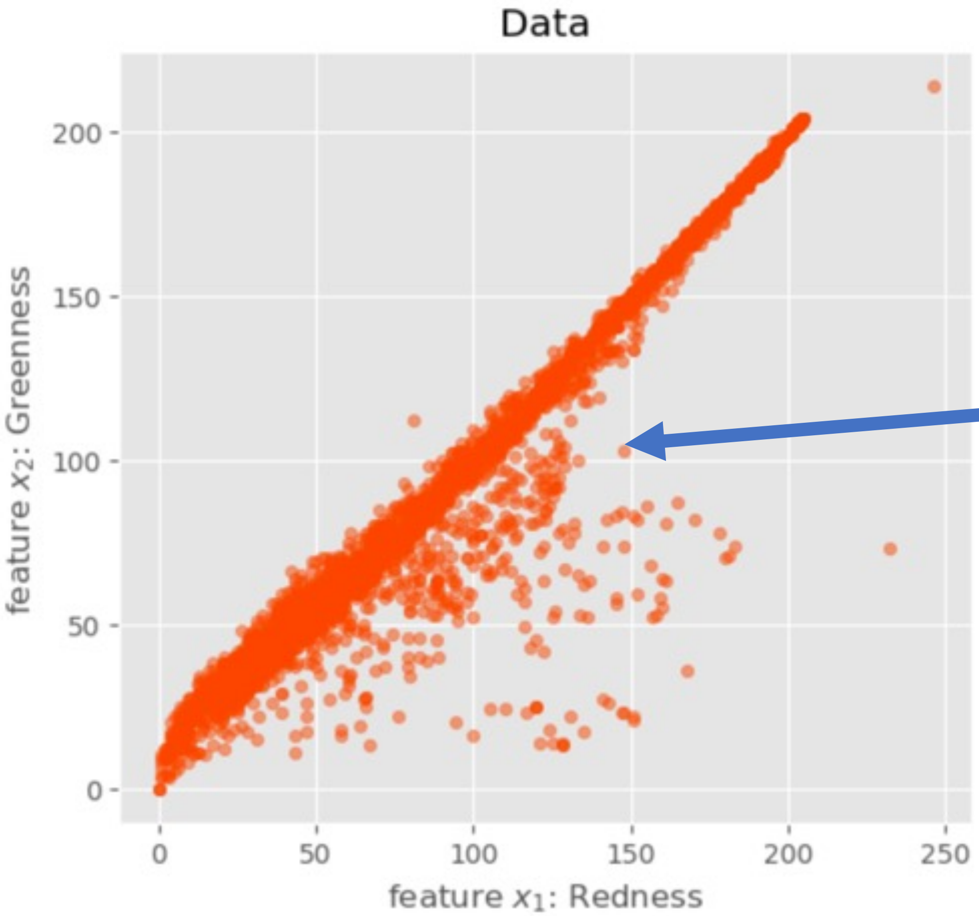
- basic idea of soft clustering
- a soft clustering algorithm
- probabilistic interpretation of algorithm
- how to choose number of clusters

Dataset = Set of Image Patches



data point

Using Two Features (Red+Green)



Hard- vs. Soft-Clustering

Output of k-means (Last Lecture)

Output of Soft-Clustering (Today!)

Soft Clustering

- datapoints $\left(\boldsymbol{x}^{(1)}, \boldsymbol{y}^{(1)}\right), \ldots, \left(\boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)}\right)$
- i-th datapoint characterized by n features

$$\boldsymbol{x}^{(i)} = \left(x_1^{(i)}, \ldots, x_n^{(i)}\right)$$

- i-th datapoint characterized by k label values

$$\boldsymbol{y}^{(i)} = \left(y_1^{(i)}, \ldots, x_k^{(i)}\right)$$

Degree of Belonging

- i-th datapoint characterized by k label values

$$\mathbf{y}^{(i)} = \left(y_1^{(i)}, \dots, x_k^{(i)}\right)$$

- $y_1^{(i)}$ degree of i-th datapoint belonging to cluster 1
- $y_2^{(i)}$ degree of i-th datapoint belonging to cluster 2
- …
- $y_k^{(i)}$ degree of i-th datapoint belonging to cluster k

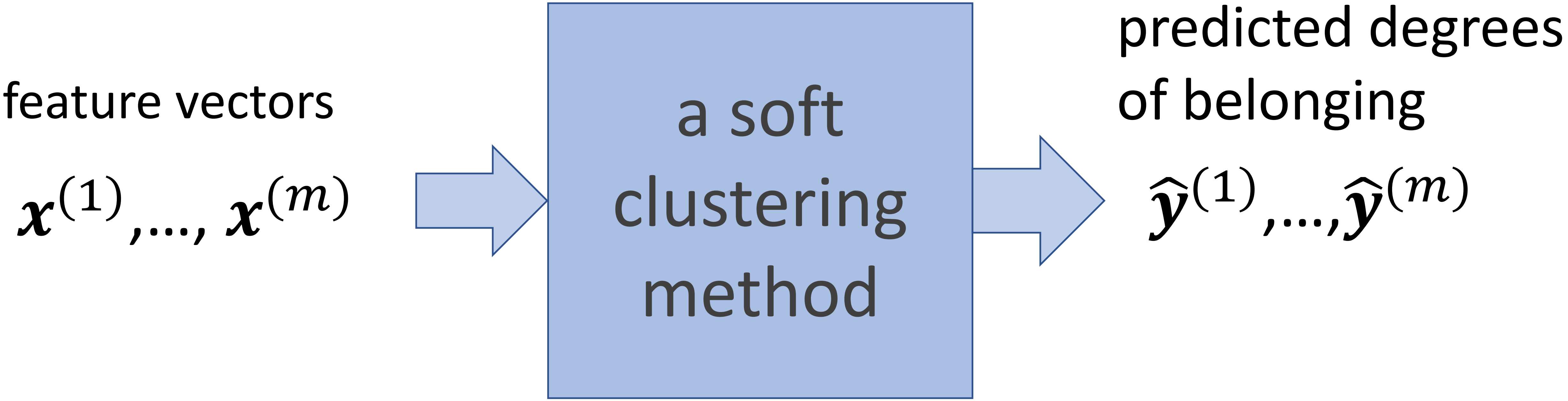
Probabilistic Interpretation

- $y_c^{(i)}$ degree of i-th datapoint belonging to cluster c
- Interpret $y_c^{(i)}$ as probability

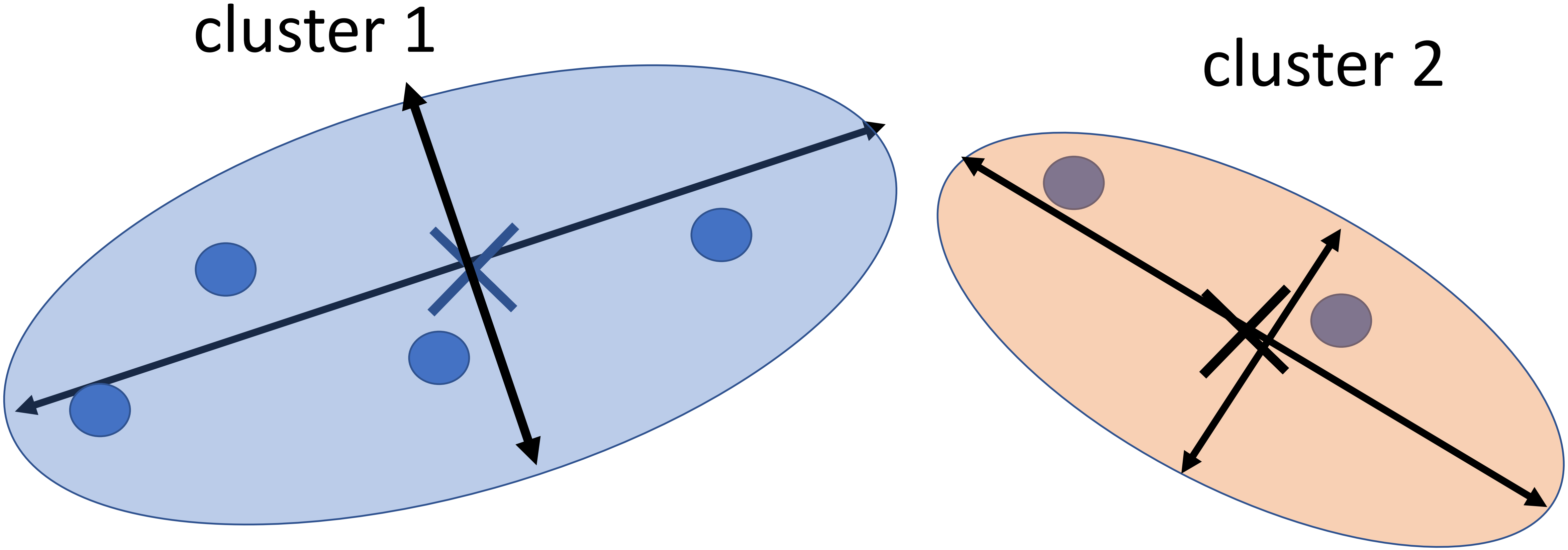
P(“ i-th datapoint belongs to cluster c”)

- $y_c^{(i)}$ can be any number between 0 and 1 (e.g., $y_c^{(i)}$ =0.33)
- $\sum_{c=1}^{k} y_c^{(i)} = 1$ (i-th datapoint must belong to some cluster)
- hard clustering requires $y_c^{(i)}$ is either 0 or 1

Soft Clustering Methods

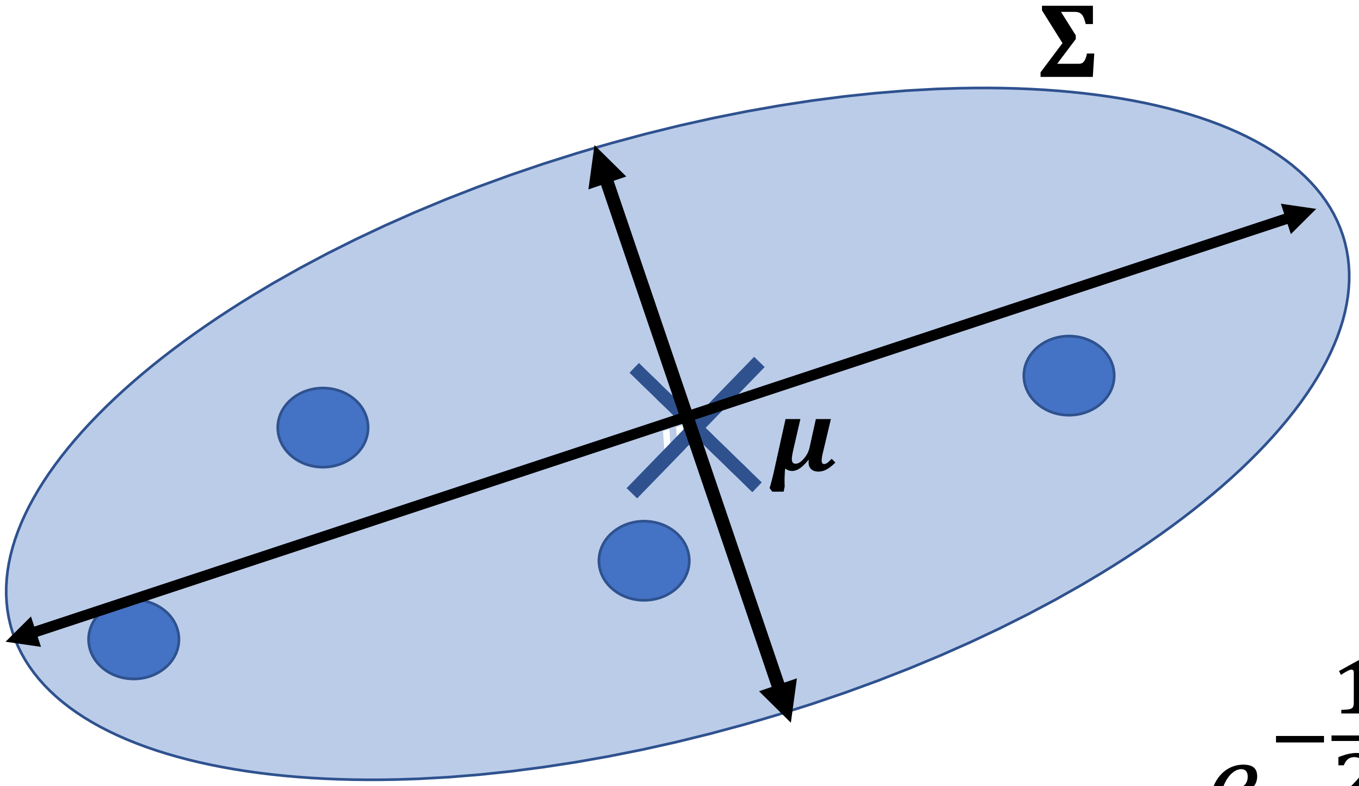


Represent Clusters by Gaussians



Gaussian mixture model (GMM)

Gaussian Distribution

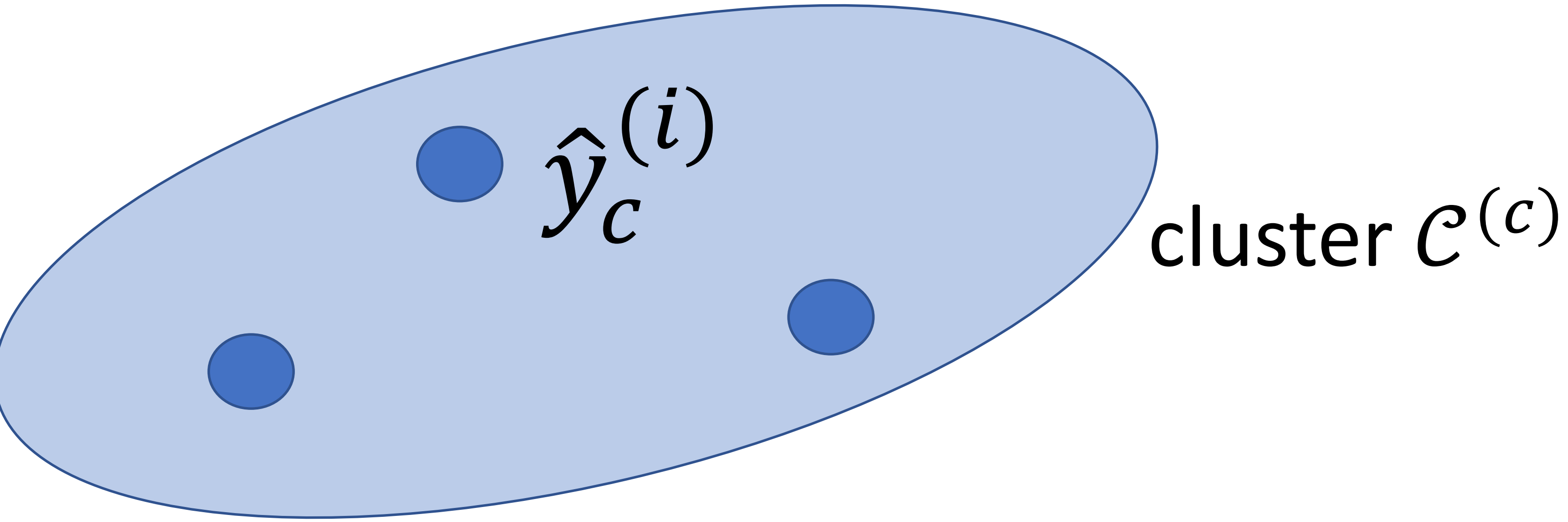


mean vector $\boldsymbol{\mu}$

covariance matrix $\boldsymbol{\Sigma}$

$$p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{\sqrt{(2\pi)^n\det(\boldsymbol{\Sigma})}}$$

Cluster Spread



$$\frac{1}{m^{(c)}}\sum_{i=1}^{m}\hat{y}_c^{(i)}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}^{(c)}\right)^{T}\left(\boldsymbol{\Sigma}^{(1)}\right)^{-1}\left(\mathbf{x}^{(i)}-\boldsymbol{\mu}^{(c)}\right)$$

effective cluster size $m^{(c)} := \sum_{i=1}^{m}\hat{y}_c^{(i)}$

Update Cluster Mean and Covariance

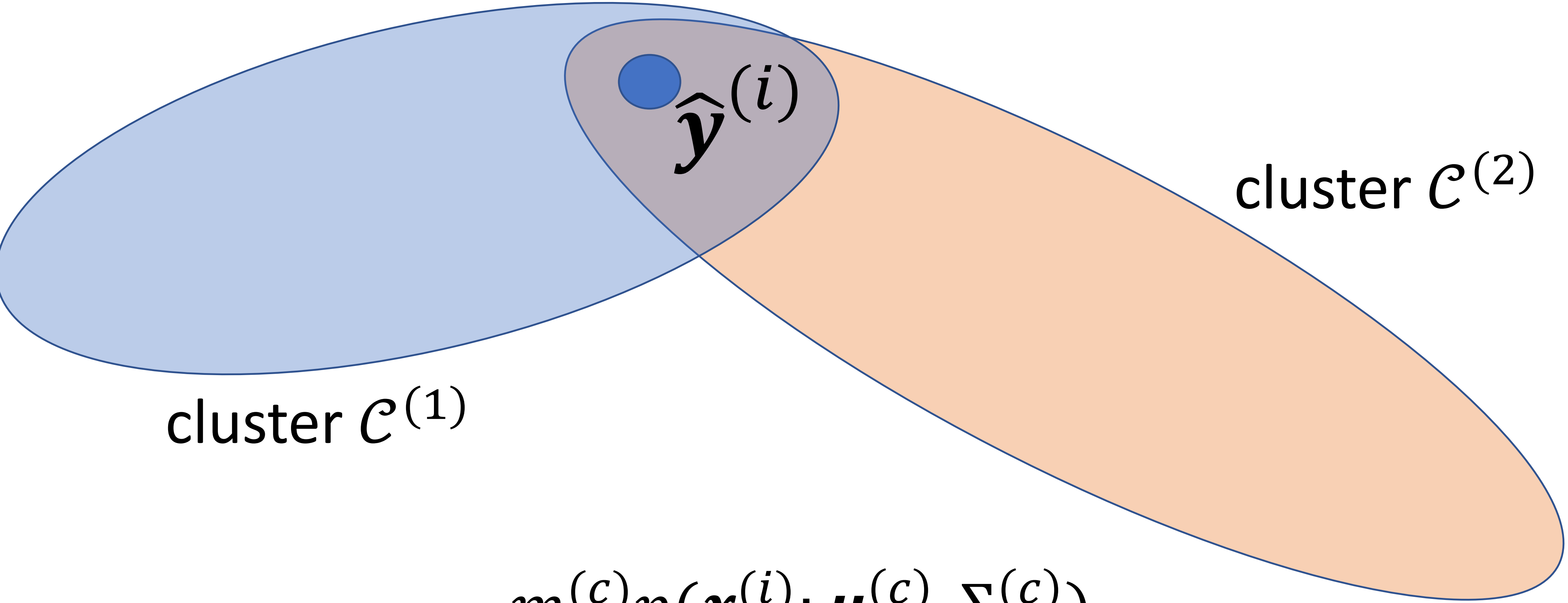
for given (soft) cluster assignments $\hat{y}_c^{(i)}$ chose cluster means and cov. to min. cluster spreads

$$\boldsymbol{\mu}^{(c)} := \frac{1}{m^{(c)}} \sum_{i=1}^{m} \hat{y}_c^{(i)} \, \boldsymbol{x}^{(i)}$$

for all c =1,…,k

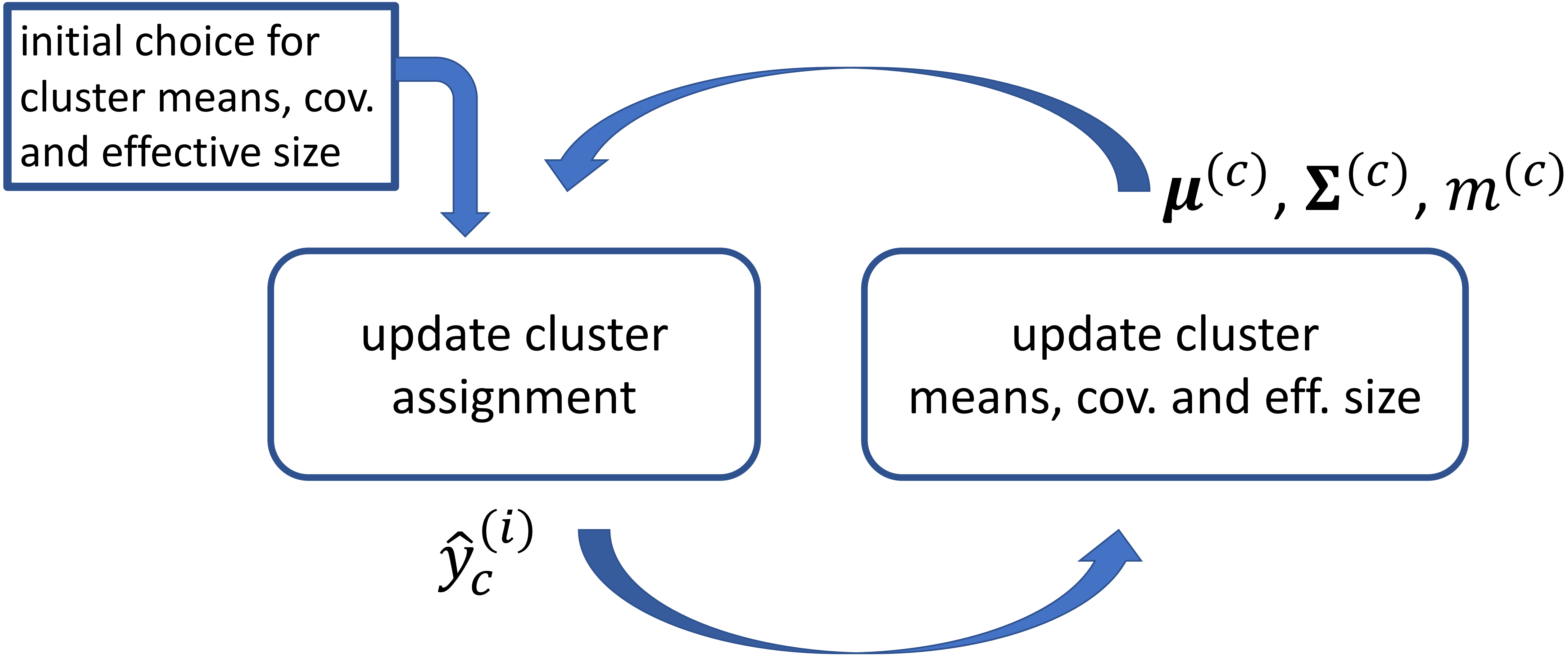
$$\boldsymbol{\Sigma}^{(c)} := \frac{1}{m^{(c)}} \sum_{i=1}^{m} \hat{y}_c^{(i)} \left(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}^{(c)}\right)\left(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}^{(c)}\right)^T$$

Cluster Assignment Update



$$\hat{y}_c^{(i)} := \frac{m^{(c)} p(\boldsymbol{x}^{(i)}; \boldsymbol{\mu}^{(c)}, \Sigma^{(c)})}{\sum_{c'=1}^{k} m^{(c')} p(\boldsymbol{x}^{(i)}; \boldsymbol{\mu}^{(c')}, \Sigma^{(c')})}$$

A Soft-Clustering Algorithm



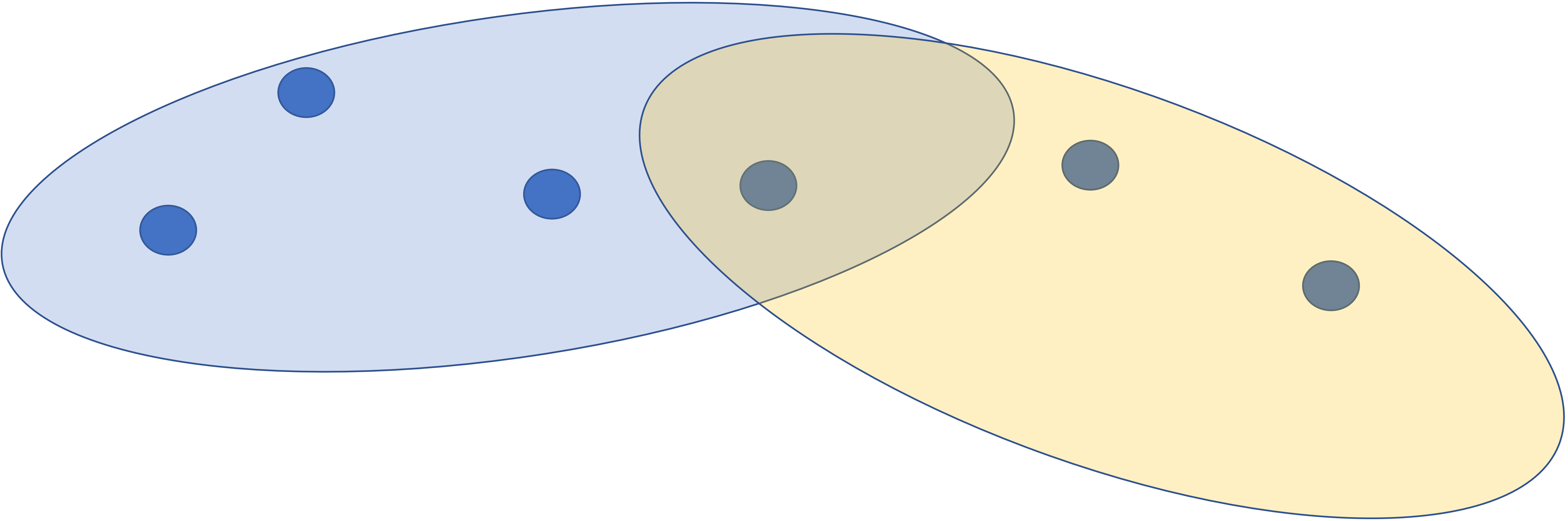
A Soft-Clustering Algorithm

- Input: $\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(m)}, k, \{\boldsymbol{\mu}^{(c)}, \boldsymbol{\Sigma}^{(c)}, m^{(c)}\}$

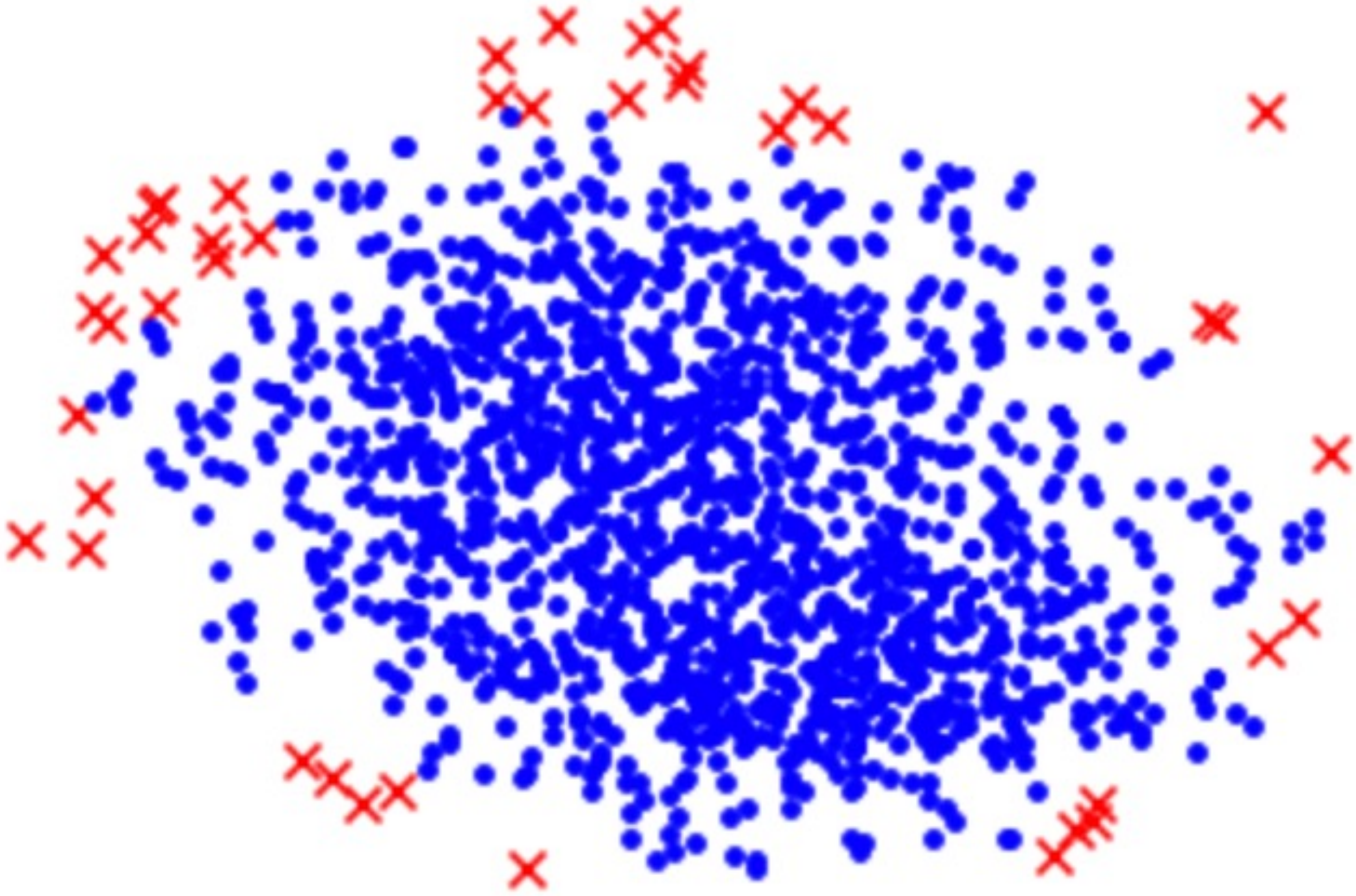
1. update soft cluster assignments $\hat{y}_c^{(i)}$
2. update cluster params $\boldsymbol{\mu}^{(c)}, \boldsymbol{\Sigma}^{(c)}, m^{(c)}$
3. go to 1. unless "finished"

- Output: $\hat{y}_c^{(i)}, \boldsymbol{\mu}^{(c)}, \boldsymbol{\Sigma}^{(c)}, m^{(c)}$

Typical Cluster Shapes



this cluster structure still out of reach!



When to Stop?

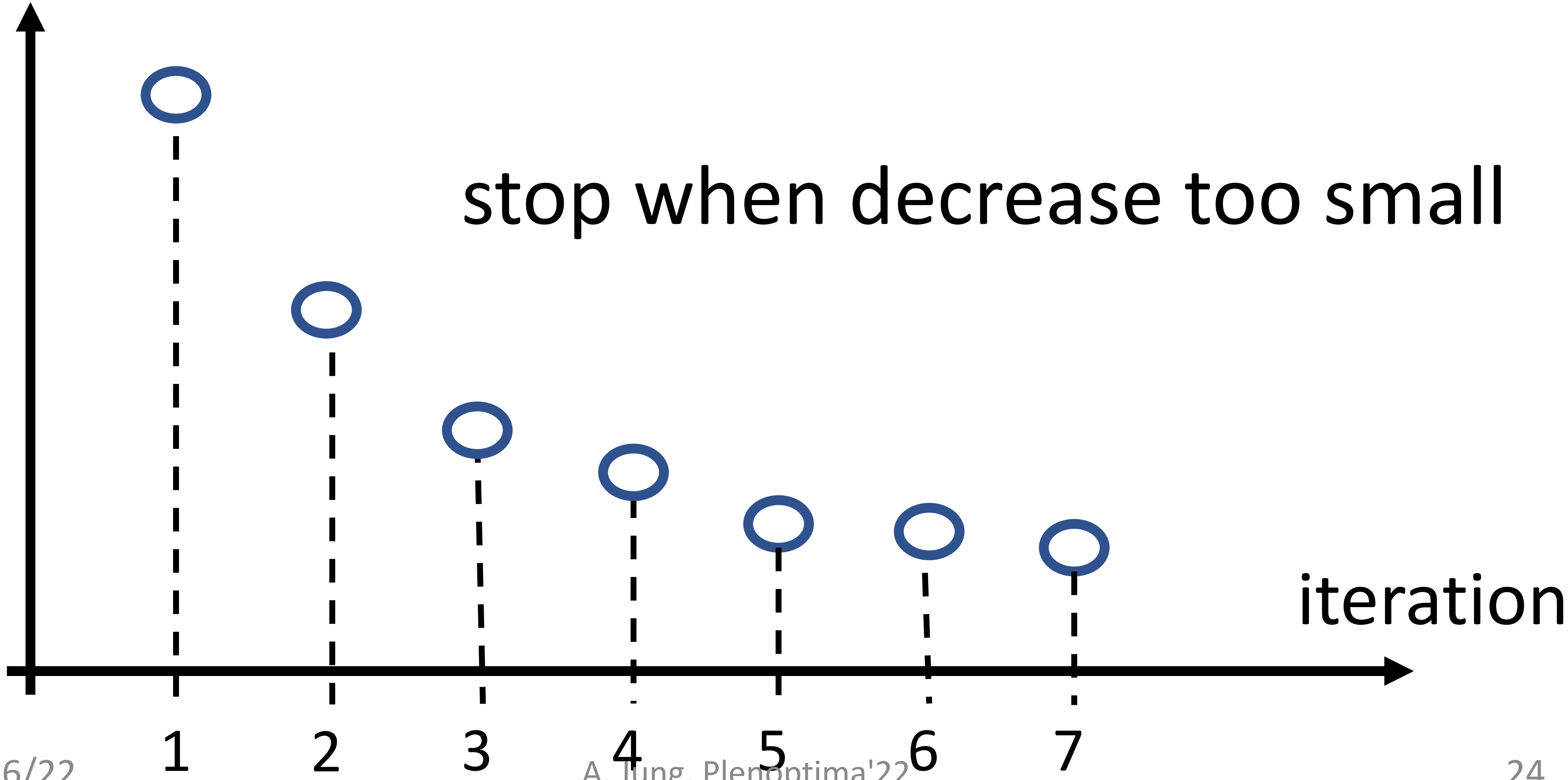
Soft-Clustering Error

$$\mathcal{E}\left(\{\boldsymbol{\mu}^{(c)}\},\{\boldsymbol{\Sigma}^{(c)}\},\{m^{(c)}\}\ \right) :=$$

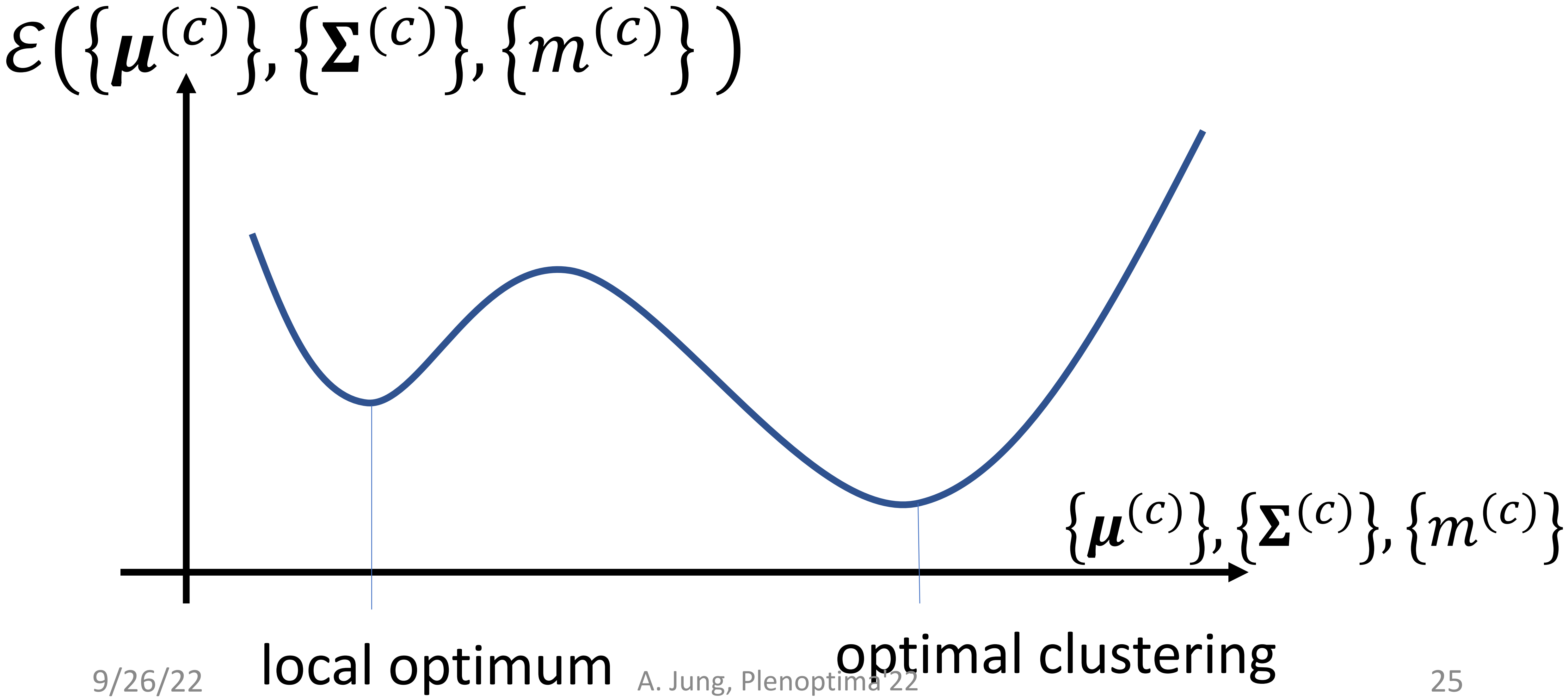
$$-\sum_{i=1}^{m}\log\sum_{c=1}^{k}\frac{m^{(c)}}{m}\,p\left(\boldsymbol{x}^{(i)};\boldsymbol{\mu}^{(c)},\boldsymbol{\Sigma}^{(c)}\right)$$

this is negative logarithm of probability to "see" datapoints under Gaussian mixture model

$$\mathcal{E}\left(\{\boldsymbol{\mu}^{(c)}\},\{\boldsymbol{\Sigma}^{(c)}\},\{m^{(c)}\}\right)$$



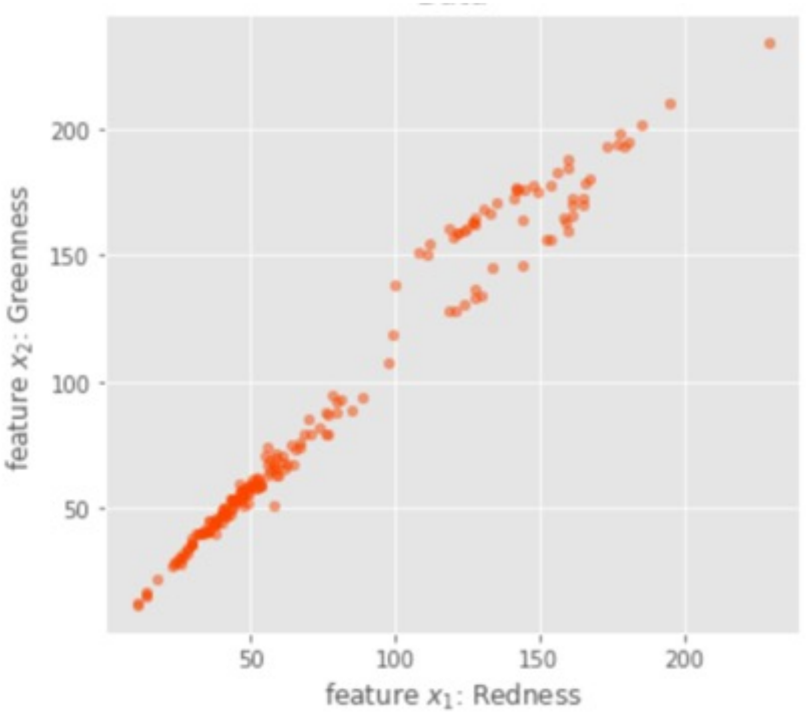
Non-Convexity of Soft-Clustering Error



Initialization is Crucial

- soft clustering depends crucially on init. means
- repeat several times with different init.

How to choose number k of clusters?

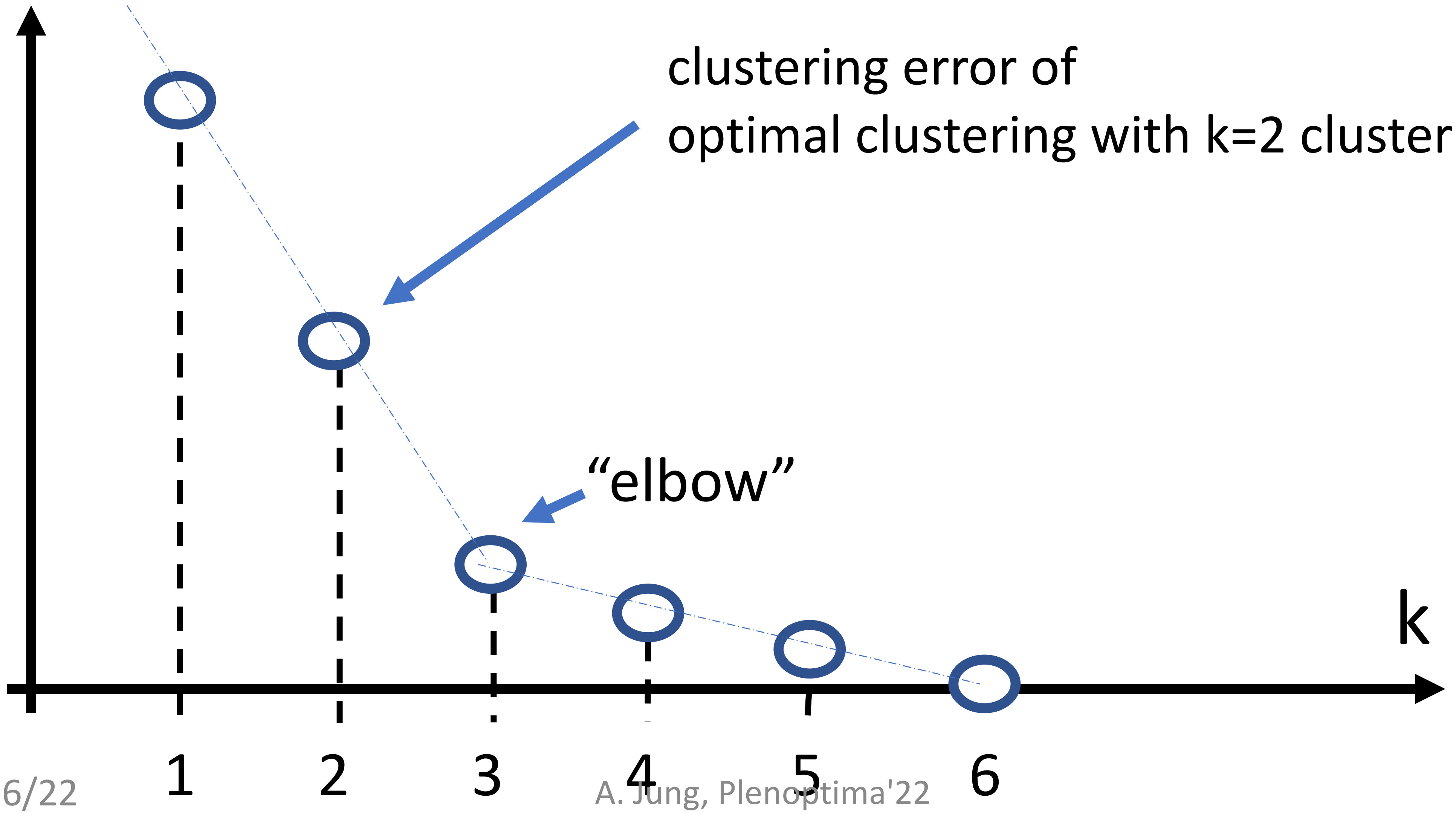


- defined by application (img. seg.)
- desired compression rate
- "elbow-method"

For/Background Segmentation k=2

Cluster 1 = Background, Cluster 2=Foreground

Elbow Method



Choose k by Validation Error

- clustering an be used as pre-processing for follow-up regression method

- try different values of k and pick the one resulting in smallest validation error

To Sum Up

- represent clusters by Gaussian distributions
- soft clustering algorithm fits GMM
- iterative optimization of soft-clustering error
- trapped in local minimum for bad initialization

Thank You!