# Medical Image Analysis and Processing

## Medical Image Segmentation

## Pixel Classification

Emad Fatemizadeh

Distance/online Course: Session 18

Date: 27 April 2021, 7th Ordibehesht 1400

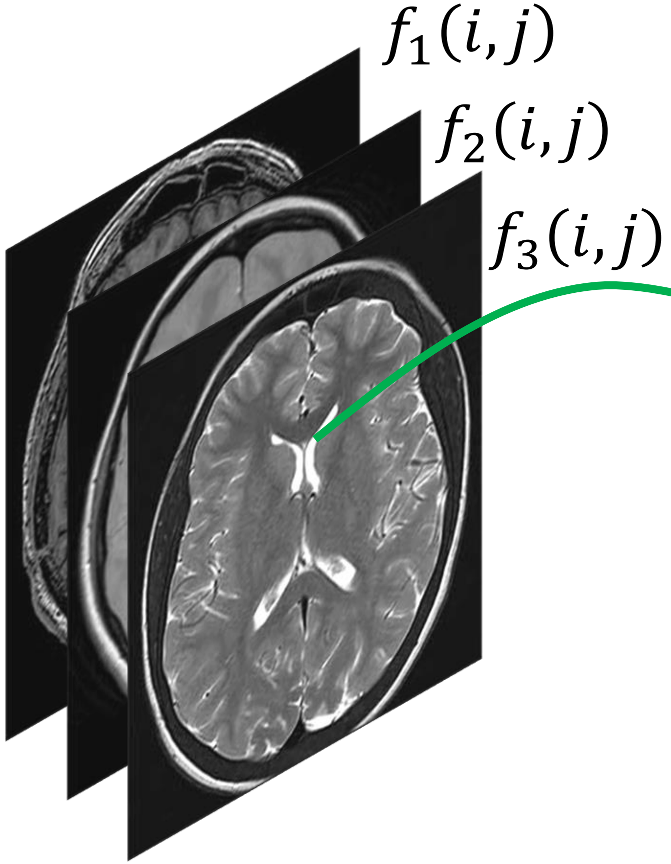
# Contents

# Pixel Classification -Definition

- In pixel classification techniques, segmentation depends upon feature space using pixel attributes:
  - Positions,
  - Gray level (s),
  - Filter responses,
  - Statistical features ,
  - Local features (SIFT, ORB, LBP, …)
- Each pixels represents by a vector of features: $x \in \mathbb{R}^D$

# Pixel Classification -Definition

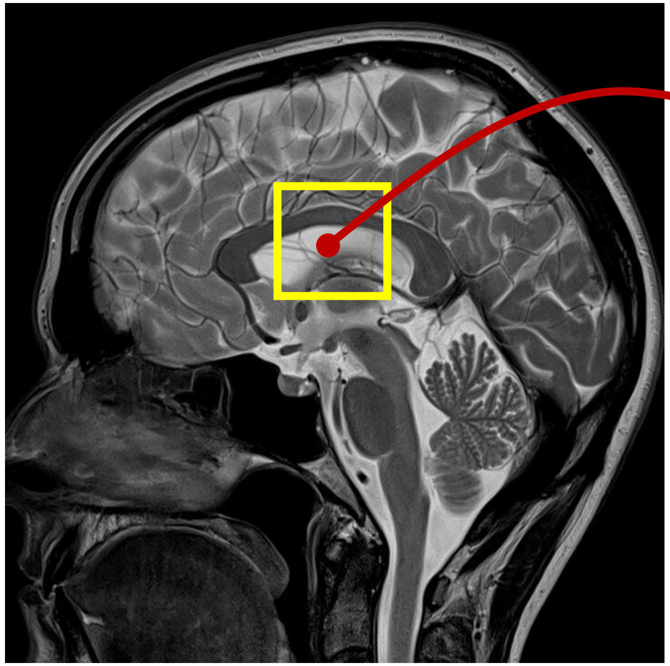
› Multimodal images raw features:




$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} f_1(i,j) \\ f_2(i,j) \\ f_3(i,j) \end{bmatrix}$$

# Pixel Classification -Definition

› Single (Multi) modal computation features:



$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$$

# Pixel Classification - Strategies

- There are two main catergories:
  - Supervised
    - Labeled images (atlas) is available (pixel-by-pixel)
  - Unsupervised
    - A single slice or volume is available

# Necessary Tools

- Mathematical tools:
  - Probability density function estimation
  - Bayesian decision theorem
  - Machine Learning (Supervised and Unsupervised)

# Probability density function estimation

- There is two approaches:
  - Parametric,
  - Non-Parametric,
- Parametric pdf estimation:
  - Assume a parametric statistic models for data, and estimate the parameters of model, most known method is Maximum Likelihood Estimation (MLE)
- Non-Parametric:
  - There is no assumption about data, most known is **K**ernel **D**ensity **E**stimation (KDE) or Parzen-Rosenblat window methods.

# Maximum Likelihood Estimation (MLE)

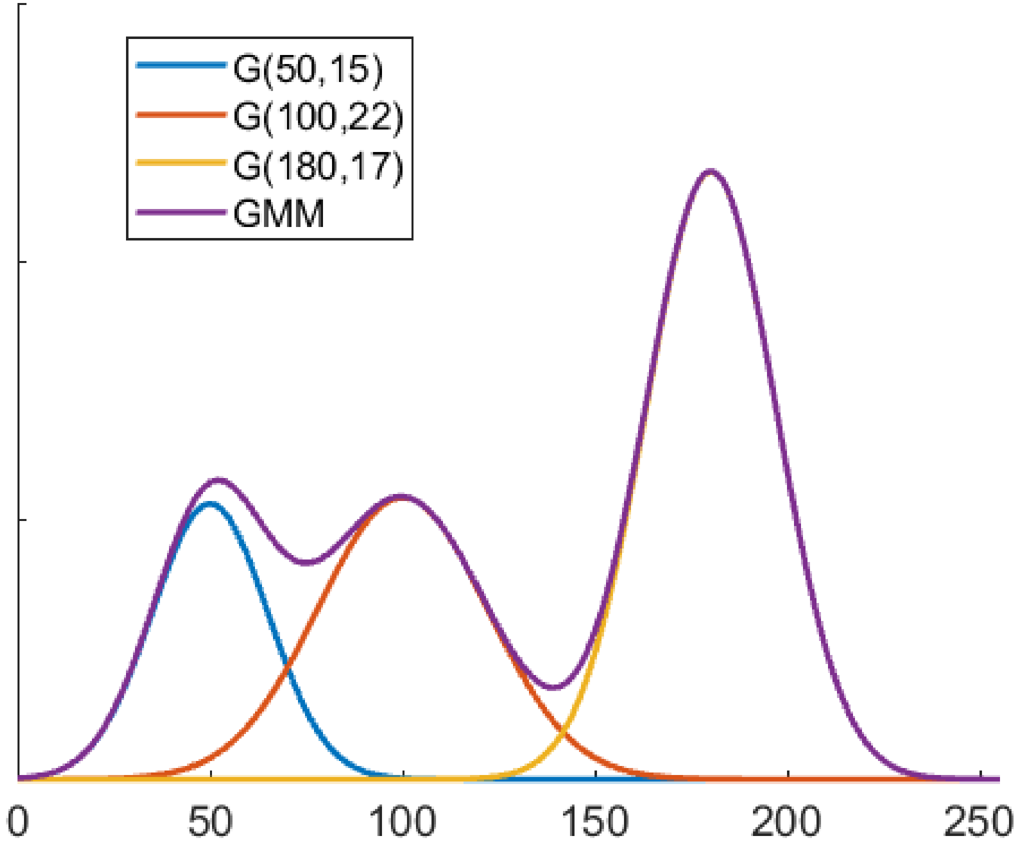
- $input\ data$: $\{\boldsymbol{x}_i\}_{i=1}^{N} \in \mathbb{R}^D$ ($i.i.d$ observation)
- $assumption$: $\boldsymbol{x}_i \sim p(\boldsymbol{x};\boldsymbol{\theta})$
- $\widehat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\boldsymbol{x}_i;\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n} log\big(p(\boldsymbol{x}_i;\boldsymbol{\theta})\big)$
- Most popular pdf model in segmentation is *Gaussian Mixture Models* (GMM):

$$p(\boldsymbol{x};\boldsymbol{\theta}) = \sum_{k=1}^{m} \pi_k G(\boldsymbol{x};\boldsymbol{\mu}_k,\boldsymbol{\Sigma_k}), \qquad \sum_{k=1}^{m} \pi_k = 1$$

- Where $m$ is number of mixture, $\boldsymbol{\mu}_k$ is center, $\boldsymbol{\Sigma_k}$ is covariance matrix, and $\pi_k$ is mixing probability of mixture $k$.

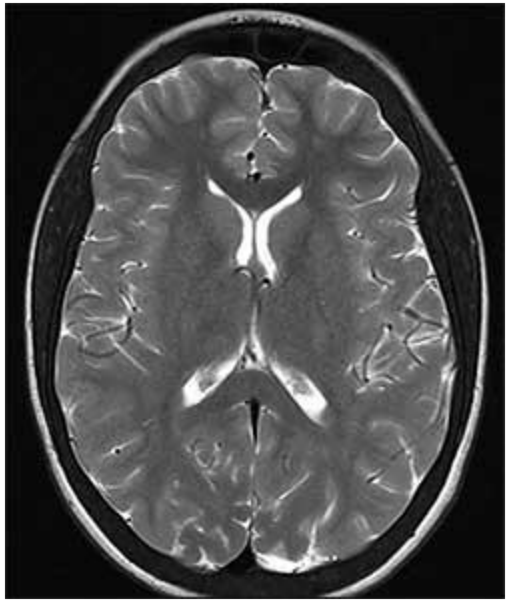
# GMM illustration (synthetic data)

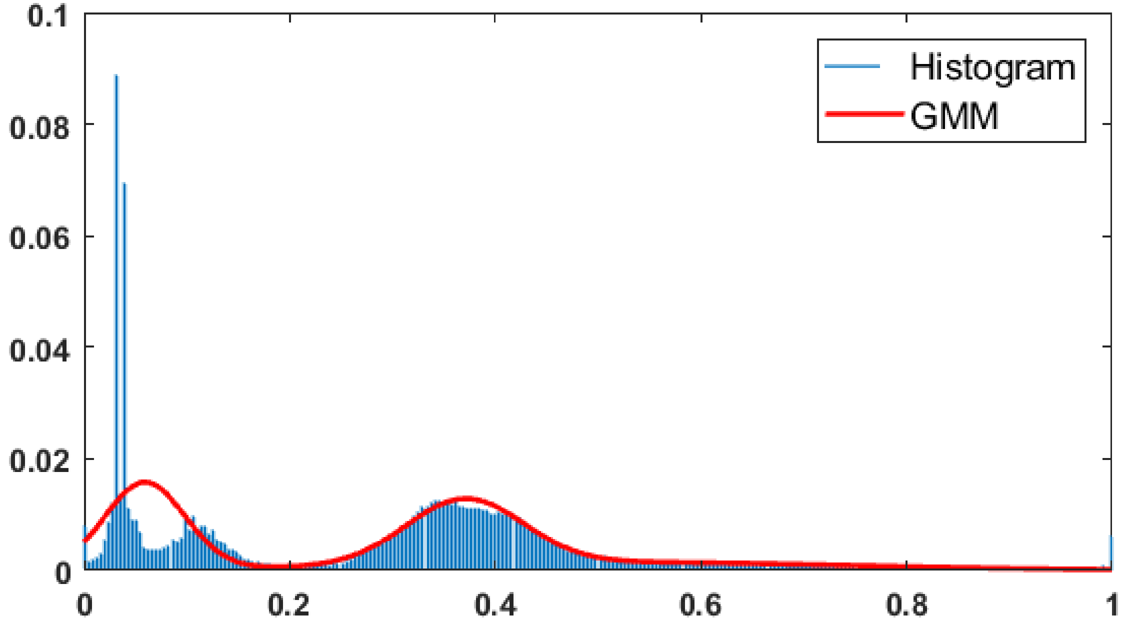
› GMM with three mixture (m=3)



# GMM illustration (real data)

› GMM example:





# Kernel Density Estimation (KDE) or Parzen Windows



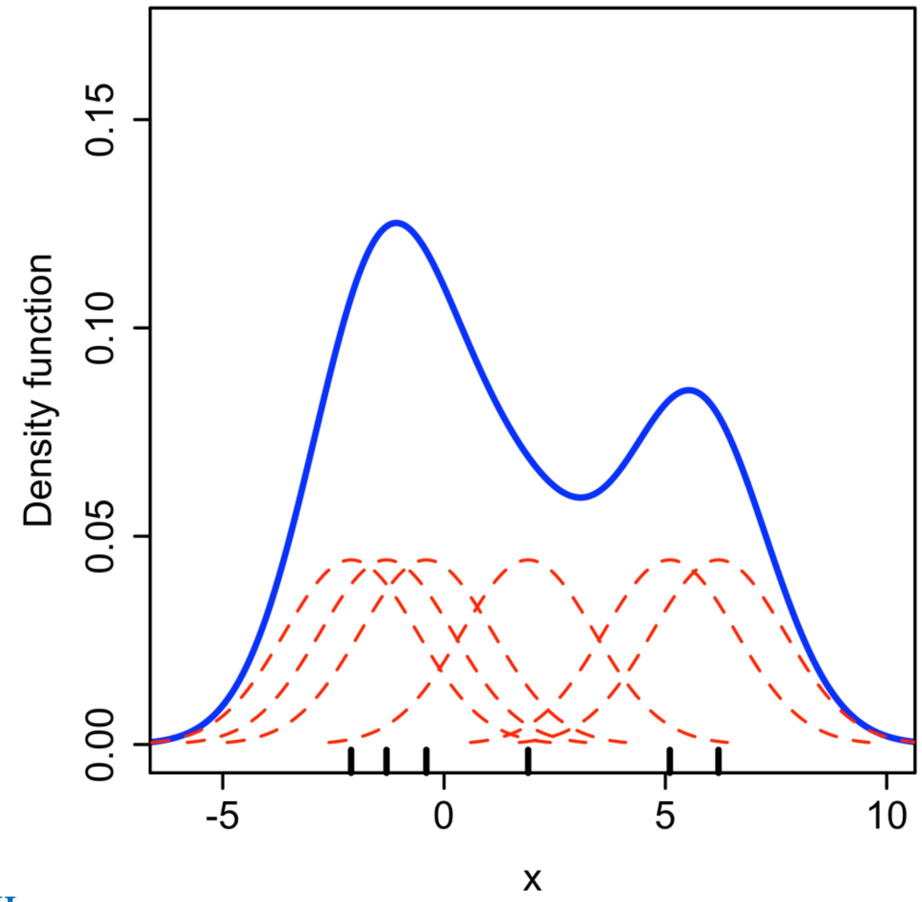
- Input data: $\{\boldsymbol{x}_i\}_{i=1}^{N} \in \mathbb{R}^D$
- A non-negative, non-increasing and piece-wise continuous kernel, $k(x)$, and $K(\boldsymbol{x}) = c_D k(\|\boldsymbol{x}\|^2)$, where $c_D$ is normalization constant so that $K(\boldsymbol{x})$ integrates to 1
- A smoothing bandwidth: $h$

$$\hat{p}(x) = \frac{1}{Nh^D} \sum_{i=1}^{N} K\left(\frac{\boldsymbol{x}_i - \boldsymbol{x}}{h}\right)$$

- Most used kernel function, $k(x)$, is gaussian, $N(0,1)$

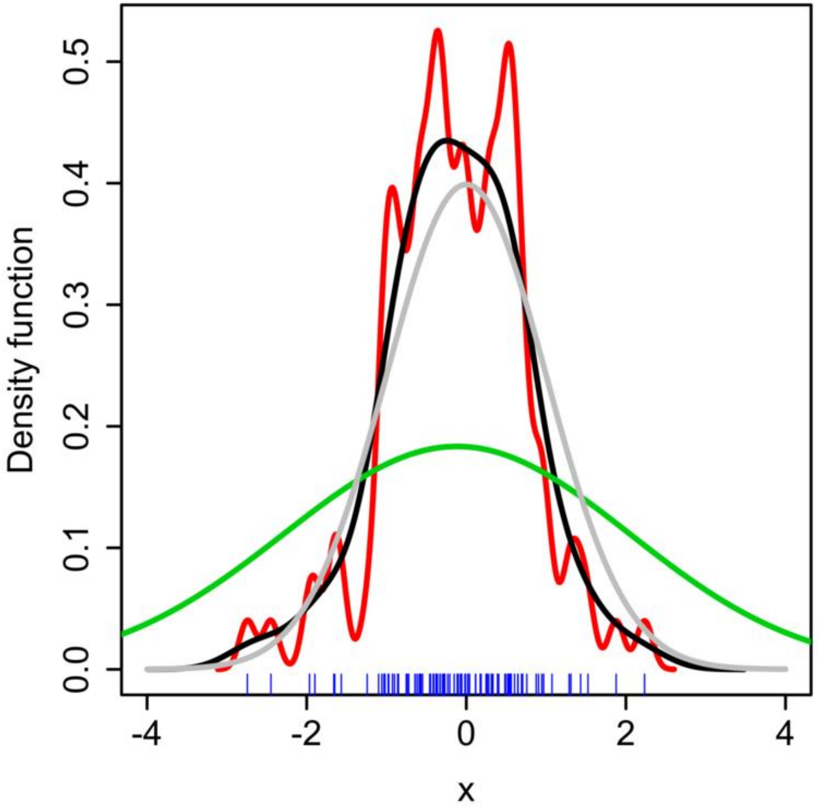
# KDE illustration (Synthetic data)

› KDE with 6 observation



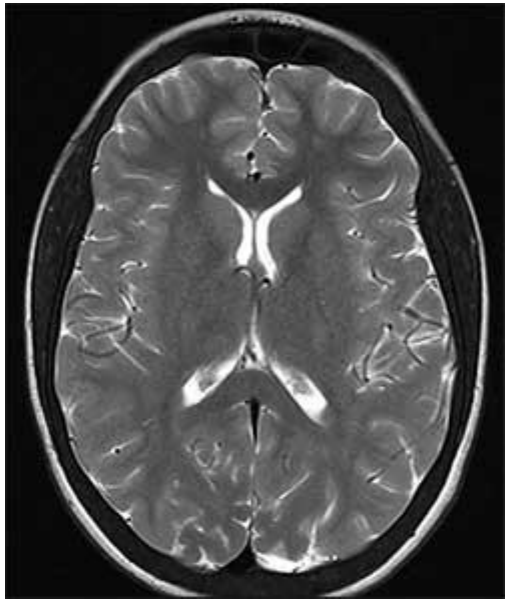
# KDE illustration – bandwidth effect

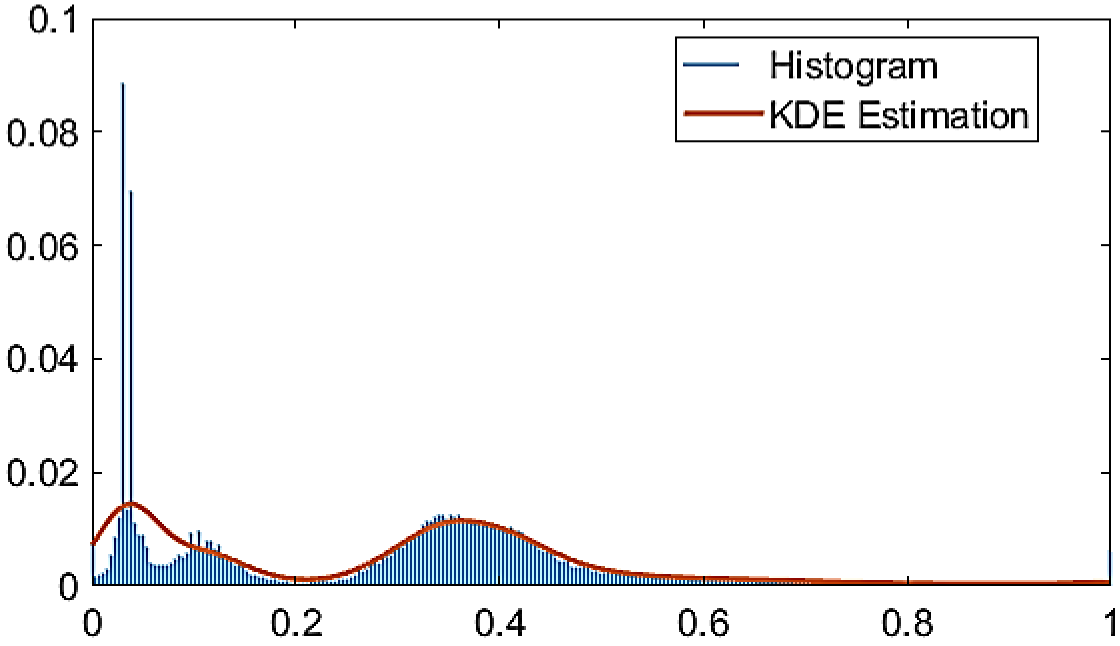
› KDE estimation for 4 different values for *h*



# KDE illustration (real data)

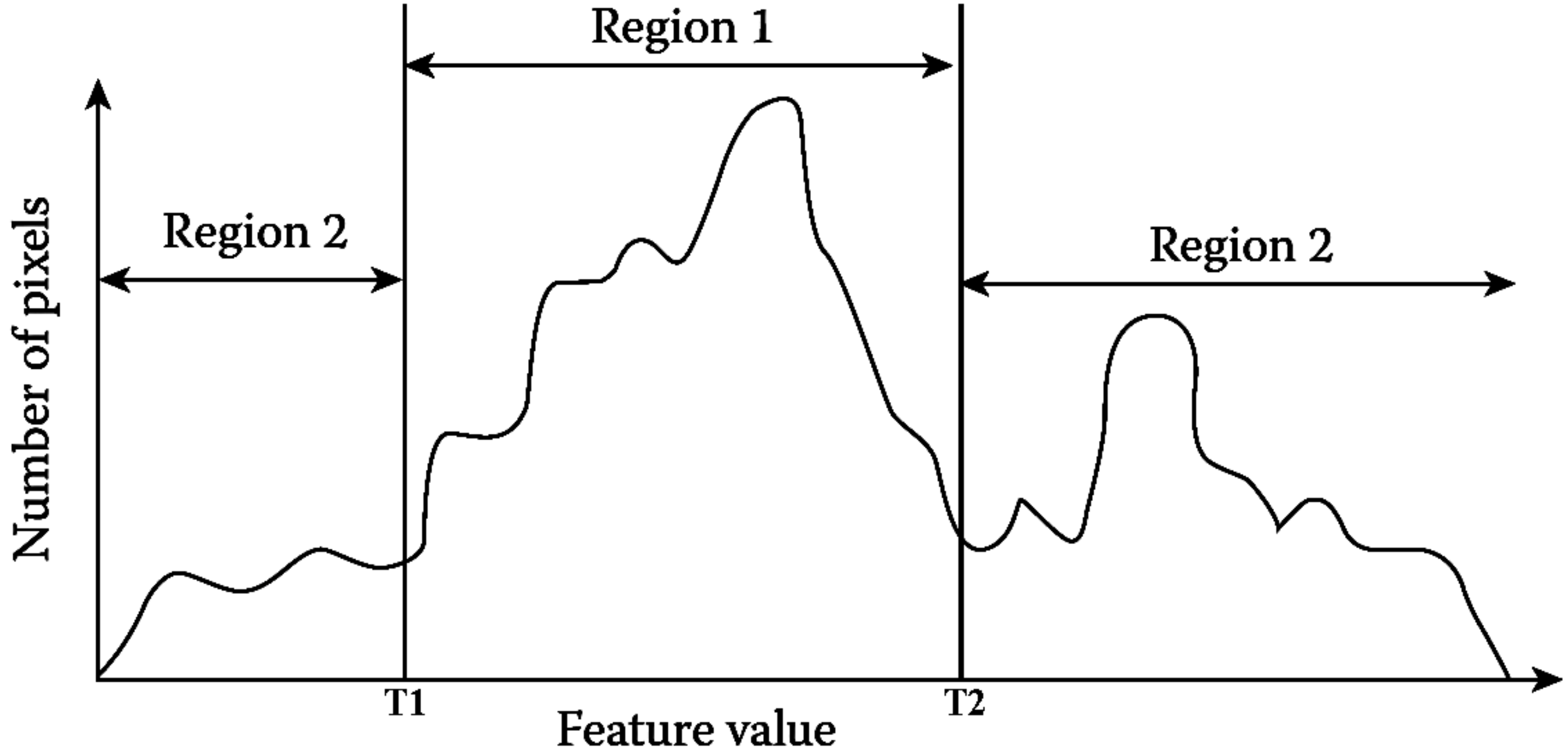
› KDE example:





# Thresholding

- Thresholding: Segment *scalar* images by creating a binary partitioning of the image intensities.



# Thresholding

- How to determine Threshold(s):
  - Supervised
  - Unsupervised

# Optimal Supervised threshold estimation

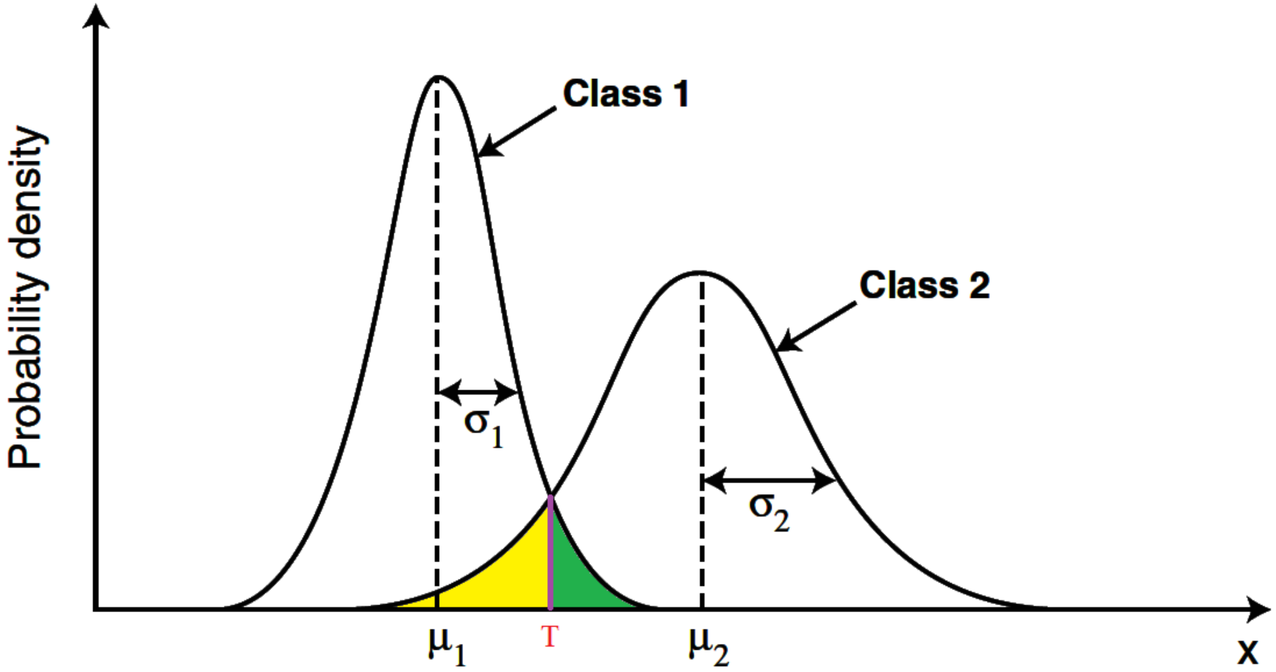
- Optimal Supervised threshold estimation:
- Assume we have two segments with known *monomodal* distribution: $\{p(x;\boldsymbol{\theta}_i)\}_{i=1}^{2}$ and prior probabilities $\{\pi_i\}_{i=1}^{2}$
- It can be shown (*Bayes* decision theorem) the optimal threshold is solution of the following equation:

$$\pi_1 p(T;\boldsymbol{\theta}_1) = \pi_2 p(T;\boldsymbol{\theta}_2)$$

$$S_2(x_i) = \begin{cases} 1, & x(i,j) \geq T \\ 0, & x(i,j) < T \end{cases} \text{ and } S_1(x_i) = \begin{cases} 1, & x(i,j) < T \\ 0, & x(i,j) \geq T \end{cases}$$

# Optimal Supervised threshold estimation

› *Bayes* decision minimize: "FN+FP" area



# Optimal Unsupervised threshold estimation:

- Distribution of two segment estimated via *GMM* or *xMM*!

$$p(x;\theta) = \pi_1 G(x;\mu_1,\sigma_1^2) + \pi_2 G(x;\mu_2,\sigma_2^2)$$

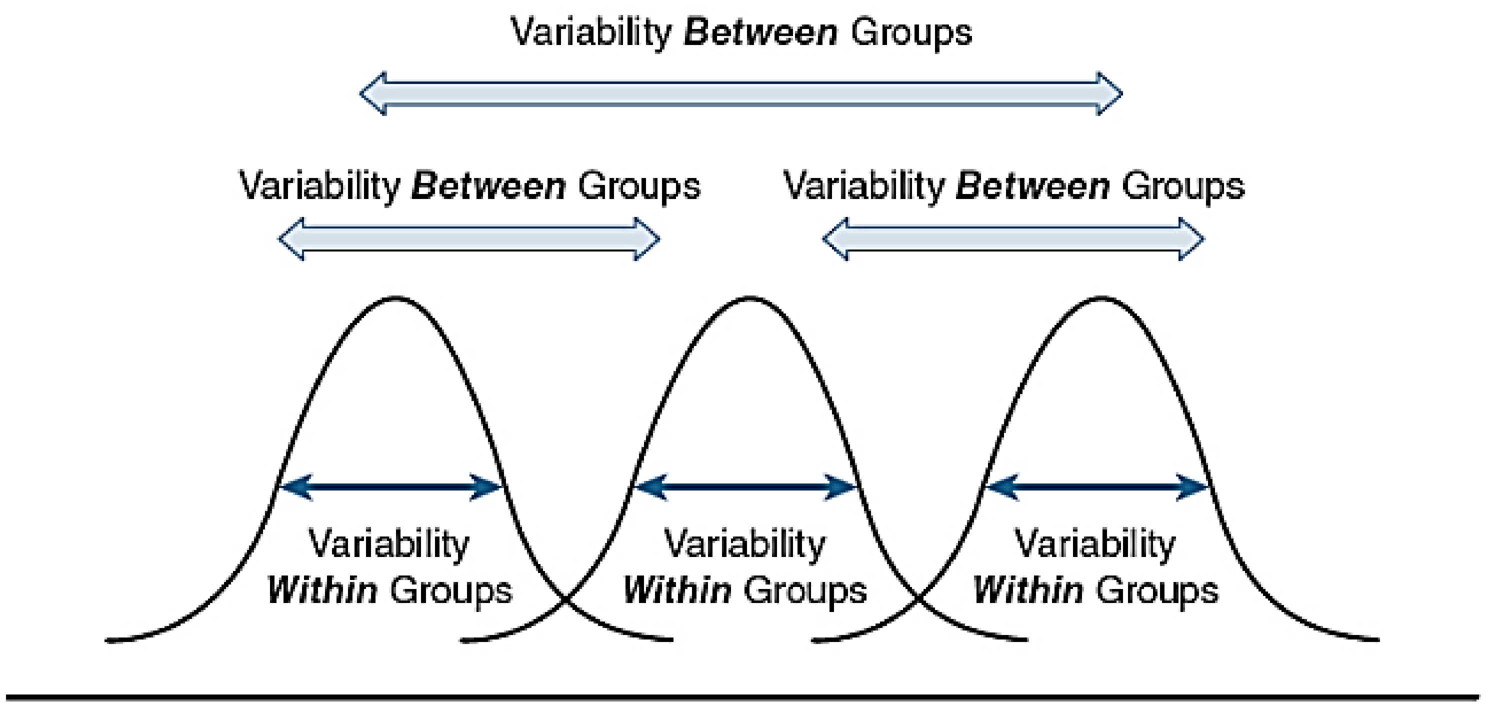
- The optimal threshold is solution of this equation:

$$\pi_1 G(T;\mu_1,\sigma_1^2) = \pi_2 G(T;\mu_2,\sigma_2^2)$$

$$S_2(x_i) = \begin{cases} 1, & x(i,j) \geq T \\ 0, & x(i,j) < T \end{cases} \text{ and } S_1(x_i) = \begin{cases} 1, & x(i,j) < T \\ 0, & x(i,j) \geq T \end{cases}$$

# Otsu Thresholding

› *Otsu* main idea: Minimize within-segments while maximize between-segment distance



# Otsu Thresholding

- Otsu methods - Optimal unsupervised threshold estimation:
- Find (exhaustive search) threshold(s) that maximize any of the three discriminant criteria:

$$\frac{\sigma_B^2}{\sigma_W^2}, \frac{\sigma_T^2}{\sigma_W^2}, \frac{\sigma_B^2}{\sigma_T^2}$$

- $\sigma_B^2$: Between-Segment variances
- $\sigma_W^2$: Within-Segment variances
- $\sigma_T^2$: Total variances ($\sigma_T^2 = \sigma_W^2 + \sigma_B^2$)

# Otsu Thresholding

- Let $\{p_i\}_{i=0}^{L-1}$ denote normalized histogram of input image, $g$, and, suppose that we select a threshold $T(k) = k,\ 0 < k < L - 1$, we have two segments:

$$S_1: [0, k], S_2: [k + 1, L - 1]$$

- with prior probabilities:

$$P_1(k) = \sum_{i=0}^{k} p_i, P_2(k) = \sum_{i=k+1}^{L-1} p_i$$

- and means and variances:

$$\{m_i(k)\}_{k=1}^{2}, \{\sigma_i^2(k)\}_{k=1}^{2}$$

- Global means is: $m_G(k) = P_1(k)m_1(k) + P_2(k)m_2(k)$

# Otsu Thresholding

- Between-segment variance:

$$\sigma_B^2(k) = P_1(k)\big(m_1(k) - m_G(k)\big)^2 + P_2(k)\big(m_2(k) - m_G(k)\big)^2$$

- Within-segment variance

$$\sigma_W^2(k) = P_1(k)\sigma_1^2(k) + P_2(k)\sigma_2^2(k)$$

- Total variance:

$$\sigma_T^2(k)=\sigma_W^2(k)+\sigma_B^2(k)$$

- Then, the optimum threshold (Otsu) is the value, $k^*$, that maximizes one of following:

$$\left\{\frac{\sigma_B^2(k)}{\sigma_W^2(k)}\right\}_{k=0}^{L=1}, \left\{\frac{\sigma_T^2}{\sigma_W^2(k)}\right\}_{k=0}^{L=1}, \left\{\frac{\sigma_B^2(k)}{\sigma_T^2(k)}\right\}_{k=0}^{L=1}$$

# Kittler-Illingworth Thresholding

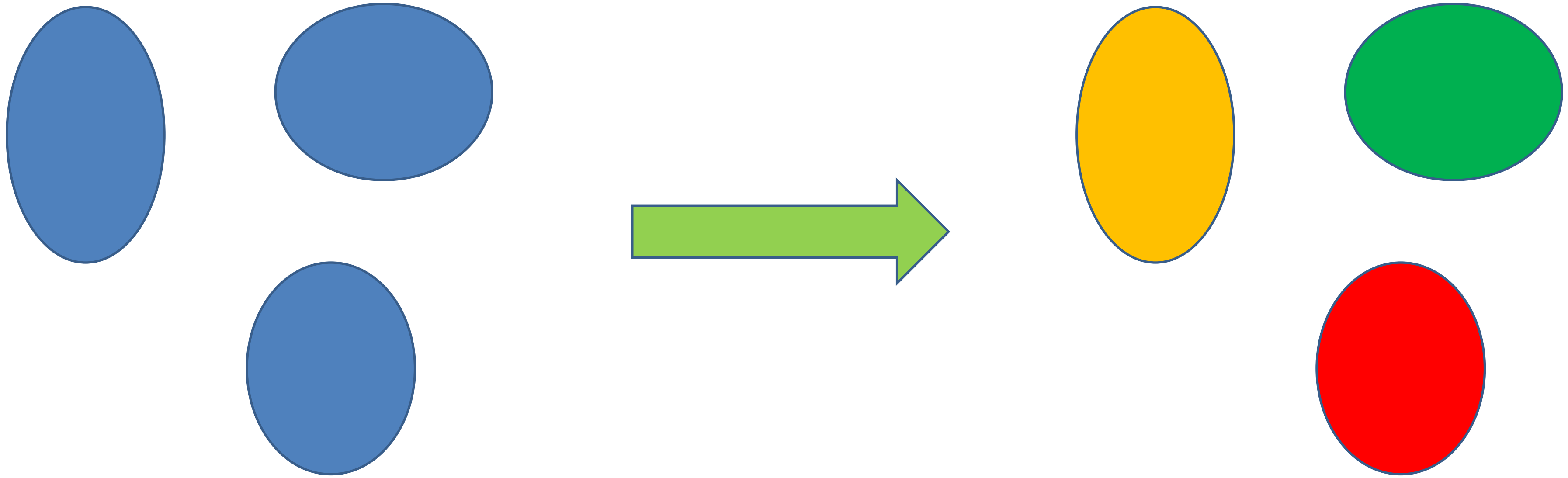
› The *Kittler-Illingworth* optimum threshold is the value, $k^*$, that minimize :

$$1 + 2\left(P_1(k) log\sigma_1(k) + P_2(k) log\sigma_2(k)\right) - 2\left(P_1(k) log P_1(k) + P_2(k) log P_2(k)\right)$$

› Extension to multiple thresholds is possible!

# Pixel Classification – Clustering

- A well known methods for image segmentation.
- What is data clustering (unsupervised learning):

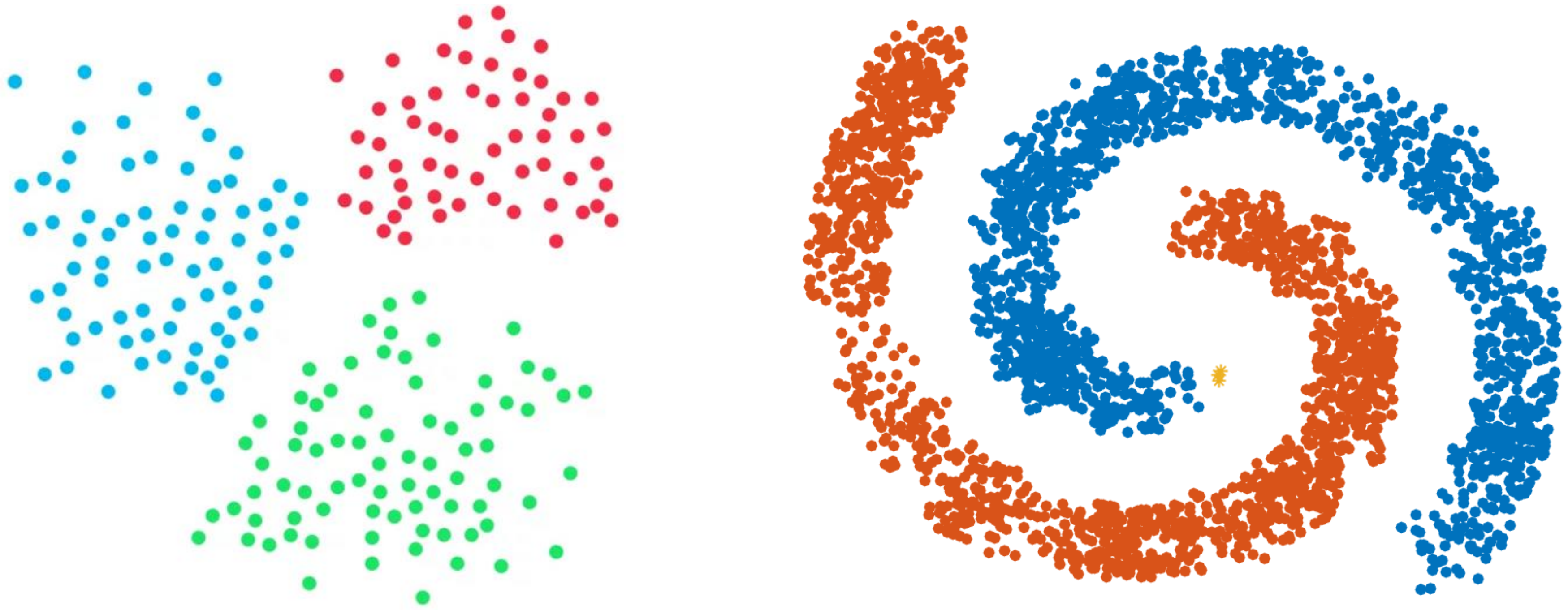


# Pixel Classification – Clustering

- Any clustering method is useful:
  - K-means,
  - Mean-Shift,
  - Fuzzy K-means,
  - Hierarchical Clustering,
  - Self Organization Map (SOM),
  - …

# Clustering Illustration

› Easy task (left) and hard task (Right)

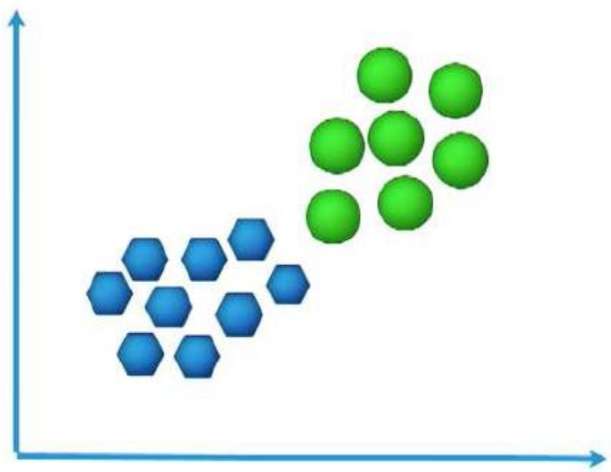


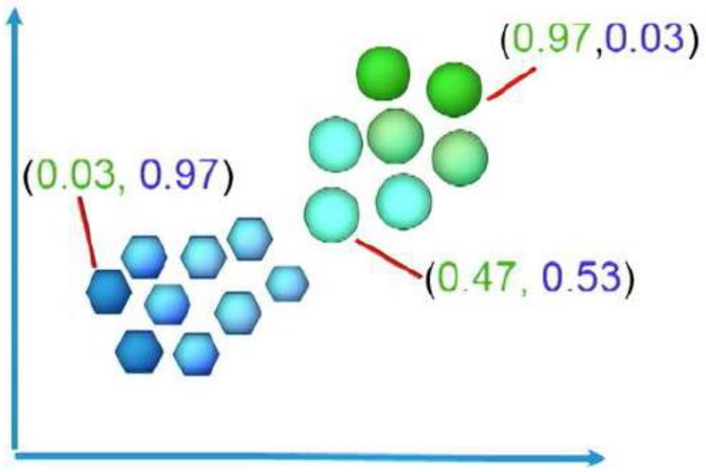
# Clustering Definition

- Clustering is data point partitioning:
- $input: \ X = \{x_i\}_{i=1}^{N}, K: \# \ of \ clusters$
- $output: \ \{C_j\}_{j=1}^{K}, X = \bigcup_{j=1}^{K} C_j$
- Goal: *within-cluster similarity* and *between-cluster dissimilarity*
- Too many options for *similarity* and *dissimilarity* are introduced

# Hard vs Soft Clustering

› Clustering may be hard (left) or soft (right):





# Hard Clustering

- Definition by membership function:
- $u_{ij} \in \{0,1\}$
- $u_{ij} = \begin{cases} 1, & x_i \in C_j \\ 0, & x_i \notin C_j \end{cases}$,
- $\sum_{j=1}^{K} u_{ij} = 1\,, \quad \forall i = 1, 2, \dots, N$
- $0 < \sum_{i=1}^{N} u_{ij} < N, \quad \forall j = 1, 2, \dots, K$
- Good for distinct (non-overlapping) data distribution

# Soft Clustering

- Definition by membership function:
- $u_{ij} \in [0,1]$
- $0 \leq u_{ij} \leq 1$
- $\sum_{j=1}^{K} u_{ij} = 1 \,, \quad \forall i = 1, 2, \dots, N$
- $0 < \sum_{i=1}^{N} u_{ij} < N, \quad \forall j = 1, 2, \dots, K$
- Good for overlapping data distribution

# The End

› AnY QuEsTiOn?