# Segmentation:

## Regions & K-means

Dr. Tushar Sandhan

# Introduction

- Goal of segmentation


Image credit: J. Jordan

# Introduction

- Goal of segmentation


Image credit: J. Jordan

# Region growing

- Partition the image $I$ into $m$ regions
  - every pix belong to some region
  - each pix is assigned to only one region
  - all pix in a region, share similar property
  - all pix in diff. regions have distinct properties
  - Prerequisites:
    1. seed points
    2. similarity measures
    3. stopping criterion

# Region growing

- Partition the image $I$ into $m$ regions
  - every pix belong to some region
  - each pix is assigned to only one region
  - all pix in a region, share similar property
  - all pix in diff. regions have distinct properties
  - Prerequisites:
    1. seed points
    2. similarity measures
    3. stopping criterion

$$I = \bigcup_{i=1}^{m} R_i$$

# Region growing

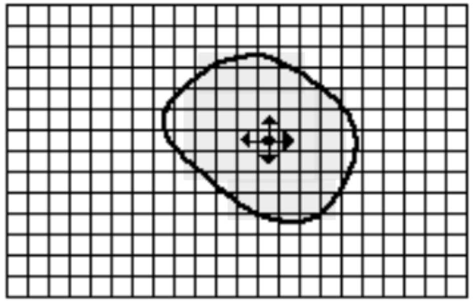
- Partition the image $I$ into $m$ regions
  - every pix belong to some region
  - each pix is assigned to only one region
  - all pix in a region, share similar property
  - all pix in diff. regions have distinct properties
  - Prerequisites:
    1. seed points
    2. similarity measures
    3. stopping criterion

$$I = \bigcup_{i=1}^{m} R_i \qquad R_i \bigcap R_j = \phi, \quad \forall \; i \neq j$$

# Region growing

- Partition the image $I$ into $m$ regions
  - every pix belong to some region
  - each pix is assigned to only one region
  - all pix in a region, share similar property
  - all pix in diff. regions have distinct properties
  - Prerequisites:
    1. seed points
    2. similarity measures
    3. stopping criterion

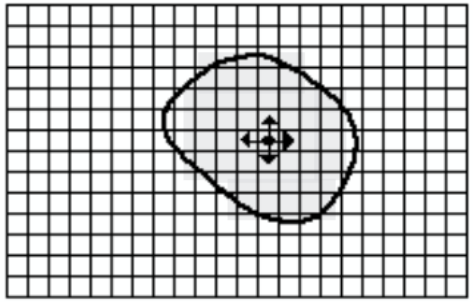
$$I = \bigcup_{i=1}^{m} R_i \qquad R_i \bigcap R_j = \phi, \quad \forall\ i \neq j$$

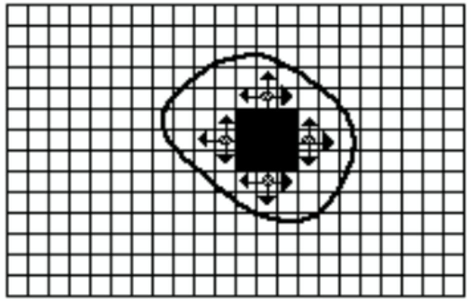


# Region growing

- Partition the image $I$ into $m$ regions
  - every pix belong to some region
  - each pix is assigned to only one region
  - all pix in a region, share similar property
  - all pix in diff. regions have distinct properties
  - Prerequisites:
    1. seed points
    2. similarity measures
    3. stopping criterion

$$I = \bigcup_{i=1}^{m} R_i \qquad R_i \bigcap R_j = \phi, \quad \forall \ i \neq j$$





after few iterations

# Region growing

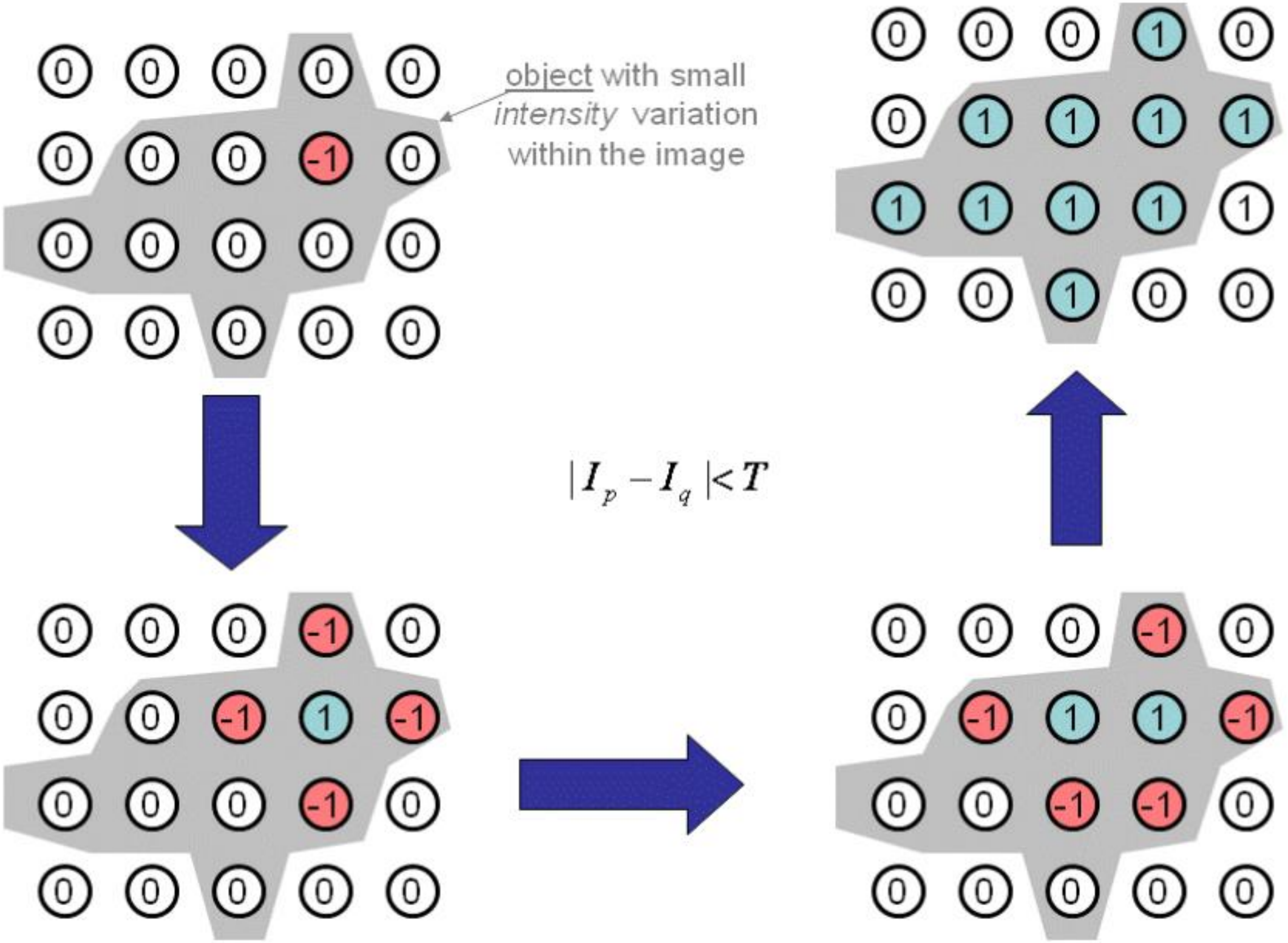
- Partition and grow

---

- Start with a seed point $s_j$ for region $R_j$
- For every $s_j$
  - Initialize mean intensity of each region: $\mu_j = s_j$
  - Initialize region: $R_j = \{s_j\}$
- For each point $p$ in $R_j$
  - Get its 4-connect neighborhood: $\mathcal{N}_i(p),\ i = 1, 2, 3, 4$
  - If $|\mathcal{N}_i(p) - \mu_j| < \tau$ , $\mathcal{N}_i(p) \notin \mathcal{R}_k \quad j \neq k$
    - $\mathcal{R}_j \leftarrow \mathcal{R}_j \bigcup \mathcal{N}_i(p)$
    - update $\mu_j$
  - Stop growing when no neighborhood pixel matches
- Move to the next seed point, until the whole image is partitioned.

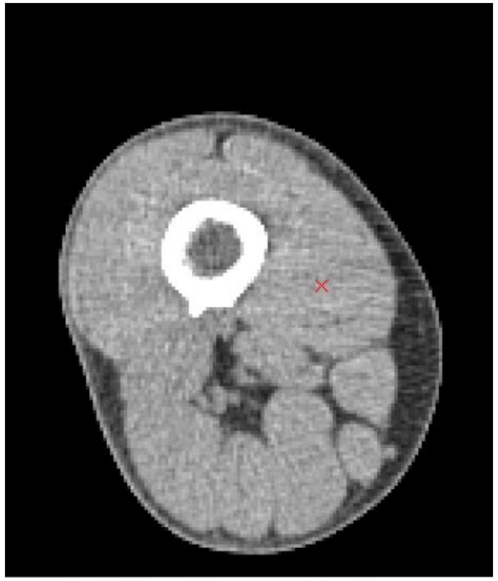
---

# Region growing



# Region growing: CT scan

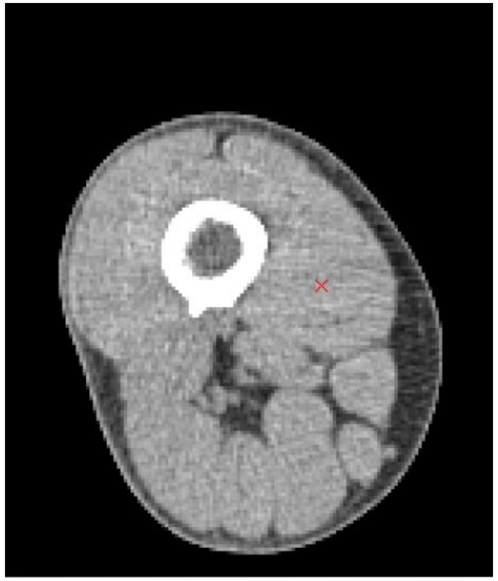
- Seed-1



stricter similarity

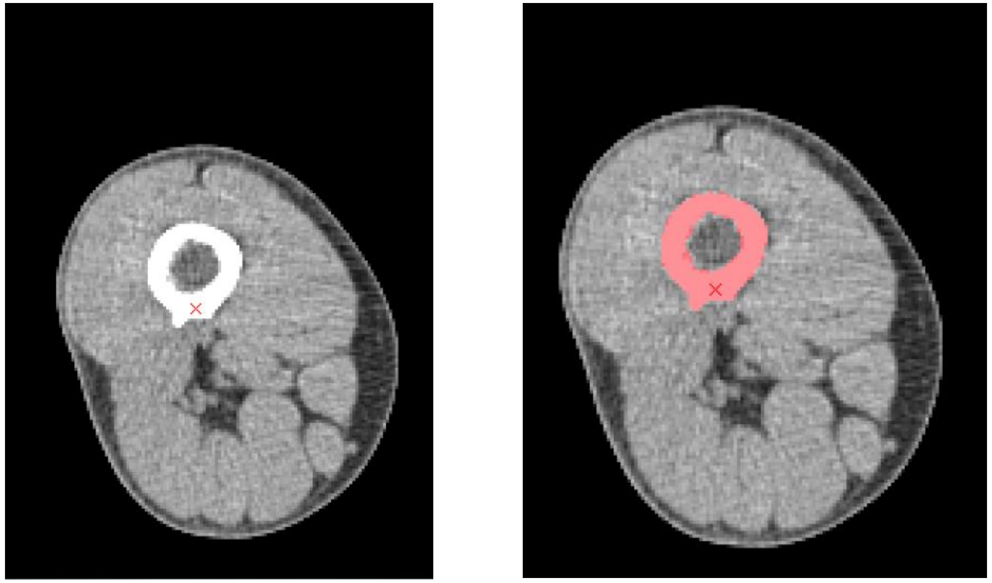
# Region growing: CT scan

- Seed-1



stricter similarity

- Seed-2



# Comparative example

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

threshold $T \geq 10$

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

threshold $T \geq 11$

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

threshold $T \geq 12$

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3 | 5 | 7 | 3 | 4 | 2 | 1 |
| 2 | 4 | 9 | 10 | 22 | 9 | 3 |
| 3 | 5 | 12 | 11 | 15 | 10 | 3 |
| 5 | 6 | 11 | 9 | 17 | 19 | 1 |
| 2 | 3 | 11 | 12 | 18 | 16 | 2 |
| 3 | 6 | 8 | 10 | 18 | 9 | 5 |
| 4 | 6 | 7 | 8 | 3 | 3 | 1 |

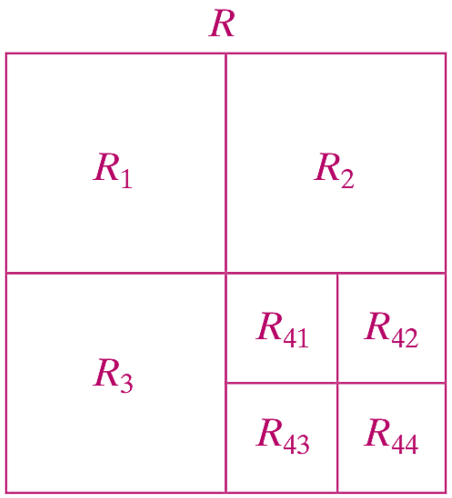
region growing with variance of 2 in respect to value 11 with reference to threshold $T \geq 11$

# Region splitting & merging

- Split
  - sub-quadrants
    - e.g. 4 parts: quadregions
    - quadtree (having leaves as quadregions or quadimages)
  - continuous splitting
    - adjacent quadimages will be having identical properties

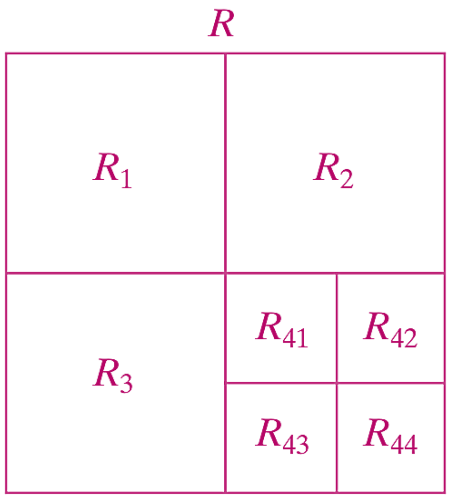
# Region splitting & merging

- Split
  - sub-quadrants
    - e.g. 4 parts: quadregions
    - quadtree (having leaves as quadregions or quadimages)
  - continuous splitting
    - adjacent quadimages will be having identical properties



# Region splitting & merging

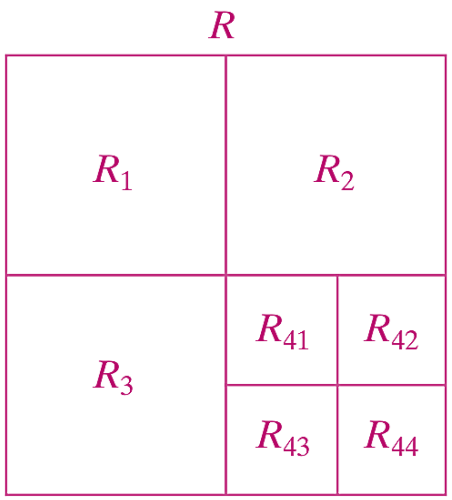
- Split
  - sub-quadrants
    - e.g. 4 parts: quadregions
    - quadtree (having leaves as quadregions or quadimages)
  - continuous splitting
    - adjacent quadimages will be having identical properties



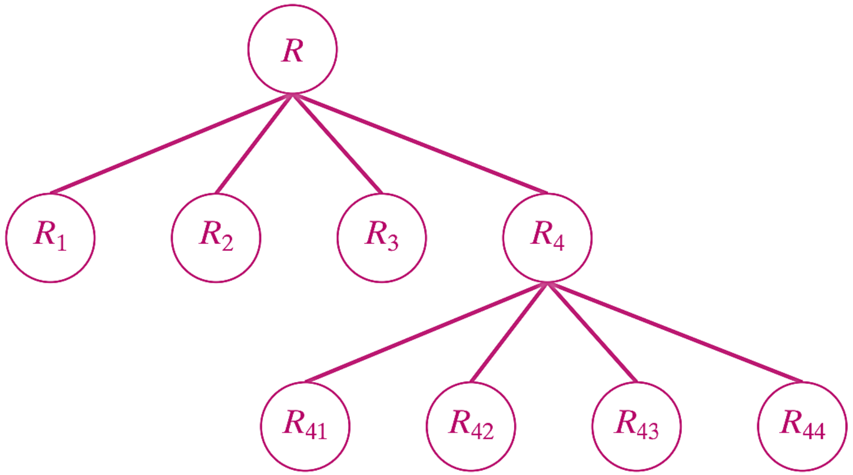

- Merging
  - quadimages that satisfy closeness in similarity criterion
    - quadimages to be merged should be adjacent
    - merging begins when no further splitting is possible

# Region splitting & merging

- Split
  - sub-quadrants
    - e.g. 4 parts: quadregions
    - quadtree (having leaves as quadregions or quadimages)
  - continuous splitting
    - adjacent quadimages will be having identical properties



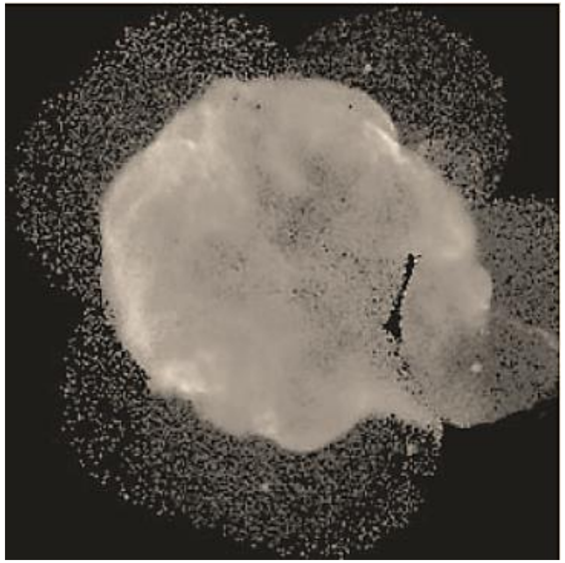

- Merging
  - quadimages that satisfy closeness in similarity criterion
    - quadimages to be merged should be adjacent
    - merging begins when no further splitting is possible



# Region splitting & merging

- Segment the ring of supernova
  - quadimages size 32x32, 16x16 & 8x8
  - variance and mean of quadimages can be used as merging criterion

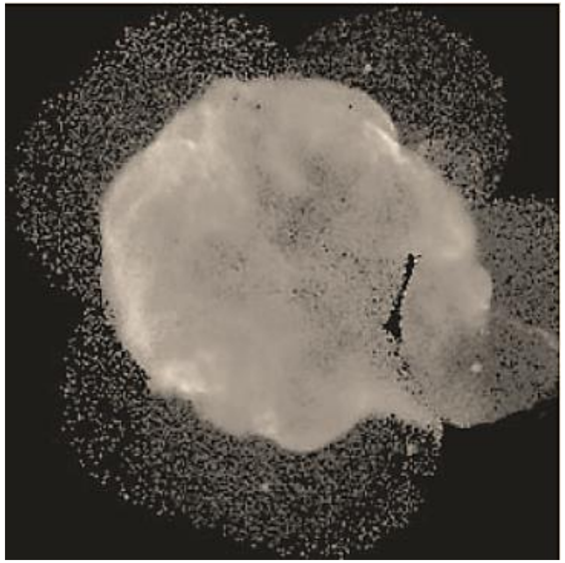
X-ray band image



# Region splitting & merging

- Segment the ring of supernova
  - quadimages size 32x32, 16x16 & 8x8
  - variance and mean of quadimages can be used as merging criterion

X-ray band image

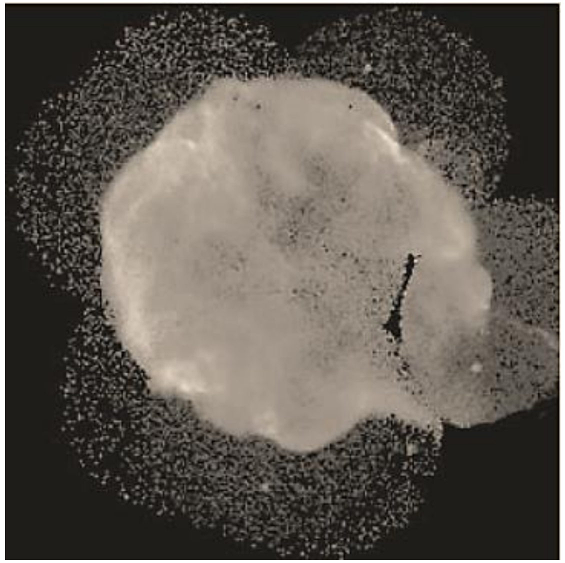


32x32

# Region splitting & merging

- Segment the ring of supernova
  - quadimages size 32x32, 16x16 & 8x8
  - variance and mean of quadimages can be used as merging criterion

X-ray band image



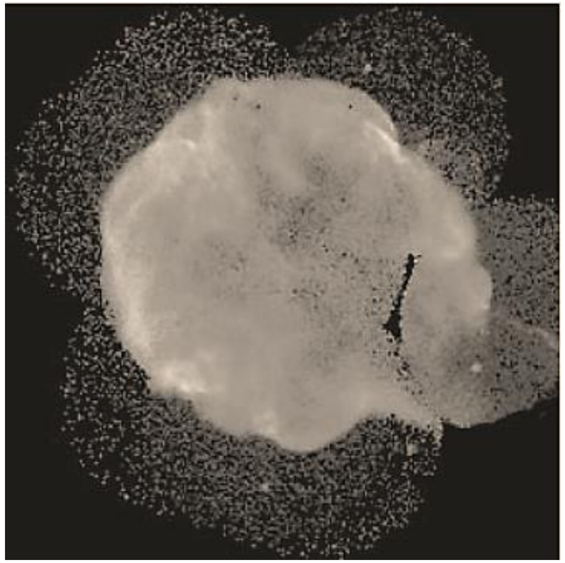
32x32

16x16

# Region splitting & merging

- Segment the ring of supernova
  - quadimages size 32x32, 16x16 & 8x8
  - variance and mean of quadimages can be used as merging criterion

X-ray band image



32x32

16x16

8x8

# Clustering

# Clustering

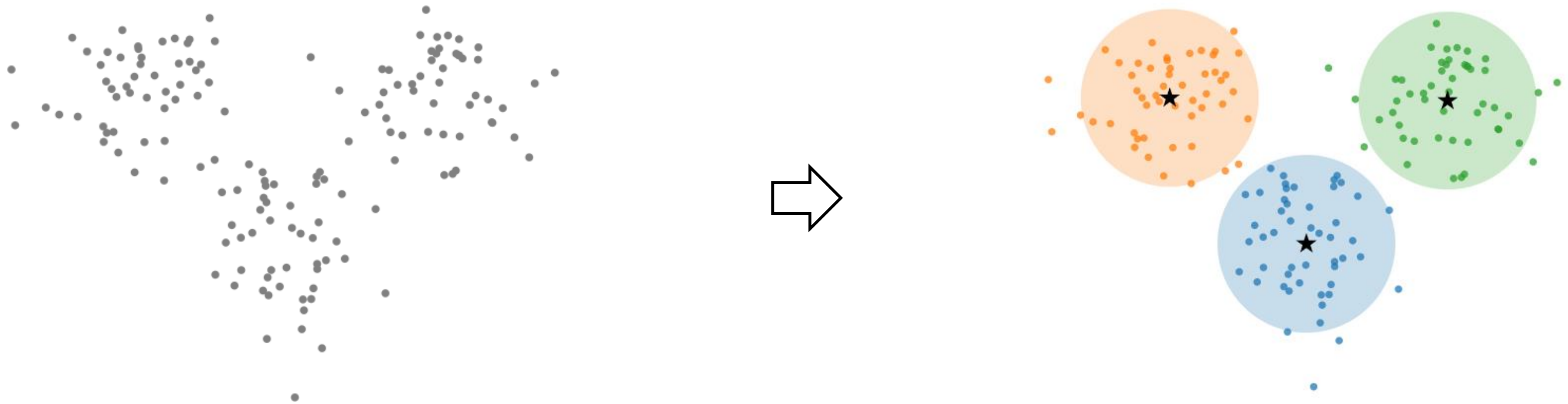
- Organizing data into multiple (#clusters) classes s.t.:
  - intra-class variance is low (high similarity)
  - inter-class variance is high (low similarity)
- Unsupervised learning paradigm
  - finding class labels directly from data
  - training data labels are not available
- What are similarity measures:
  - distance
    - e.g. euclidian, cosine
  - density
    - e.g. amount of neighbourhood

# Clustering

- Organizing data into multiple (#clusters) classes s.t.:
  - intra-class variance is low (high similarity)
  - inter-class variance is high (low similarity)
- Unsupervised learning paradigm
  - finding class labels directly from data
  - training data labels are not available
- What are similarity measures:
  - distance
    - e.g. euclidian, cosine
  - density
    - e.g. amount of neighbourhood

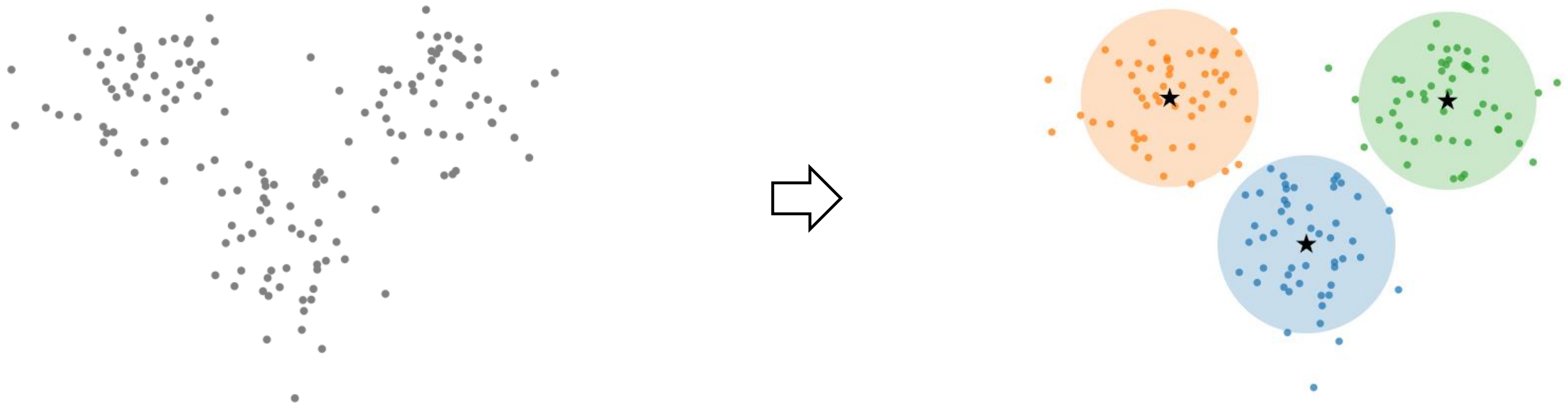
# K-Means

- K-means clustering
  - unsupervised learning method: requires data but not labels
  - useful for pattern recognition, when we don’t know what to look for
  - detects united patterns e.g. groups of text topics, regions of images
  - pros: simple iterative
  - cons: difficult to guess K



# K-Means

- K-means clustering
  - unsupervised learning method: requires data but not labels
  - useful for pattern recognition, when we don't know what to look for
  - detects united patterns e.g. groups of text topics, regions of images
  - pros: simple iterative
  - cons: difficult to guess K



# K-Means

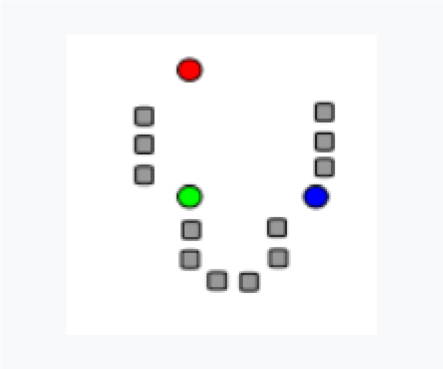
- **Input:** $x^{(1)}, x^{(2)}, ..., x^{(n)}$
- **Output:** Set of clusters $C_1, C_2, ...C_k$
- **Initialization:** Randomly pick ***k*** centroids $z^{(1)}, z^{(2)}, \ldots, z^{(k)}$
- **Itereate** until convergence or up to iterations $T$
  - **Assignment:** Assign each point to its closest centroid

    $$\text{for each } j = 1, .., k$$
    $$C_j = \{i | \text{s.t. } x^{(i)} \text{ is closest to } z^{(j)}\}$$

  - **Update:** Recompute centroids with newly assigned points

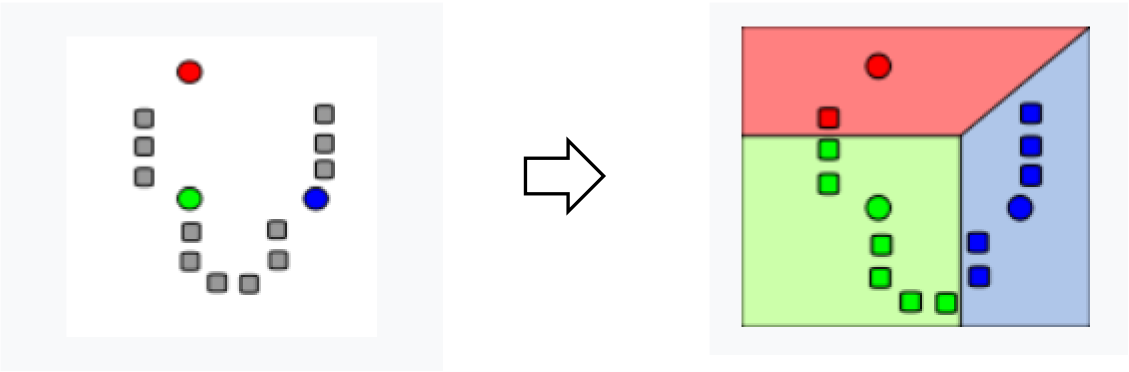
    $$z^{(j)} = \frac{1}{|C_j|} \sum_{i \in C_j} x^{(i)}$$

# K-Means



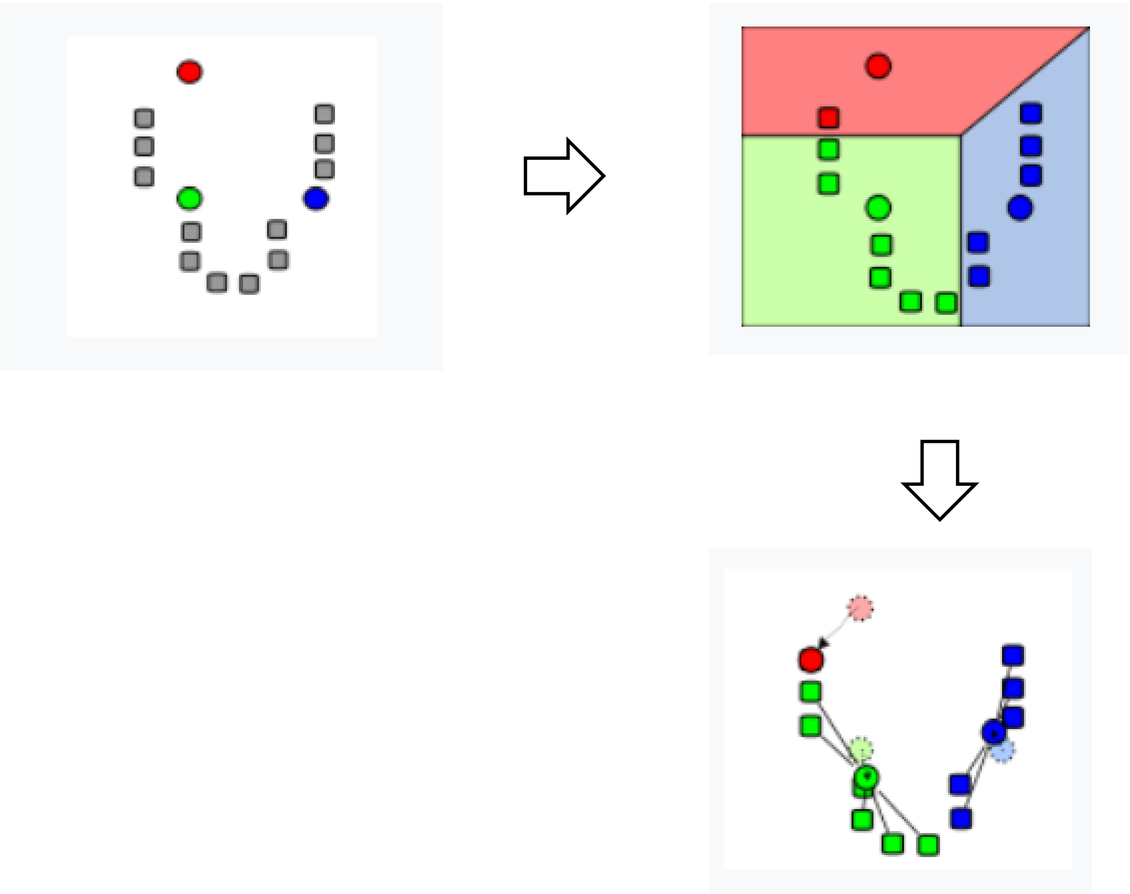
Courtesy: wiki

# K-Means



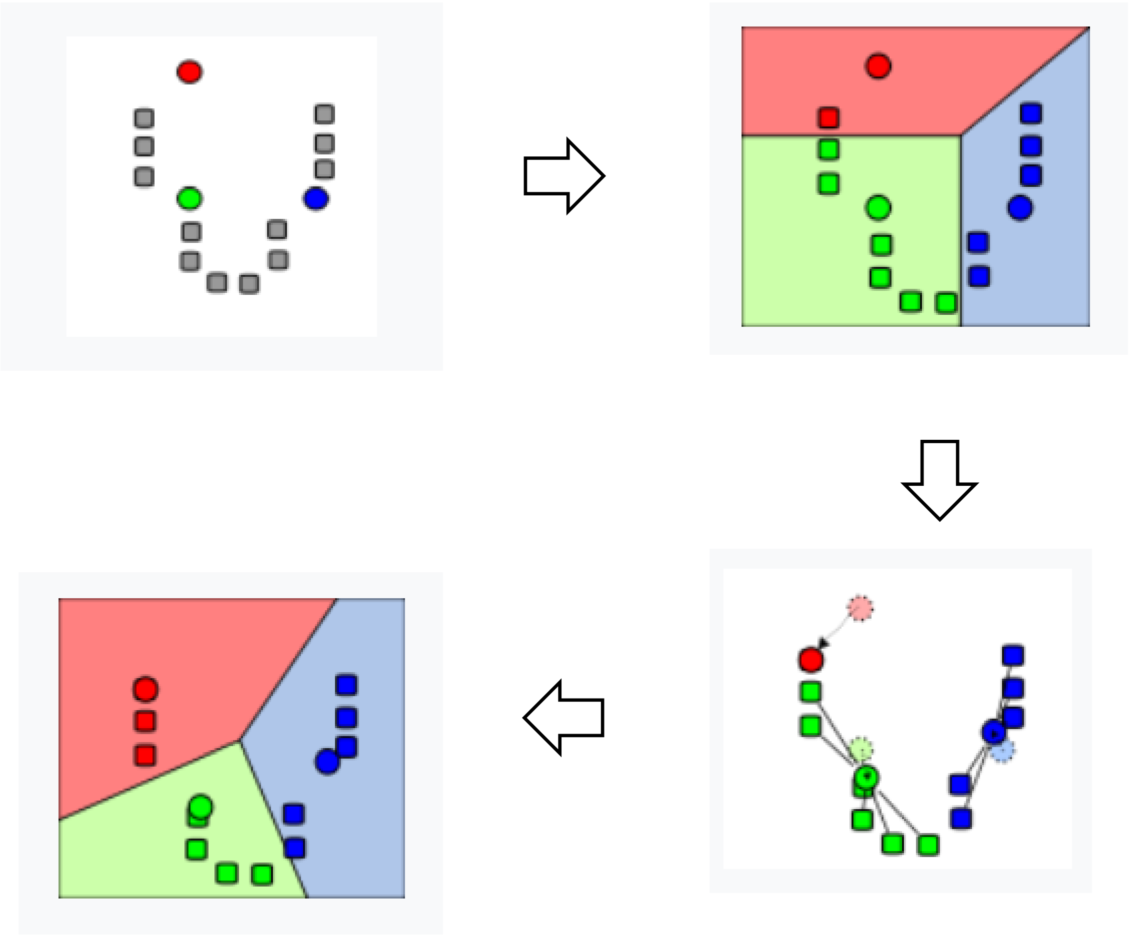
Courtesy: wiki

# K-Means



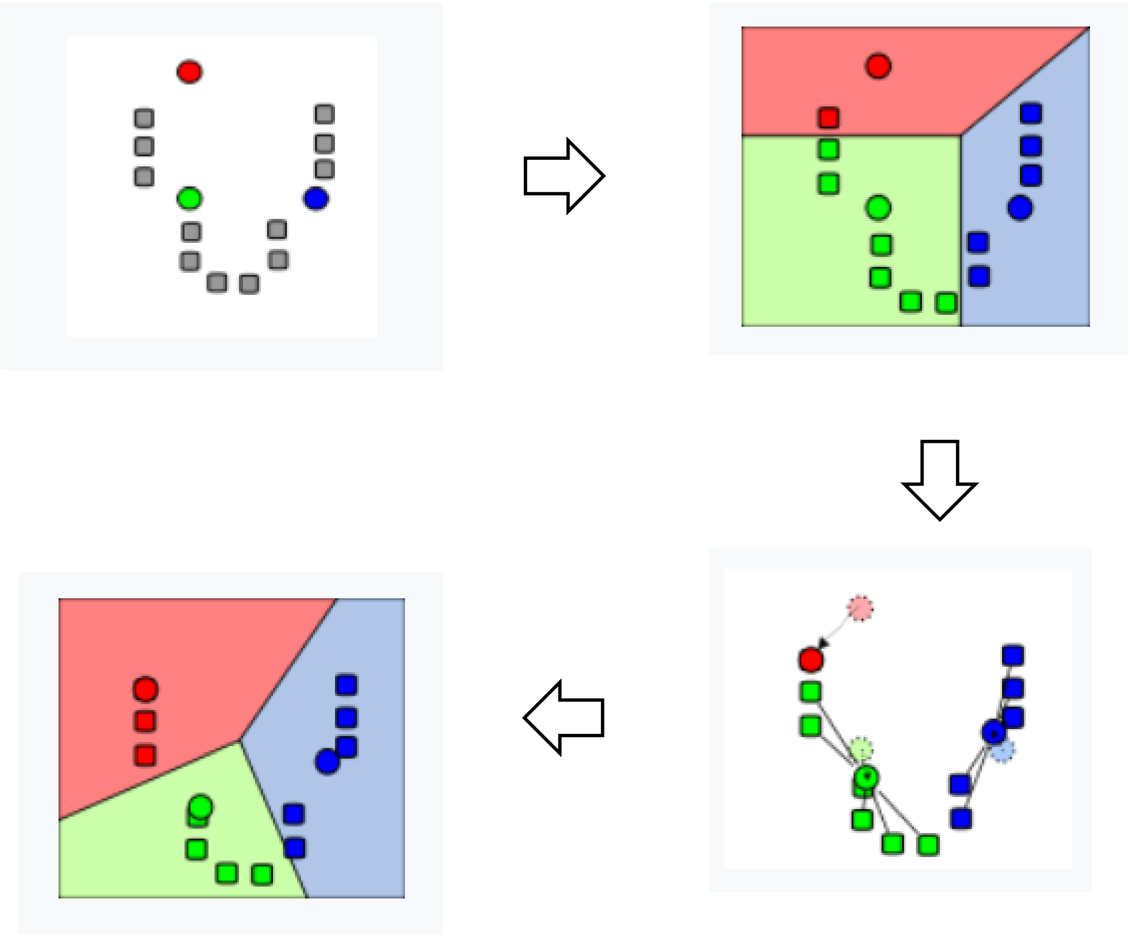
Courtesy: wiki

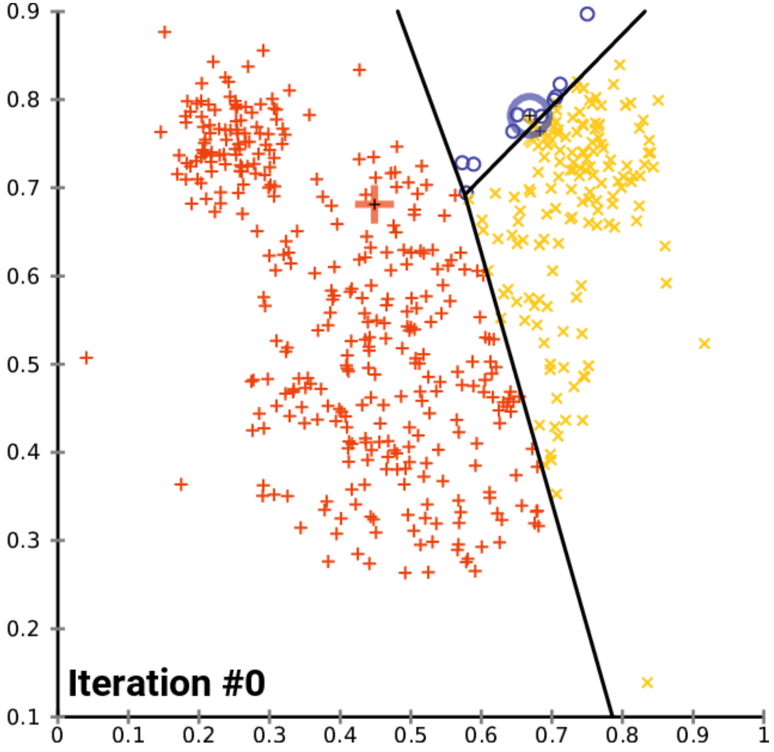
# K-Means



Courtesy: wiki

# K-Means






Courtesy: wiki

# K-Means

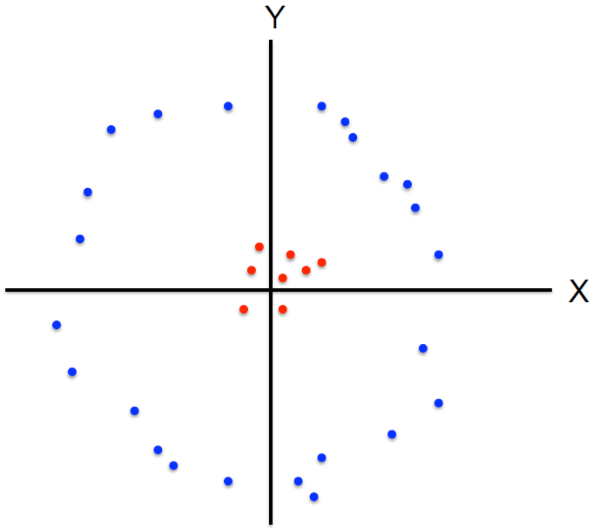
- Properties
  - guaranteed to converge in a finite iterations
    - at each iteration the error reduces
  - running time
    - data assignment to the closest cluster: $O(kN)$
    - update the means : $O(N)$
    - Total complexity : $O(kNT)$
  - global minima?

# K-Means

- Properties
  - guaranteed to converge in a finite iterations
    - at each iteration the error reduces
  - running time
    - data assignment to the closest cluster: $O(kN)$
    - update the means : $O(N)$
    - Total complexity : $O(kNT)$
  - global minima?
  - feature type selection also plays imp role
  - how should we choose initial clusters?
    - no good method available
    - do multiple runs and choose best results
  - similarity function choice?
    - euclidean
    - cosine

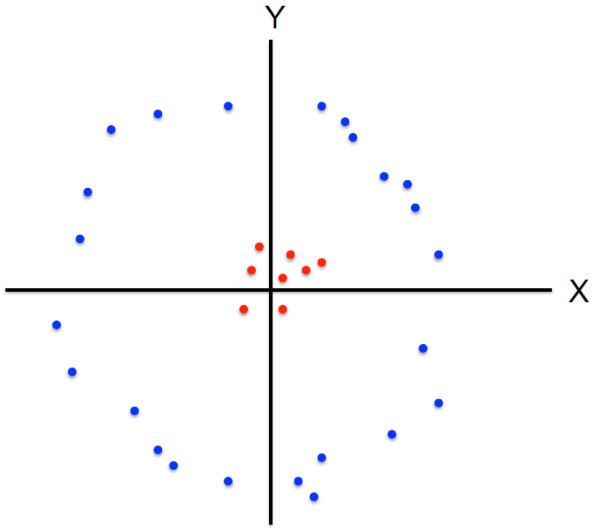
# K-Means

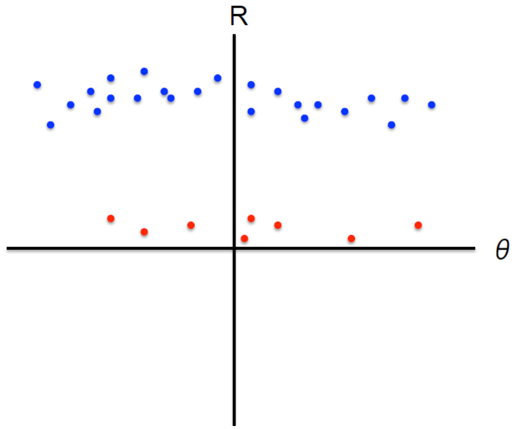
- Properties
  - guaranteed to converge in a finite iterations
    - at each iteration the error reduces
  - running time
    - data assignment to the closest cluster: $O(kN)$
    - update the means : $O(N)$
    - Total complexity : $O(kNT)$
  - global minima?
  - feature type selection also plays imp role
  - how should we choose initial clusters?
    - no good method available
    - do multiple runs and choose best results
  - similarity function choice?
    - euclidean
    - cosine



# K-Means

- Properties
  - guaranteed to converge in a finite iterations
    - at each iteration the error reduces
  - running time
    - data assignment to the closest cluster: $O(kN)$
    - update the means : $O(N)$
    - Total complexity : $O(kNT)$
  - global minima?
  - feature type selection also plays imp role
  - how should we choose initial clusters?
    - no good method available
    - do multiple runs and choose best results
  - similarity function choice?
    - euclidean
    - cosine



# K-Means

- Convergence

$$\min_{z^{(1)},\ldots,z^{(k)}} \min_{C_1,..C_k} \sum_{j=1}^{k} \sum_{i \in C_j} \|x^{(i)} - z^{(j)}\|^2$$

# K-Means

- Convergence

$$\min_{z^{(1)},\ldots,z^{(k)}} \min_{C_1,..C_k} \sum_{j=1}^{k} \sum_{i \in C_j} \|x^{(i)} - z^{(j)}\|^2$$

  - assignment: optimize $C$ with fixed $z$

$$\min_{C_1,..C_k} \sum_{j=1}^{k} \sum_{i \in C_j} \|x^{(i)} - z^{(j)}\|^2 = \sum_{i=1}^{n} \min_{j=1:k} \|x^{(i)} - z^{(j)}\|^2$$

# K-Means

- Convergence

$$\min_{z^{(1)},\dots,z^{(k)}} \min_{C_1,..C_k} \sum_{j=1}^{k} \sum_{i \in C_j} \|x^{(i)} - z^{(j)}\|^2$$

  - assignment: optimize $C$ with fixed $z$

$$\min_{C_1,..C_k} \sum_{j=1}^{k} \sum_{i \in C_j} \|x^{(i)} - z^{(j)}\|^2 = \sum_{i=1}^{n} \min_{j=1:k} \|x^{(i)} - z^{(j)}\|^2$$

  - update: fix $C$, optimize for $z$

$$J(z) = \min_{z^{(1)},\dots,z^{(k)}} \sum_{j=1}^{k} \sum_{i \in C_j} \|x^{(i)} - z^{(j)}\|^2$$

$$z^{(j)} = \frac{1}{|C_j|} \sum_{i \in C_j} x^{(i)}$$

$$\frac{\delta J(z)}{\delta z^{(j)}} \to 0$$

# K-Means

input

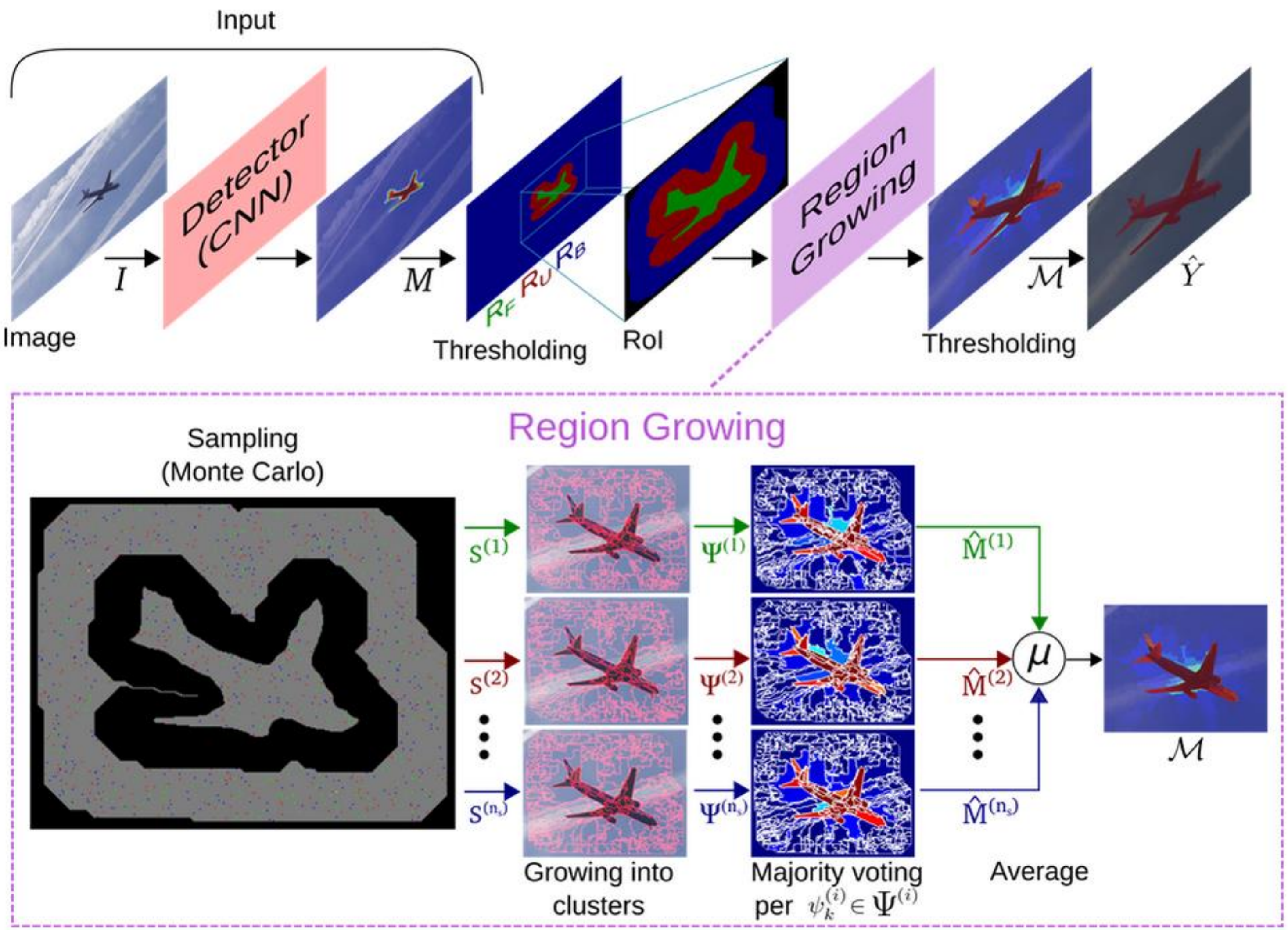
K = 2

K = 3

K = 10

Courtesy: D. Sontag

# Region growing in feature space

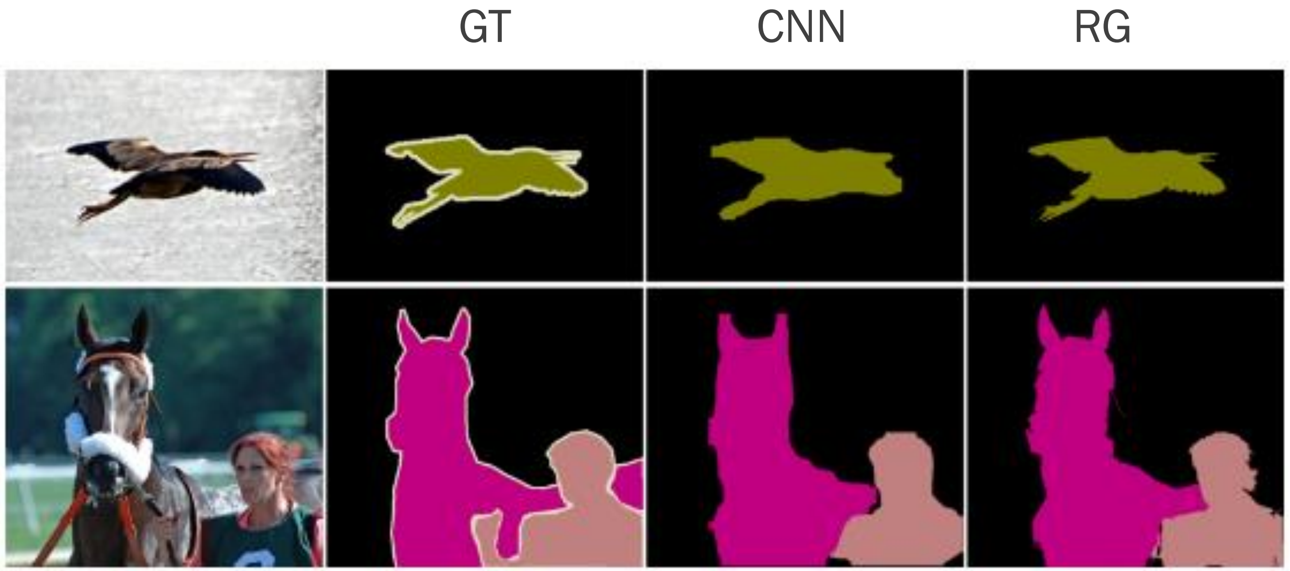



$$R_F = \{p_j | M(p_j) > \tau_F\},$$
$$R_U = \{p_j | \tau_B < M(p_j) < \tau_F\},$$
$$R_B = \{p_j | M(p_j) < \tau_B\},$$

Image credit: PA Dias

# Conclusion

- Regions
- Clustering



# Conclusion

- Regions
- Clustering

- Region growing
- Region splitting & merging
- Clustering
  - K-means clustering