Artificial Intelligence

Fiona Yan Liu

Department of Computing

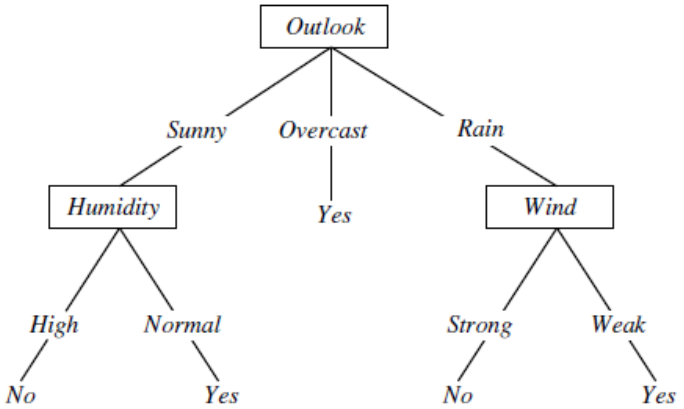
The Hong Kong Polytechnic University

Information of Teaching Staffs

- Fiona Yan Liu
  - Any question related with add/drop, lecture teaching and the final exam, please contact by yan.liu@polyu.edu.hk
- Zhi Zhang
  - Any question related with group project grading, please contact by tczhangzhi@gmail.com
- Xiang Hao
  - Any question related with lab teaching from week 1 – week 9, please contact by 22123509r@connect.polyu.hk (afternoon classes)
- Junbao Pan
  - Any question related with lab teaching from week 1 – week 9, please contact by csjbpan@comp.polyu.edu.hk (evening class)
- JIANG Zhuohang
  - Any question related with lab teaching from week 11 – week 13, please contact by 24037322r@connect.polyu.hk
- ZHANG Yujing
  - Any question related with quiz grading from quiz 4 – quiz 9, please contact by 24068516r@connect.polyu.hk
- Changhong Jing
  - Any question related with quiz grading from quiz 1 – quiz 3, please contact by 23123218r@connect.polyu.hk
  - General query

Learning Outcomes of Lecture 04

- Understand the related concepts of knowledge-based agent
- Master how to use knowledge base to solve the problem
- Understand the related concepts of decision tree
- Master how to utilize the decision tree to construct knowledge base



Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Agenda

- Machine learning
- Clustering
- Classification

Agenda

- Machine learning
- Clustering
- Classification

Machine Learning

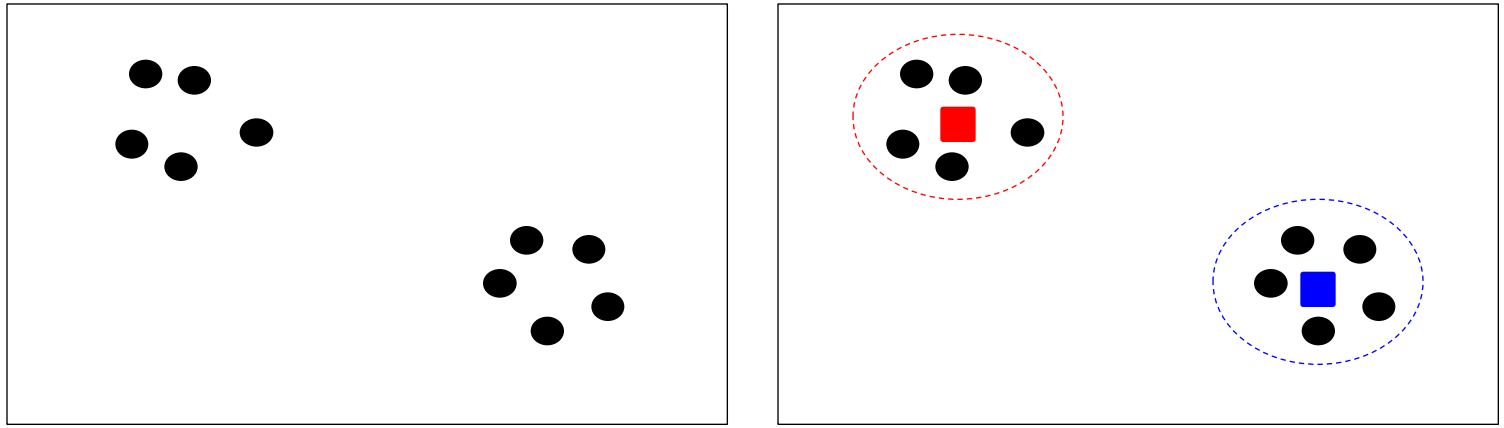
- Definition
  - A branch of artificial intelligence
  - Computational model that can learn from experience in the environment with respect to improve the performance of some tasks
- According to the feedback from environment
  - Unsupervised learning
  - Supervised learning
  - Reinforcement learning

Agenda

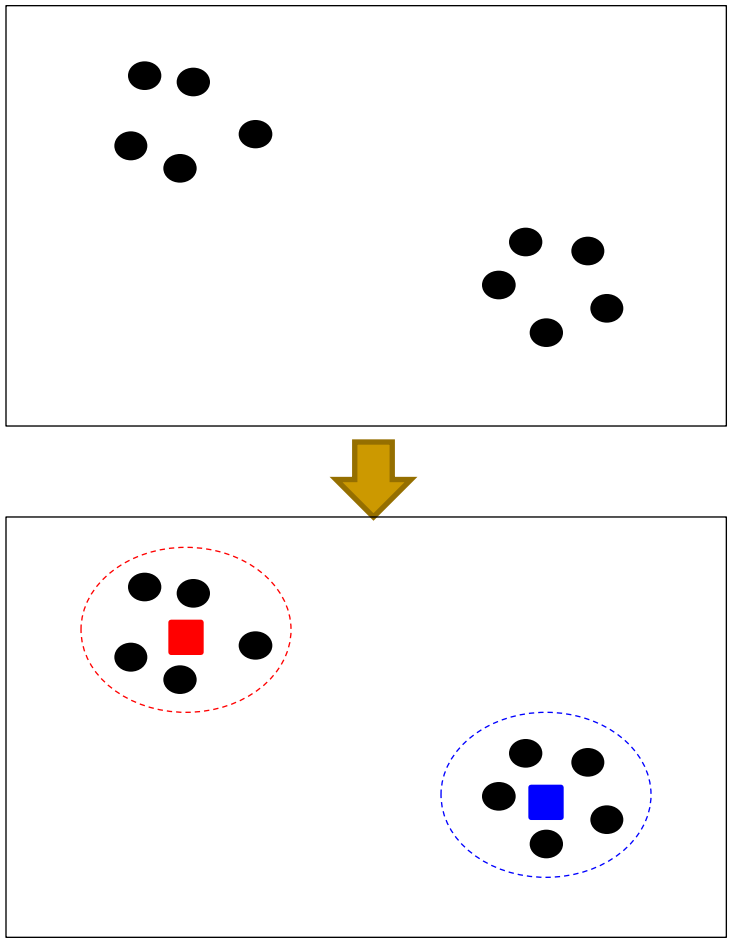
- Machine learning
- Clustering
- Classification

Clustering

- Clustering
  - Partition a given dataset into groups based on specified features so that the data points within a group are more similar to each other than the points in different groups
  - A typical unsupervised learning problem
  - Different clustering algorithms lead to different results



K-means clustering



Given

a set of observation $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$,
where each observation is
a *d*-dimensional real vector

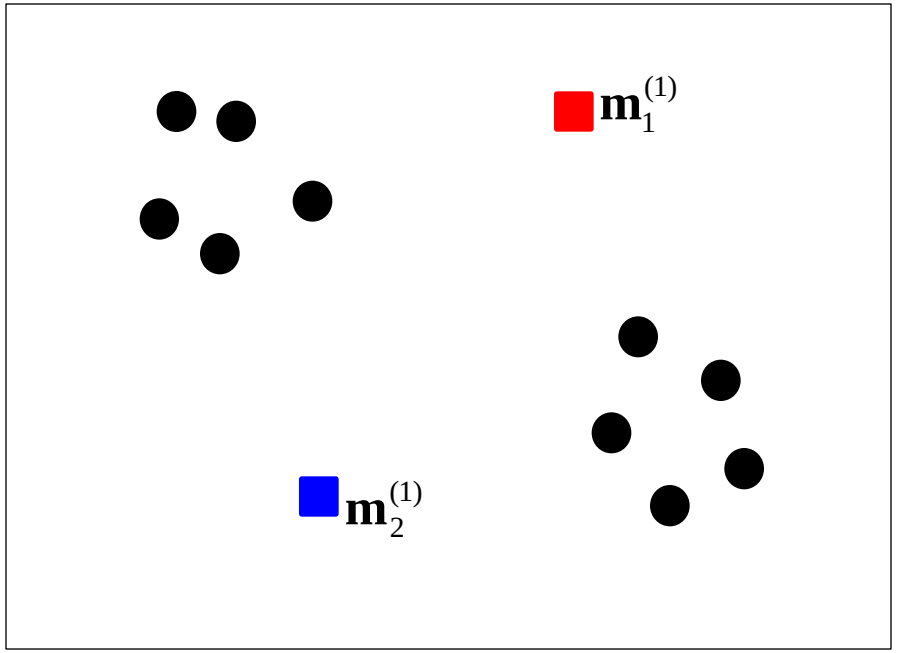
Aim to

partition observations into *K* sets
($K < N$), $\mathbf{S} = \{S_1, \cdots S_K\}$
so as to minimize
the within-cluster sum of squares:

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{K} \sum_{\mathbf{x}_j \in S_i} \| \mathbf{x}_j - \mu_i \|^2$$

where μ_i is the mean of S_i.

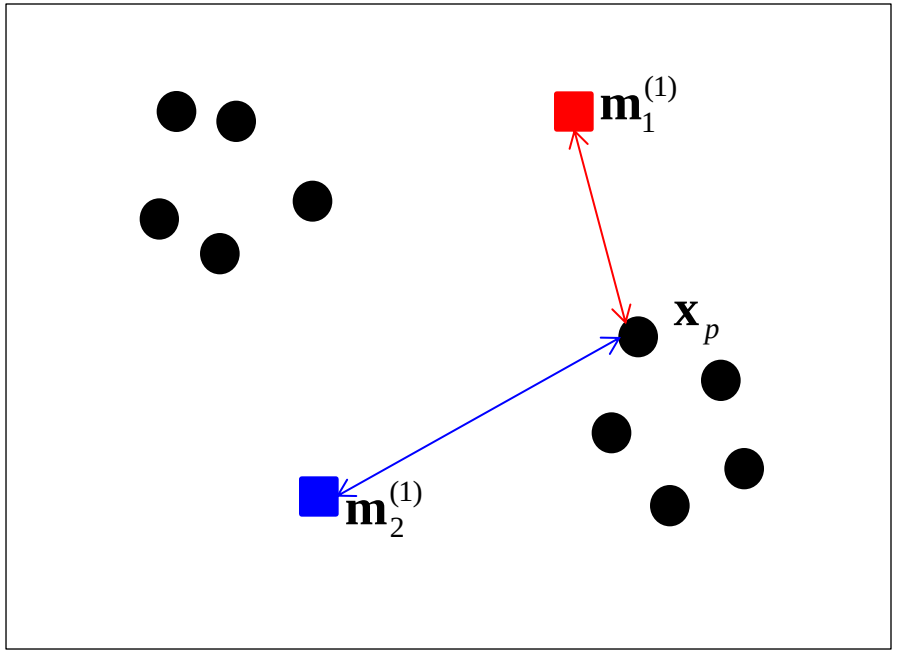
K means: details



K means begin with an initial guess to the centers: $\mathbf{m}_1^{(1)}$ and $\mathbf{m}_2^{(1)}$

K means: details

First iteration



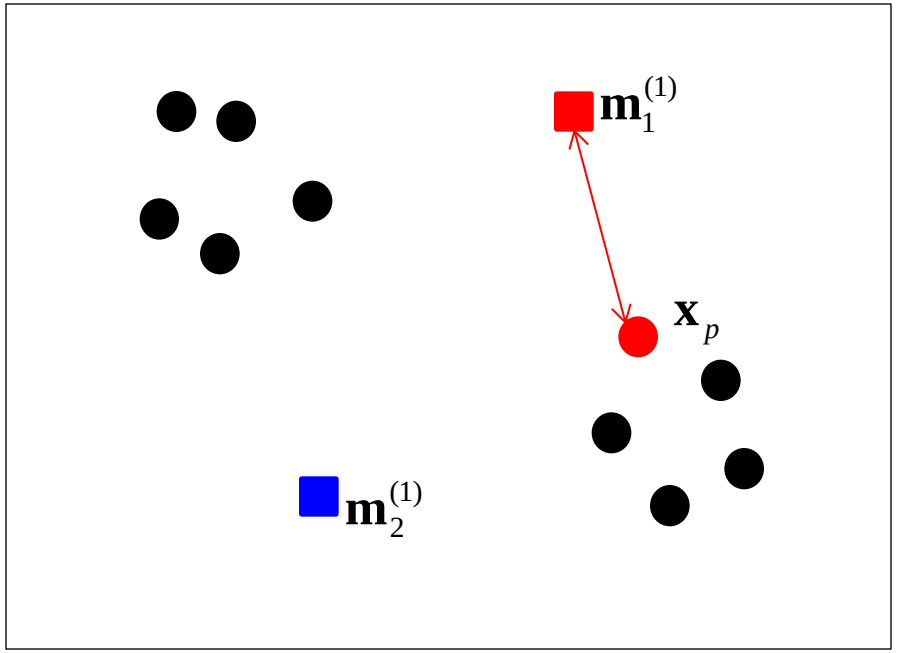
Calculate the Euclidean distance between each point and the centers

$$d^2(\mathbf{x}_p, \mathbf{m}_1^{(1)}) = \| \mathbf{x}_p - \mathbf{m}_1^{(1)} \|^2 = \sum_i (x_{pi} - m_{1i}^{(1)})^2$$

$$d^2(\mathbf{x}_p, \mathbf{m}_2^{(1)}) = \| \mathbf{x}_p - \mathbf{m}_2^{(1)} \|^2 = \sum_i (x_{pi} - m_{2i}^{(1)})^2$$

K means: details

First iteration



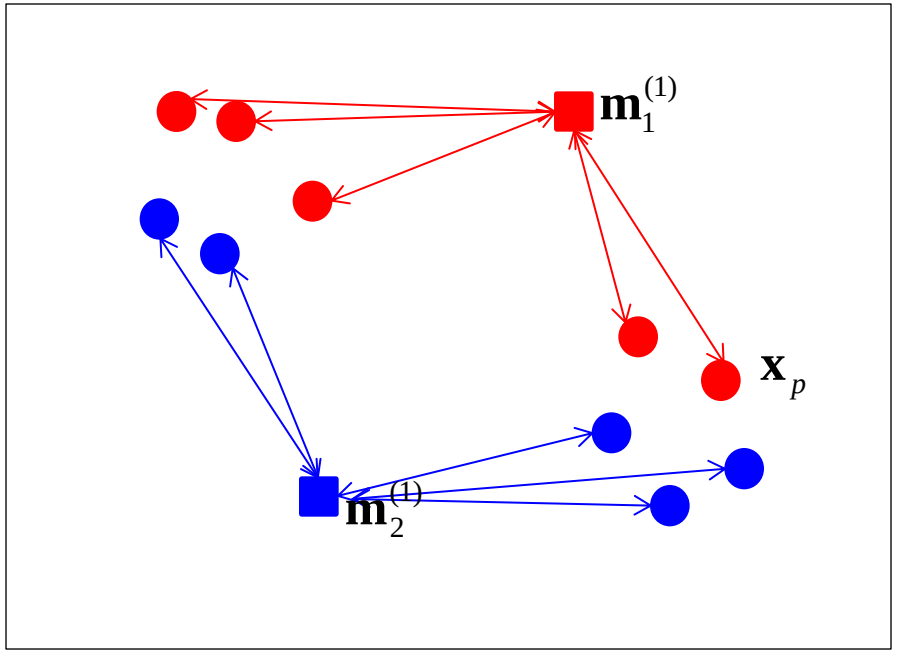
Assign the data point to the center with smallest distance:

$$d^2(\mathbf{x}_p, \mathbf{m}_1^{(1)}) < d^2(\mathbf{x}_p, \mathbf{m}_2^{(1)})$$

Then $\mathbf{x}_p$ is assigned to $\mathbf{m}_1^{(1)}$

K means: details

First iteration



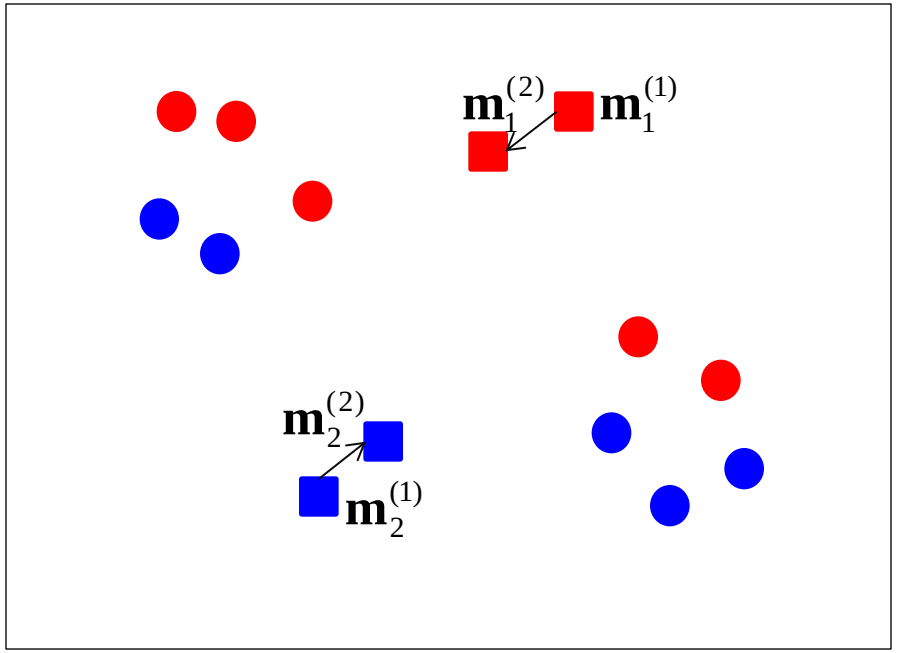
Assign all the other points to the centers in the same way as $\mathbf{x}_p$

$$S_1^{(1)} = \{\mathbf{x}_p : \| \mathbf{x}_p - \mathbf{m}_1^{(1)} \|^2 \leq \| \mathbf{x}_p - \mathbf{m}_2^{(1)} \|^2\}$$

$$S_2^{(1)} = \{\mathbf{x}_p : \| \mathbf{x}_p - \mathbf{m}_2^{(1)} \|^2 \leq \| \mathbf{x}_p - \mathbf{m}_1^{(1)} \|^2\}$$

K means: details

First iteration



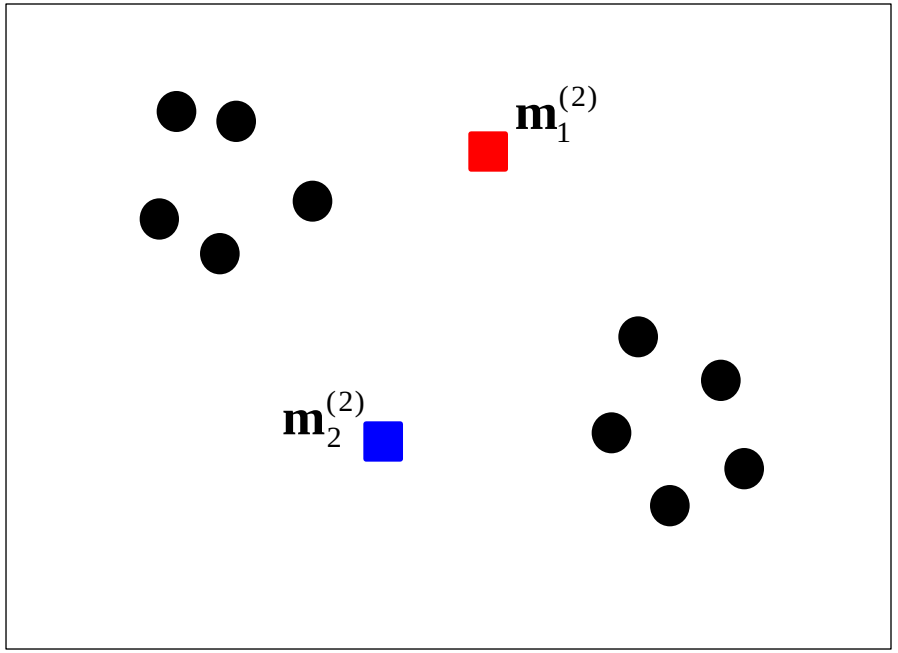
Update new centers by averaging the data points in each set:

$$\mathbf{m}_1^{(2)} = \sum_{\mathbf{x}_p \in S_1^{(1)}} \mathbf{x}_p / | S_1^{(1)} |$$

$$\mathbf{m}_2^{(2)} = \sum_{\mathbf{x}_p \in S_2^{(1)}} \mathbf{x}_p / | S_2^{(1)} |$$

K means: details

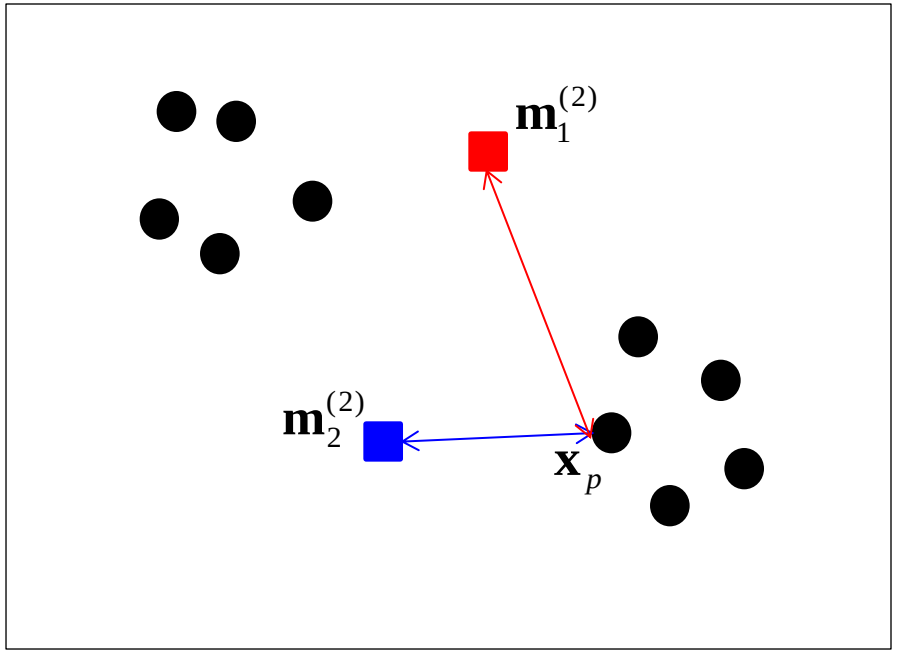
First iteration



The result after first iteration

K means: details

Second iteration



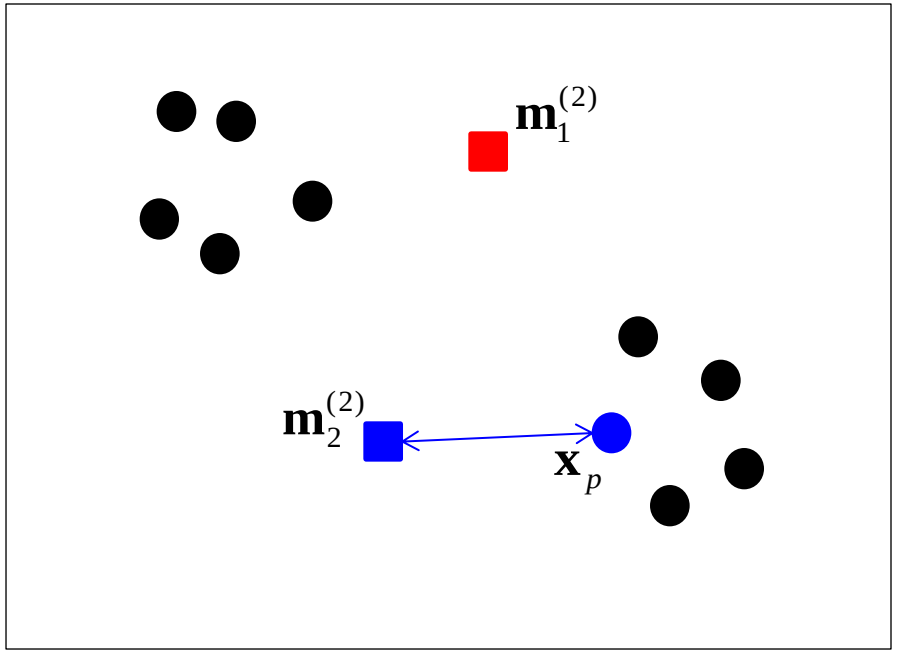
Calculate the Euclidean distance between each point and the centers

$$d^2(\mathbf{x}_p, \mathbf{m}_1^{(2)}) = \| \mathbf{x}_p - \mathbf{m}_1^{(2)} \|^2 = \sum_i (x_{pi} - m_{1i}^{(2)})^2$$

$$d^2(\mathbf{x}_p, \mathbf{m}_2^{(2)}) = \| \mathbf{x}_p - \mathbf{m}_2^{(2)} \|^2 = \sum_i (x_{pi} - m_{2i}^{(2)})^2$$

K means: details

Second iteration



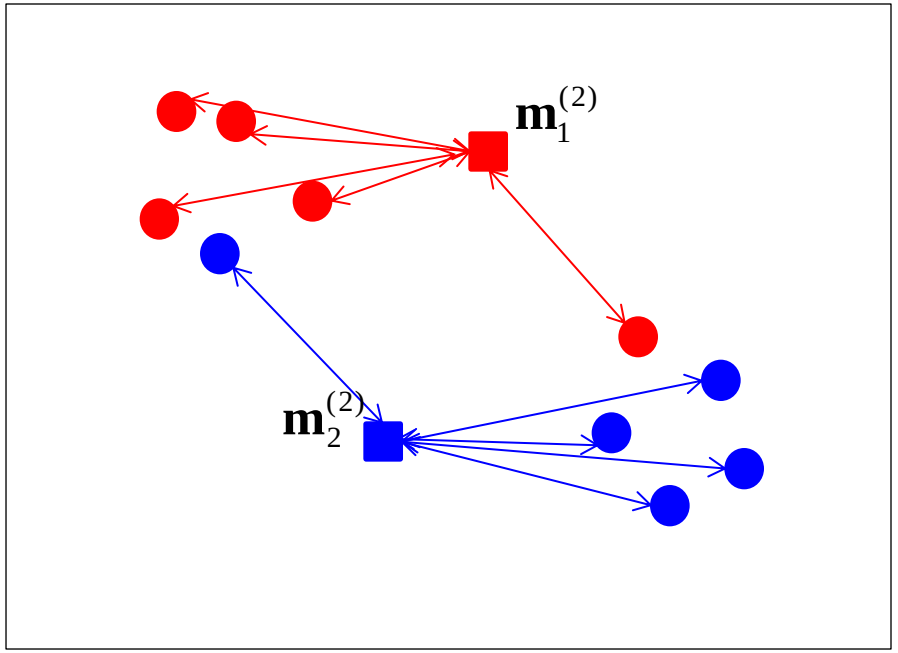
Assign the data point to the center with smallest distance:

$$d^2(\mathbf{x}_p, \mathbf{m}_2^{(2)}) < d^2(\mathbf{x}_p, \mathbf{m}_1^{(2)})$$

Then $\mathbf{x}_p$ is assigned to $\mathbf{m}_2^{(2)}$

K means: details

Second iteration

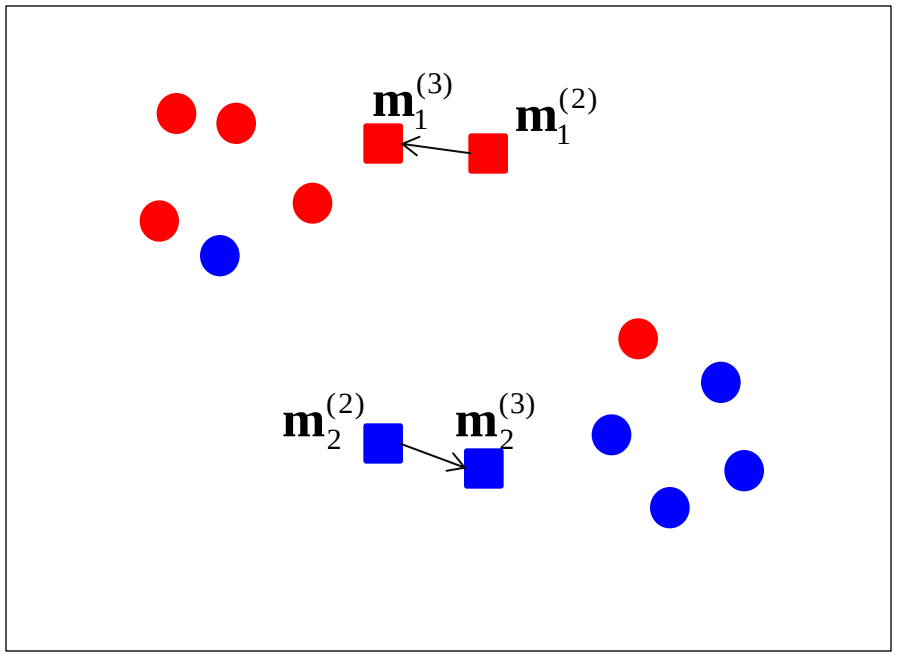


Assign all the other points to the centers in the same way as $\mathbf{x}_p$

$$S_1^{(2)} = \{\mathbf{x}_p : \| \mathbf{x}_p - \mathbf{m}_1^{(2)} \|^2 \leq \| \mathbf{x}_p - \mathbf{m}_2^{(2)} \|^2\}$$

$$S_2^{(2)} = \{\mathbf{x}_p : \| \mathbf{x}_p - \mathbf{m}_2^{(2)} \|^2 \leq \| \mathbf{x}_p - \mathbf{m}_1^{(2)} \|^2\}$$

K means: details



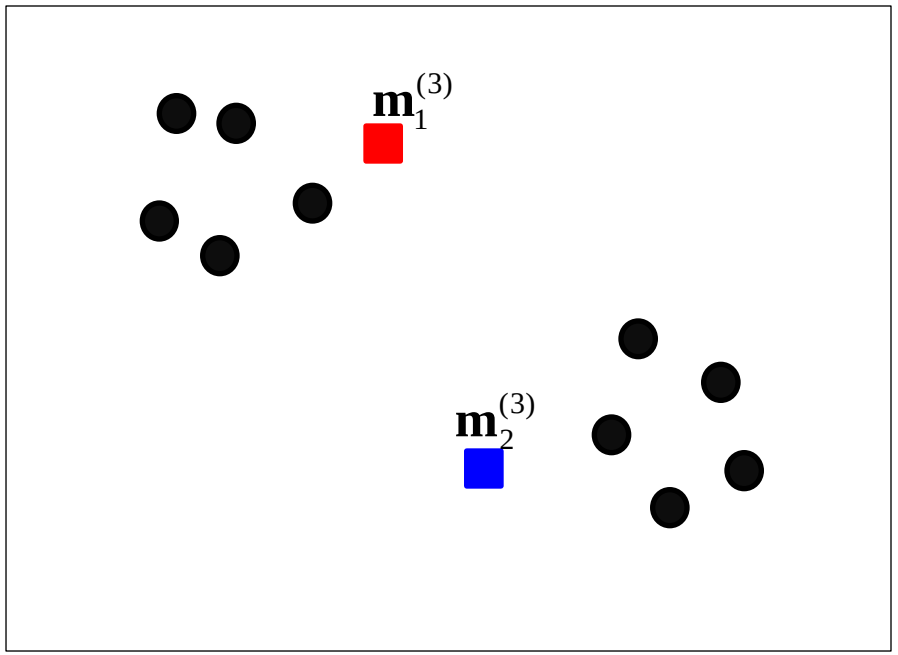
Update new centers by averaging the data points in each set:

$$\mathbf{m}_1^{(3)} = \sum_{\mathbf{x}_p \in S_1^{(2)}} \mathbf{x}_p / | S_1^{(2)} |$$

$$\mathbf{m}_2^{(3)} = \sum_{\mathbf{x}_p \in S_2^{(2)}} \mathbf{x}_p / | S_2^{(2)} |$$

K means: details

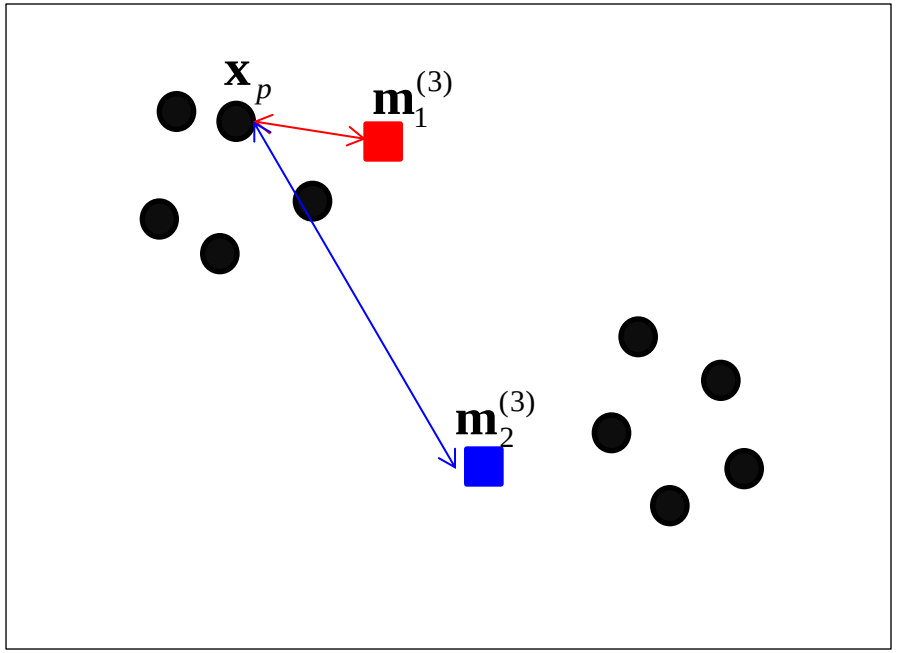
Second iteration



The result after second iteration

K means: details

Third iteration



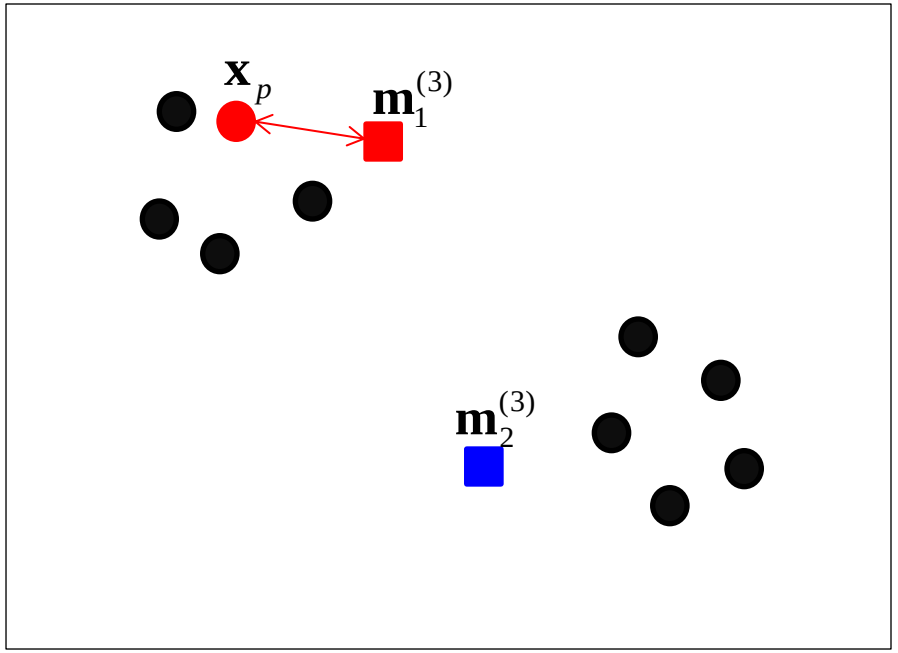
Calculate the Euclidean distance between each point and the centers

$$d^2(\mathbf{x}_p, \mathbf{m}_1^{(3)}) = \| \mathbf{x}_p - \mathbf{m}_1^{(3)} \|^2 = \sum_i (x_{pi} - m_{1i}^{(3)})^2$$

$$d^2(\mathbf{x}_p, \mathbf{m}_2^{(3)}) = \| \mathbf{x}_p - \mathbf{m}_2^{(3)} \|^2 = \sum_i (x_{pi} - m_{2i}^{(3)})^2$$

K means: details

Third iteration



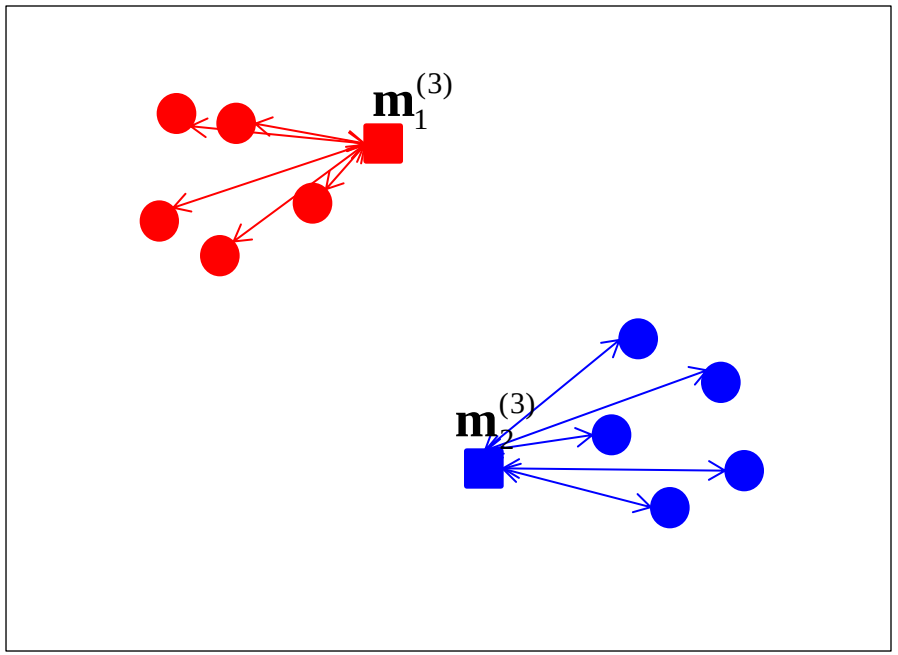
Assign the data point to the center with smallest distance:

$$d^2(\mathbf{x}_p, \mathbf{m}_1^{(3)}) < d^2(\mathbf{x}_p, \mathbf{m}_2^{(3)})$$

Then $\mathbf{x}_p$ is assigned to $\mathbf{m}_1^{(3)}$

K means: details

Third iteration

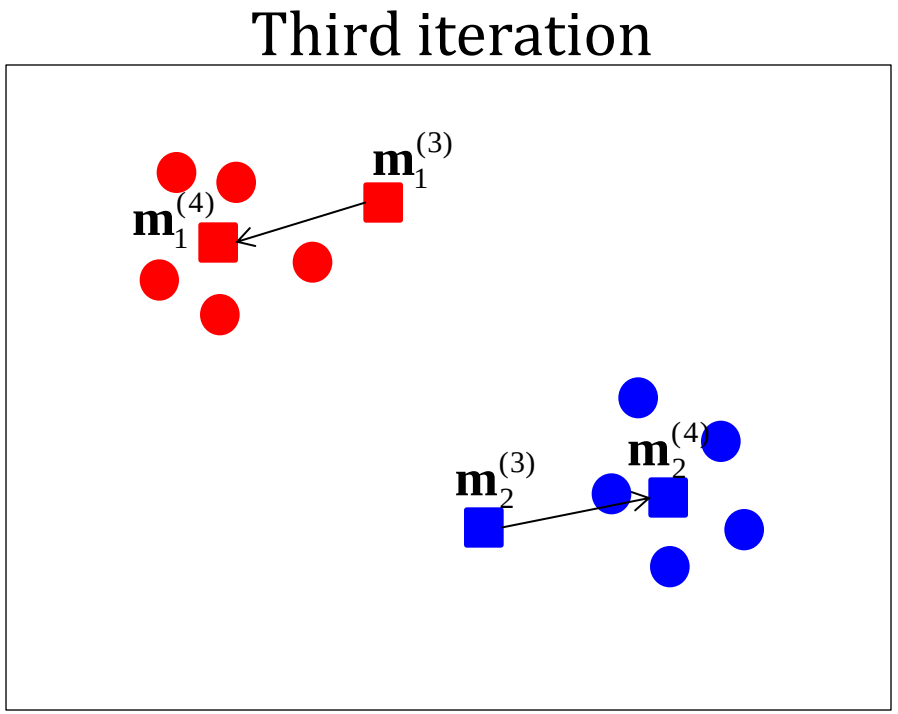


Assign all the other points to the centers in the same way as $\mathbf{x}_p$

$$S_1^{(3)} = \{\mathbf{x}_p : \| \mathbf{x}_p - \mathbf{m}_1^{(3)} \|^2 \leq \| \mathbf{x}_p - \mathbf{m}_2^{(3)} \|^2\}$$

$$S_2^{(3)} = \{\mathbf{x}_p : \| \mathbf{x}_p - \mathbf{m}_2^{(3)} \|^2 \leq \| \mathbf{x}_p - \mathbf{m}_1^{(3)} \|^2\}$$

K means: details



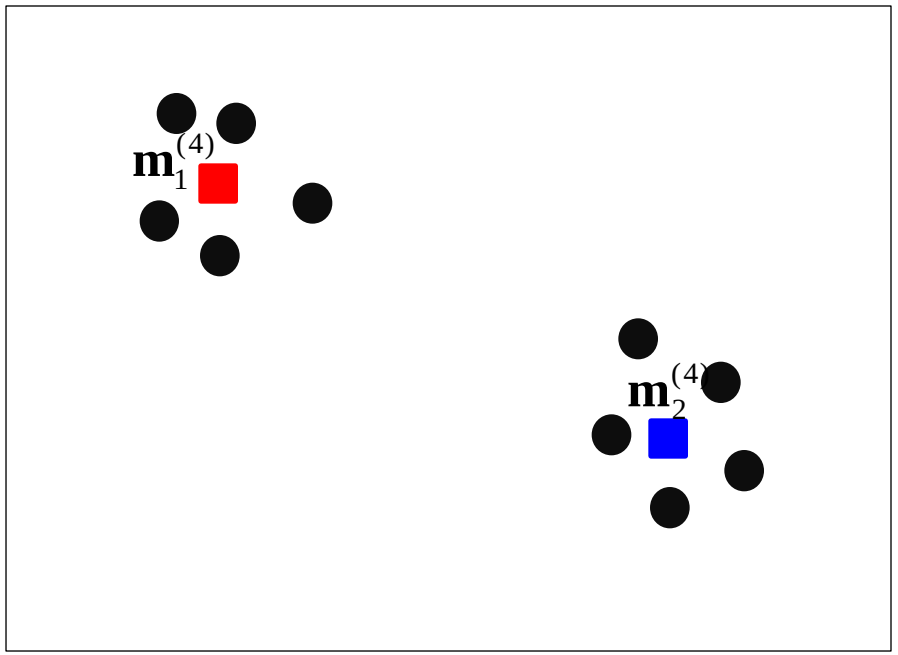
Update new centers by averaging the data points in each set:

$$\mathbf{m}_1^{(4)} = \sum_{\mathbf{x}_p \in S_1^{(3)}} \mathbf{x}_p / | S_1^{(3)} |$$

$$\mathbf{m}_2^{(4)} = \sum_{\mathbf{x}_p \in S_2^{(3)}} \mathbf{x}_p / | S_2^{(3)} |$$

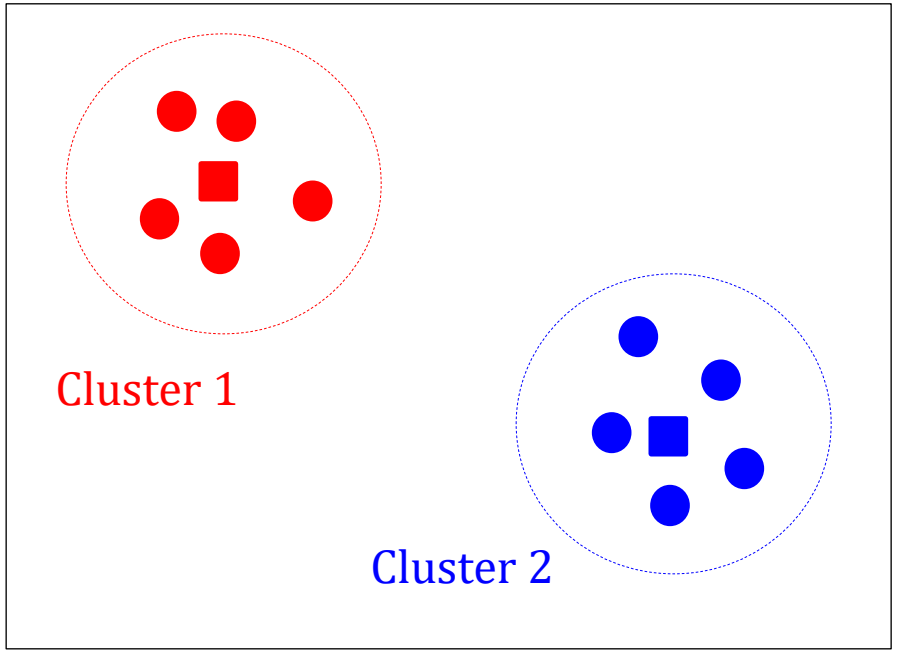
K means: details

Forth iteration



Converge !

K means: details



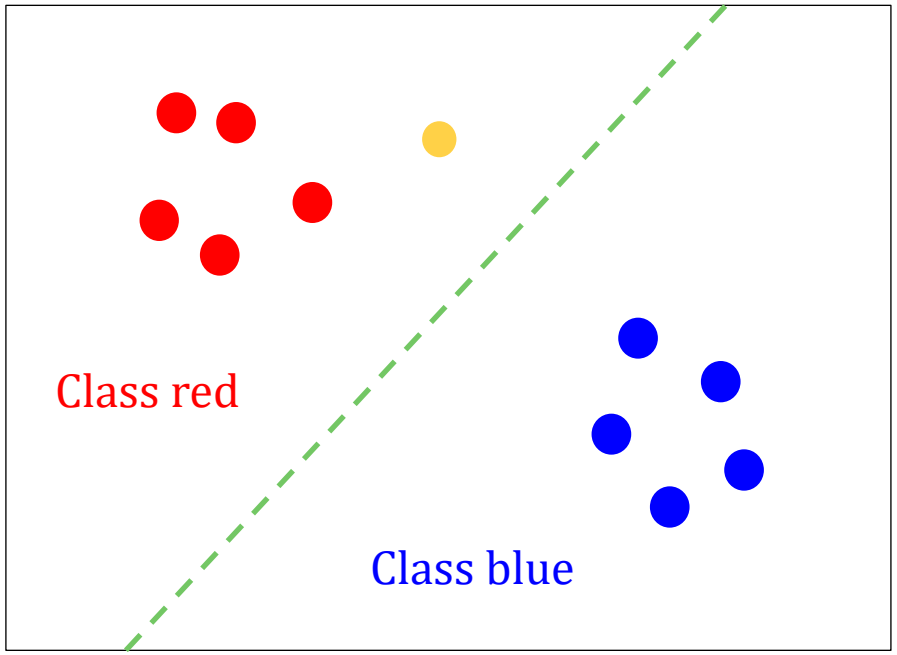
Clustering result

Agenda

- Machine learning
- Clustering
  - k means clustering
- Classification

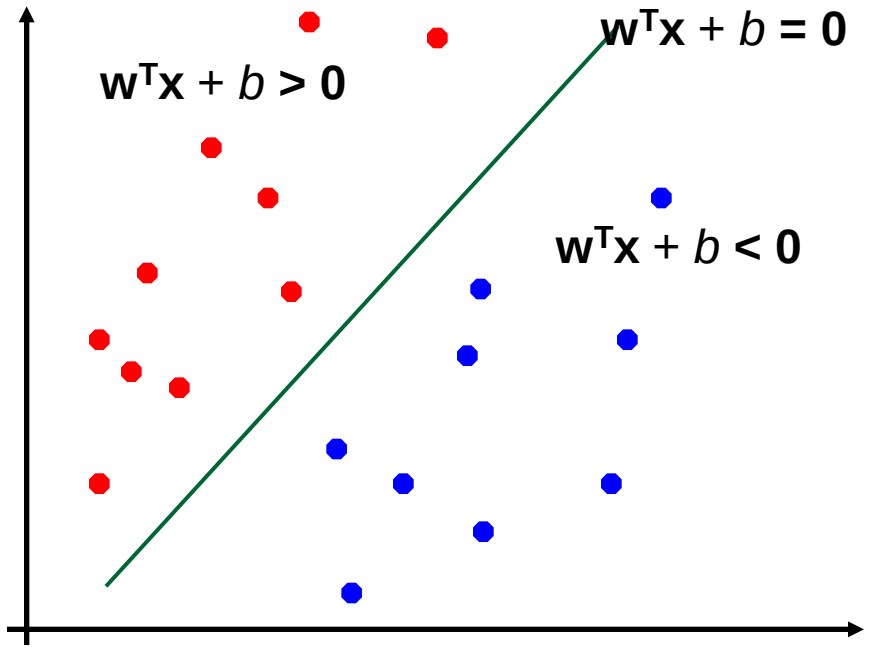
Classification

- Classification
  - Assign the data to the predefined label
- Training data
  - A set of data with predefined labels
- Test data
  - A set of data, which is used to validate the classification performance
  - The labels of the data are defined as ground truth
  - The labels from the computational model are defined as the classification results



Linear Classification

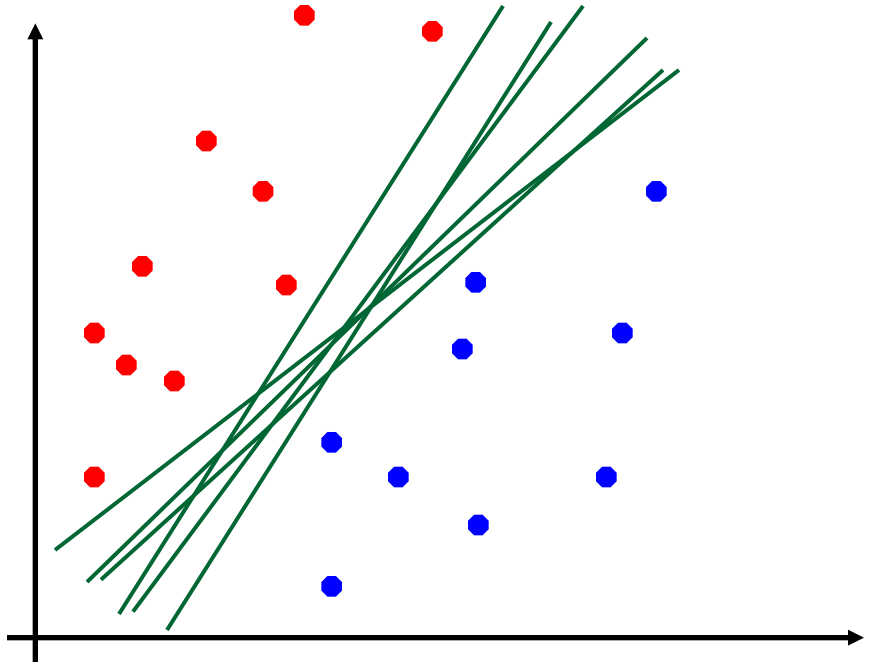
- Linear classification can be viewed as the task of finding the linear separator that can separate different classes in the feature space:



$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$

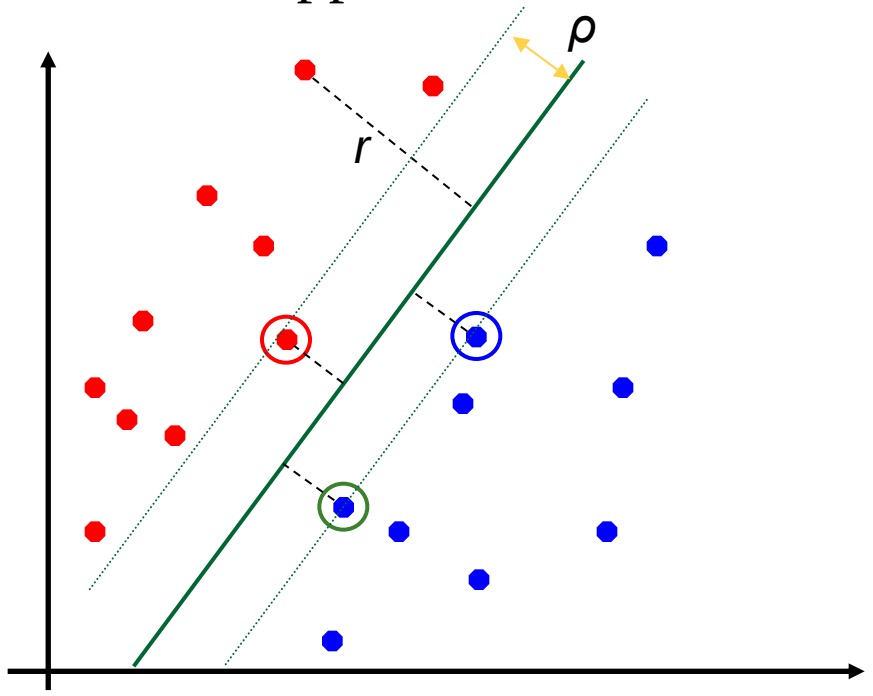
Multiple Solutions

- Which of the linear separators is the best?



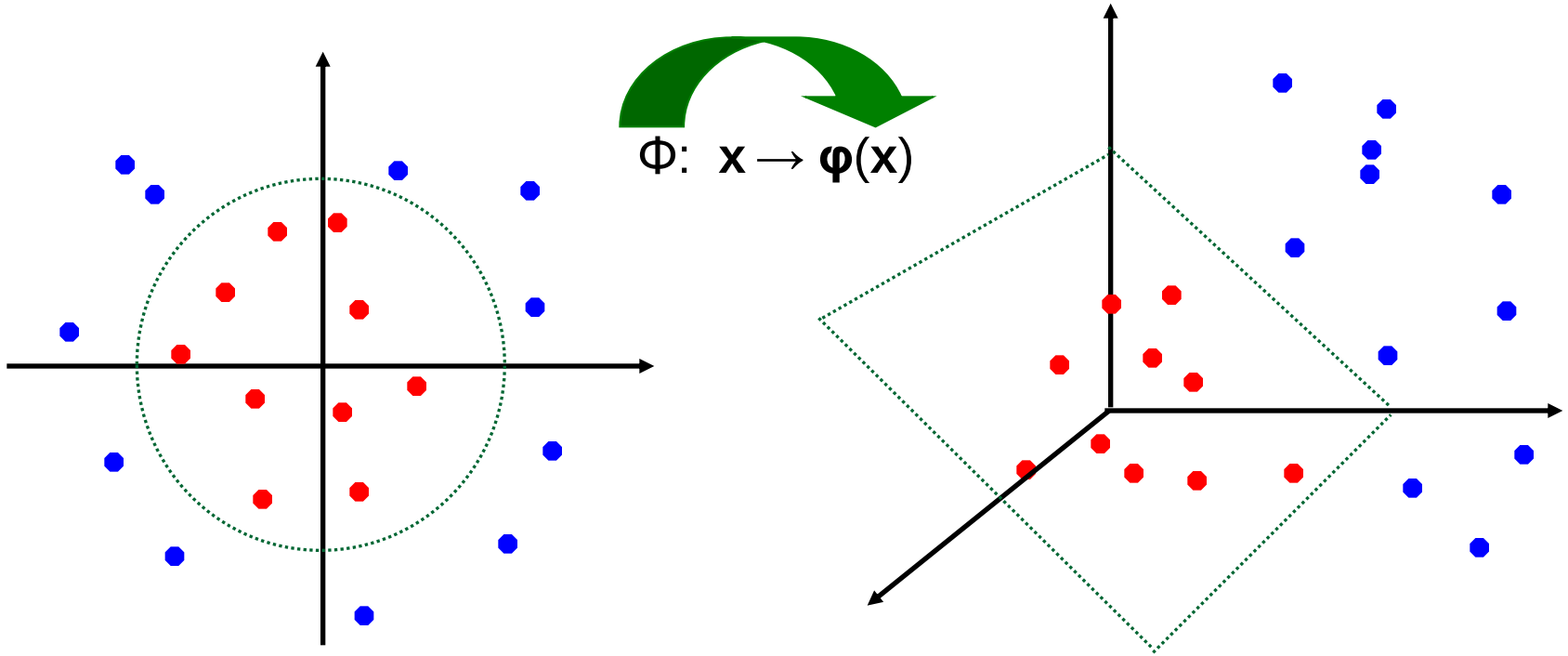
Classification Margin

- Distance from example $\mathbf{x}_i$ to the separator is $r = |(\mathbf{w}^T \mathbf{x}_i + b)| / \|\mathbf{w}\|$
- Data points closest to the hyperplane are ***support vectors***.
- ***Margin*** ρ of the separator is the distance between support vectors and the classification boundary
- Example
  - Given two data points
    - (5, 1) belonging to class red
    - (-1, -1) belonging to class blue
  - Calculate w1 and w0
    - Y = -3X+6
-



Nonlinear SVM

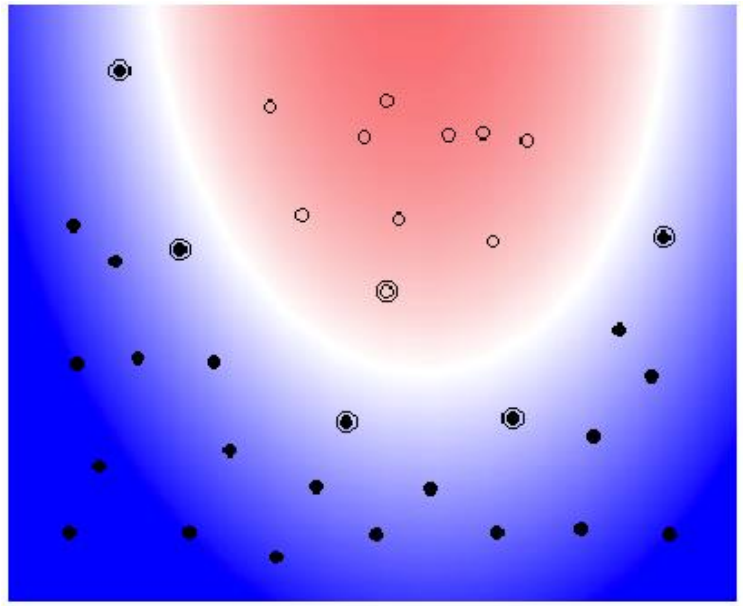
- General idea
  - The original feature space can always be mapped to some higher-dimensional feature space where the training set is separable



Examples of SVM with Different Kinds of Kernels

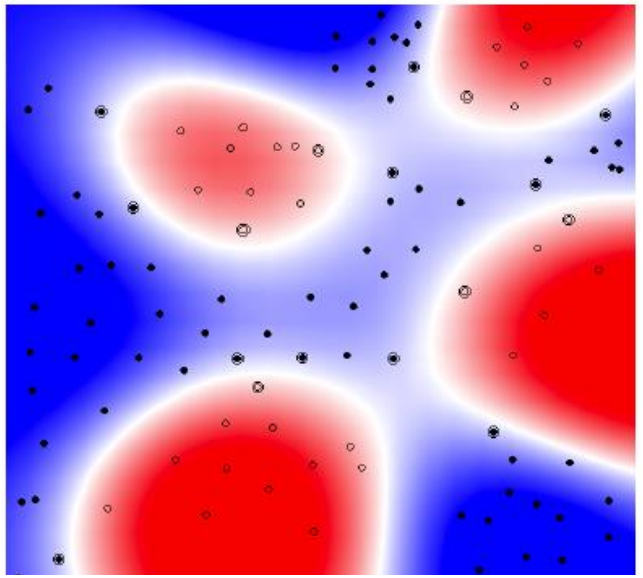
- Data points and classification boundary mapped to the feature space of the original data

Polynomial Kernel



$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = (\mathbf{x}^{(i)^T}\mathbf{x}^{(j)} + 1)^d$$

Radial-basis Function Kernel



$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp(-\frac{||\mathbf{x}^{(i)} - \mathbf{x}^{(j)}||^2}{2\sigma^2})$$