# Distance-based Classification summary

P. Papageorgiou, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.5

Artificial Intelligence & Information Analysis Lab

# Outline

- **Nearest neighbor classification**
- Supervised Learning Vector Quantization

# K-means Algorithm

- Distances between a feature vector and a class center:
  - Mahalanobis distance:

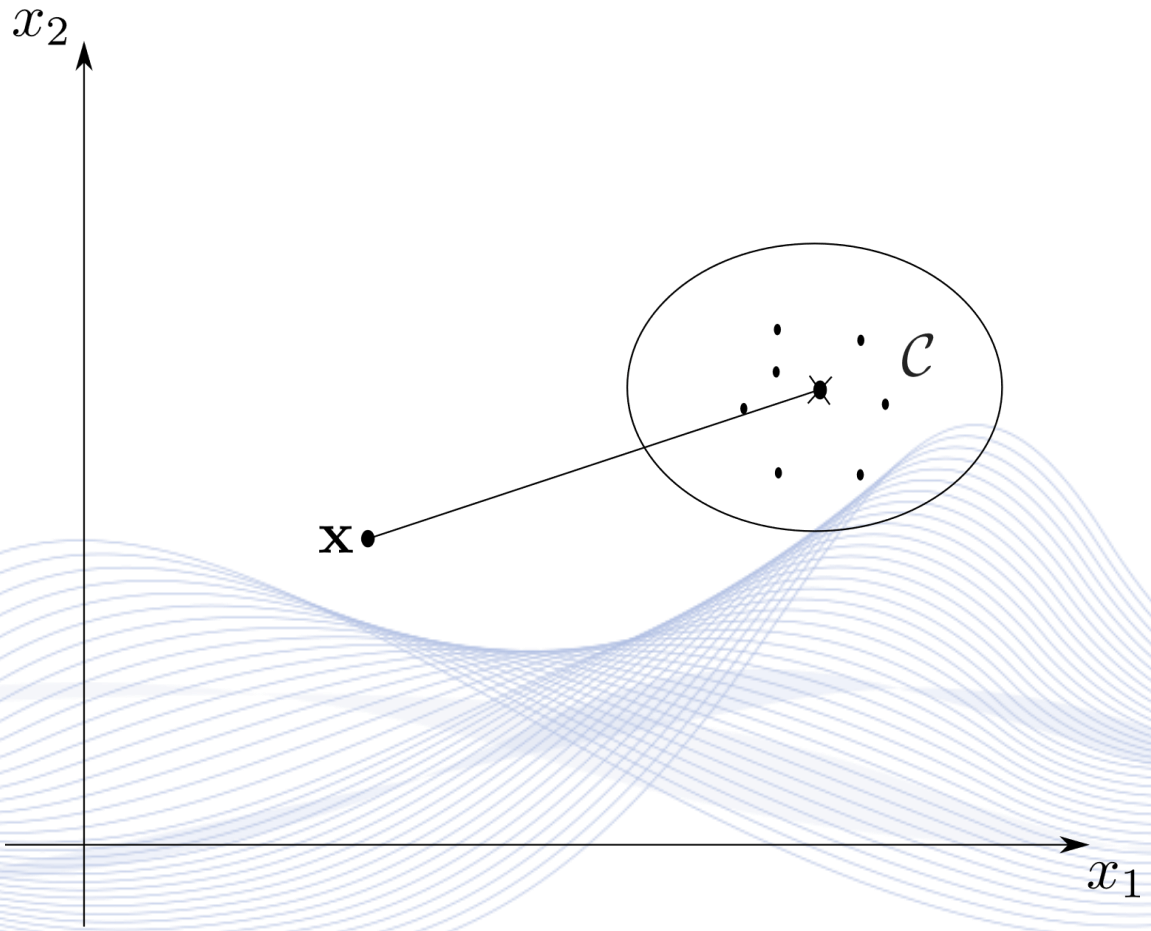
$$d(\mathbf{x}_i, \mathbf{m}_j) = (\mathbf{x}_i - \mathbf{m}_j)^T \mathbf{A} (\mathbf{x}_i - \mathbf{m}_j).$$

  - $\mathbf{A}$: symmetric, positive definite matrix.
  - Euclidean distance:

$$d(\mathbf{x}_i, \mathbf{m}_j) = (\mathbf{x}_i - \mathbf{m}_j)^T (\mathbf{x}_i - \mathbf{m}_j).$$

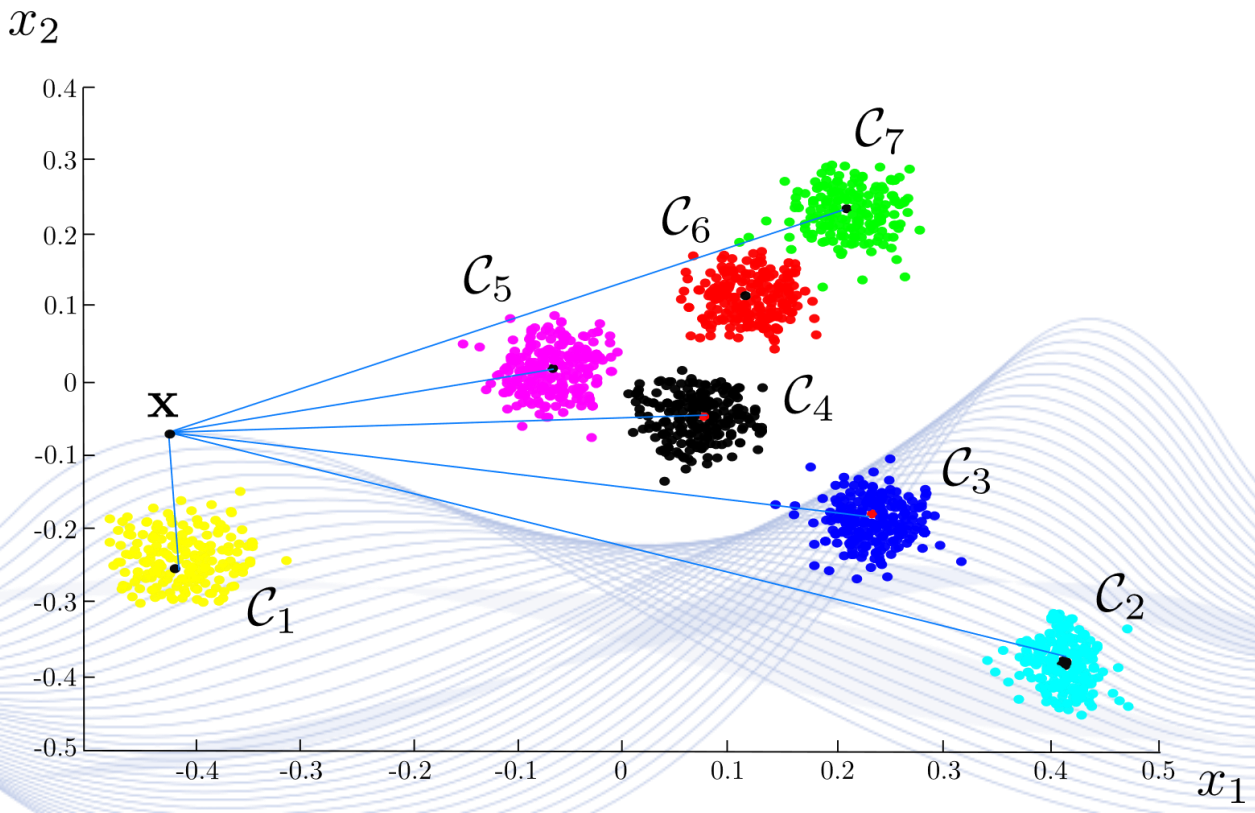
  - Minkowski distance: $d(\mathbf{x}_i, \mathbf{m}_j) = \left(\sum_{k=1}^{l} |\mathbf{x}_{ik} - \mathbf{m}_{jk}|^p\right)^{\frac{1}{p}}$
  - $\mathbf{x}_{ik}, \mathbf{m}_{jk}$ are the jth coordinates of $\mathbf{x}_i, \mathbf{m}_j$ respectively.

# Distance Measures

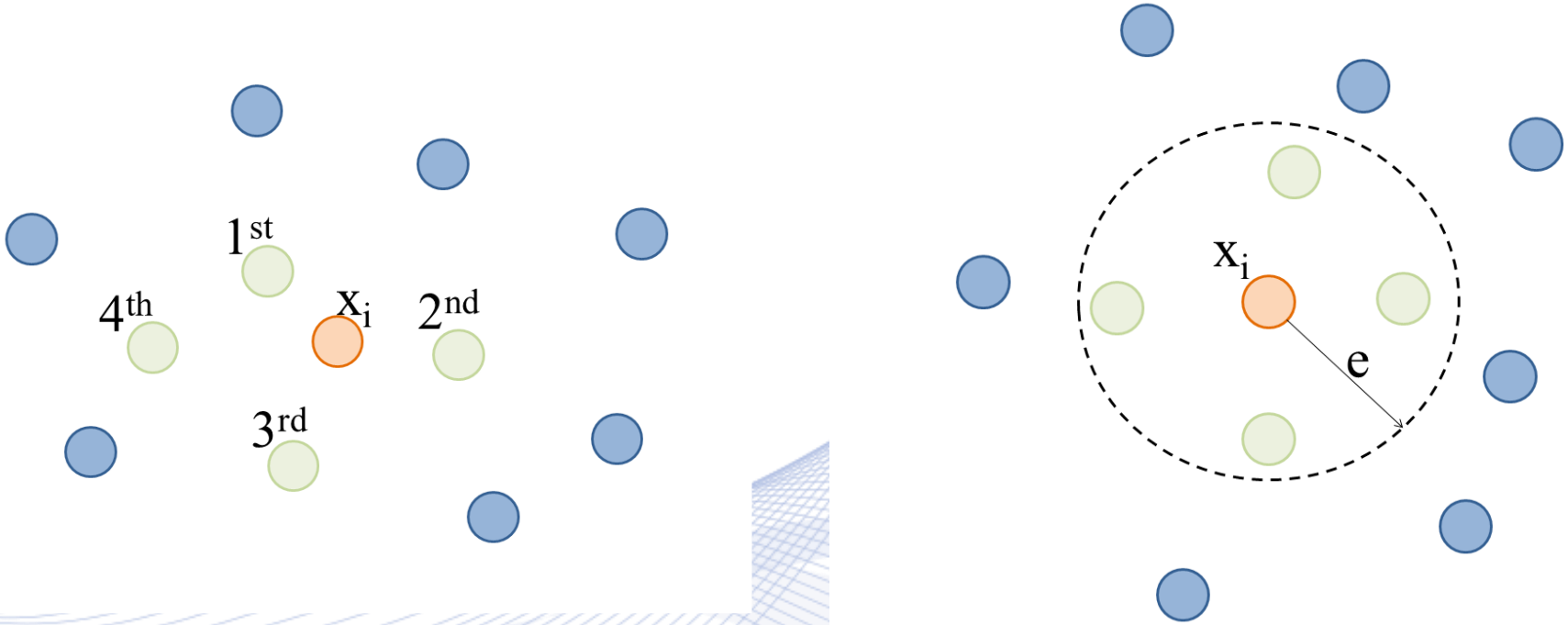



Distance between point and set (set center).

# Nearest class classification



# Nearest neighbor graphs




a) $k$-nearest neighbors graph; b) $e$-neighborhood graph.

# k-Nearest neighbor classification

- A data point $\mathbf{x}$ is to be classified to one of the classes $\mathcal{C}_i, i = 1, \dots, m$.
- A data class $\mathcal{C}_i$ is represented by a labeled data set: $\mathcal{C}_i = \{\mathbf{x}_{1i}, \mathbf{x}_{2i}, \dots, \mathbf{x}_{Ni}\}$.
- Classify $\mathbf{x}$ to the class $\mathcal{C}$, whose data vectors are most common in the $k$ −neighborhood of $\mathbf{x}$.

# Supervised Learning Vector Quantization

- A data point $\mathbf{x}$ is to be classified to one of the classes $\mathcal{C}_i, i = 1, \dots, m$.
- A data class $\mathcal{C}_i$ is represented by a labeled data set: $\mathcal{C}_i = \{\mathbf{x}_{1i}, \mathbf{x}_{2i}, \dots, \mathbf{x}_{Ni}\}$.
- Each class is represented by a class center $\mathbf{m}_i, i = 1, \dots, m$.

# Supervised Learning Vector Quantization

**Supervised LVQ training**

- $\mathbf{x}$: vector to be assigned to a class.
- Employ Euclidean distance.
- Find the optimal class centers $\mathbf{m}_i, i = 1, \ldots, m$.
- Find the closest class center $\mathbf{m}_k$:

$$d(\mathbf{x}, \mathbf{m}_k) = \min_i \{d(\mathbf{x}, \mathbf{m}_i)\}, \forall i \neq k \ .$$

# Supervised Learning Vector Quantization

- Winning class  center updating:

$$\mathbf{m}_k(t+1) = \mathbf{m}_k(t) + a(t)[\mathbf{x} - \mathbf{m}_k(t)]$$

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t), \qquad \text{for } i \neq k,$$

- $0 \leq a(t) \leq 1.$

# Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr