Data Mining

Classification II – Other Methods (Part B)
K-Nearest Neighbor Classifier

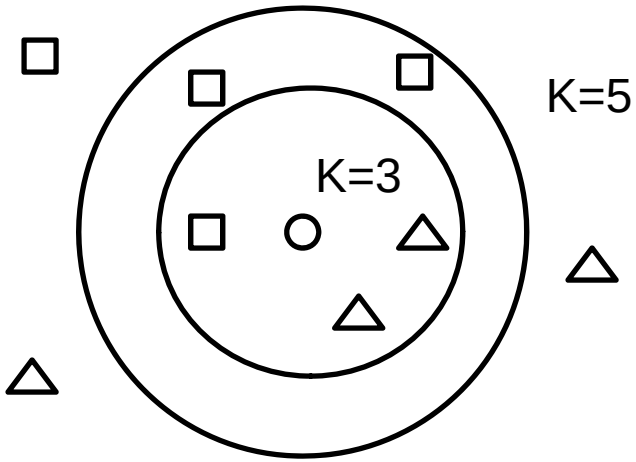
Dr. Jason T.L. Wang, Professor

Department of Computer Science

New Jersey Institute of Technology

K-Nearest Neighbor (KNN) Algorithm

- Classical algorithm in data mining
- An algorithm that classifies unlabeled objects based on the majority class of nearest neighbors



KNN Algorithm

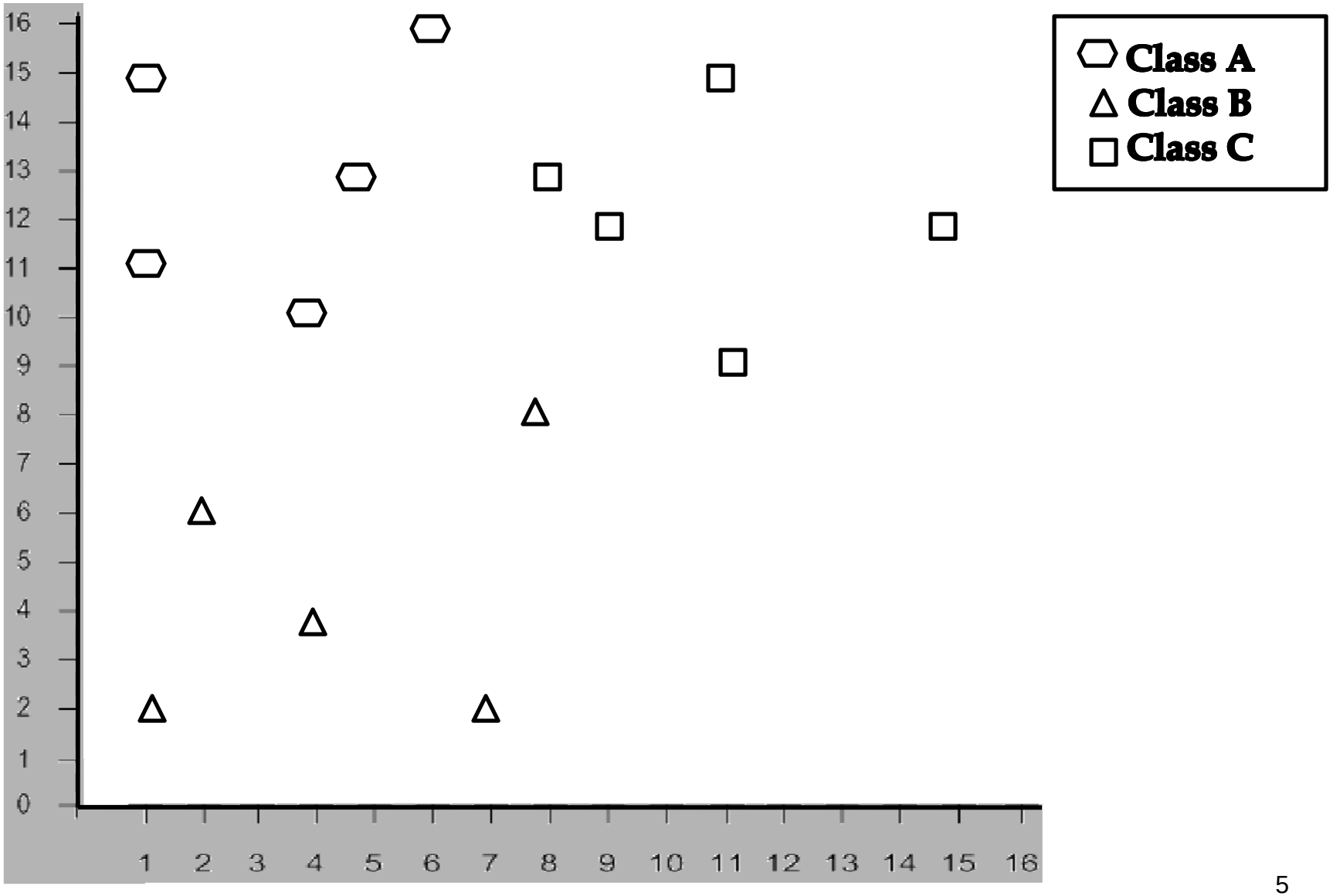
- Given a training dataset, KNN classifier learns the training dataset.

- K nearest neighbors are determined from the training set by calculating Euclidean distance.

KNN Algorithm

- K nearest neighbors are chosen that have minimum distance from the unlabeled test object.
- Once the nearest neighbors are determined, this unlabeled test object will be classified according to the rule of majority.
- Example: If K is 3 and 2 nearest neighbors belong to class A and 1 to B, then the unlabeled test object is assigned to class A.

Example KNN (Training Dataset)

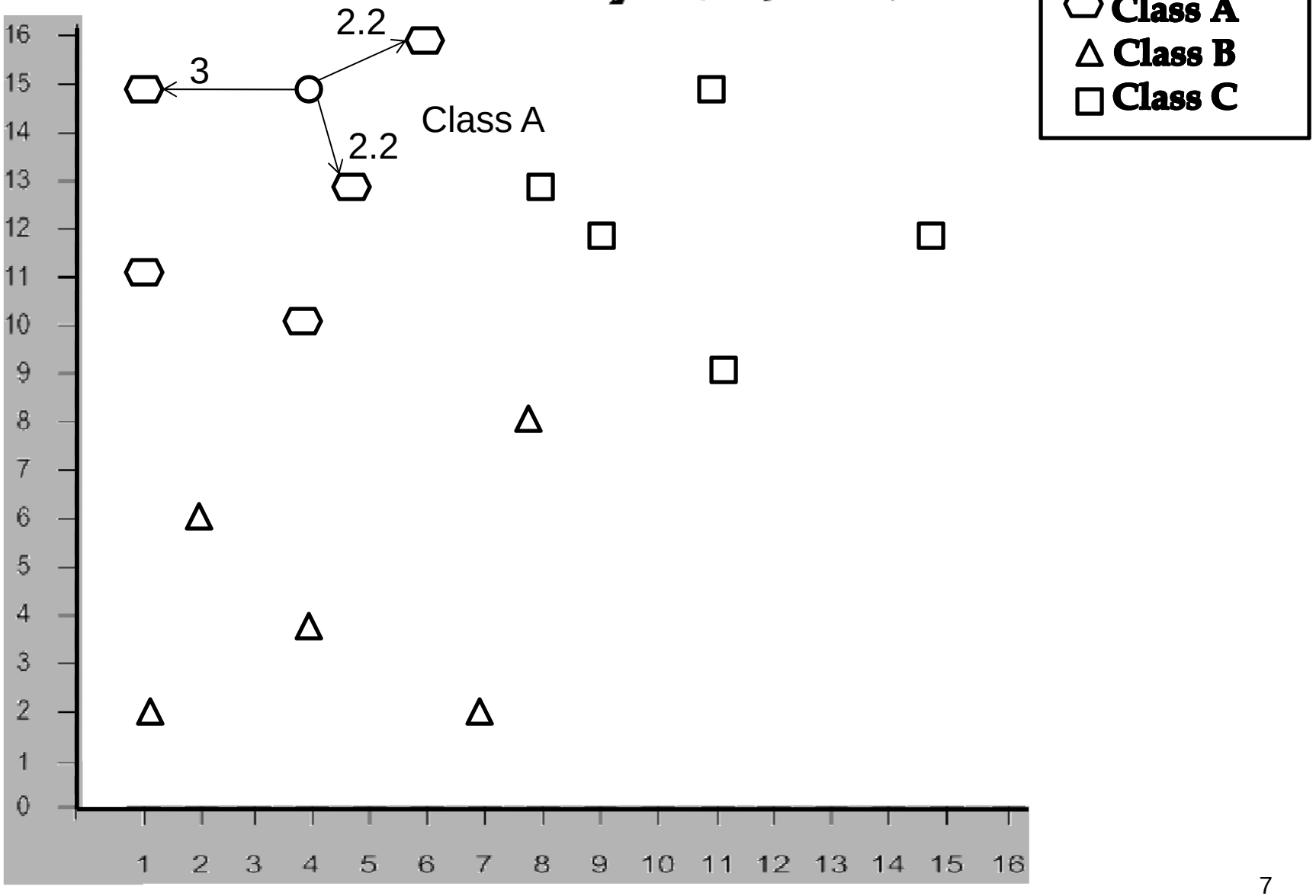


Data for Classification

- We need to classify the following five objects {(4,15), (1,9), (11,12), (6,9), (2,4)}

- K = 3

Classify (4,15)

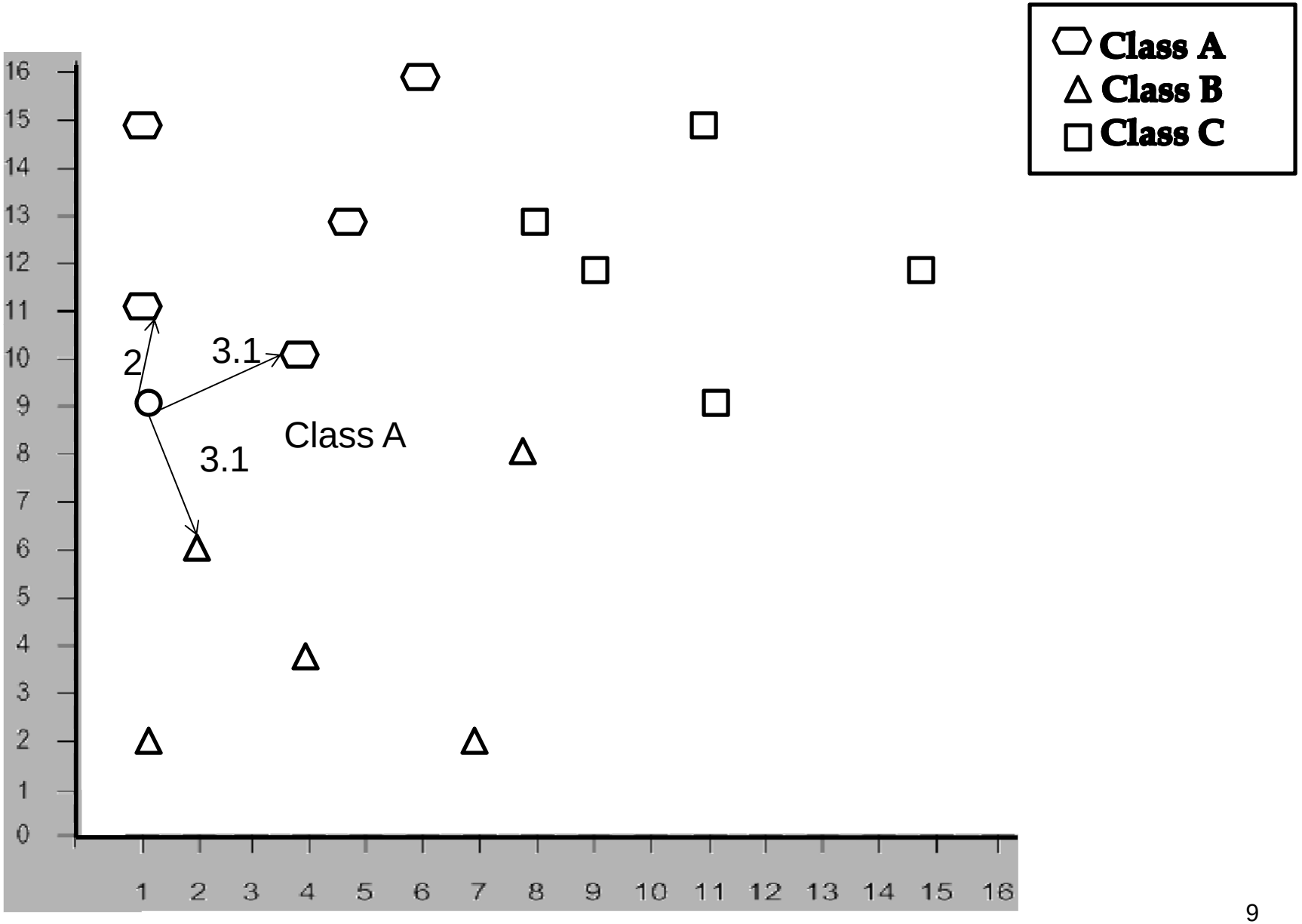


Classify point (4,15) cont.

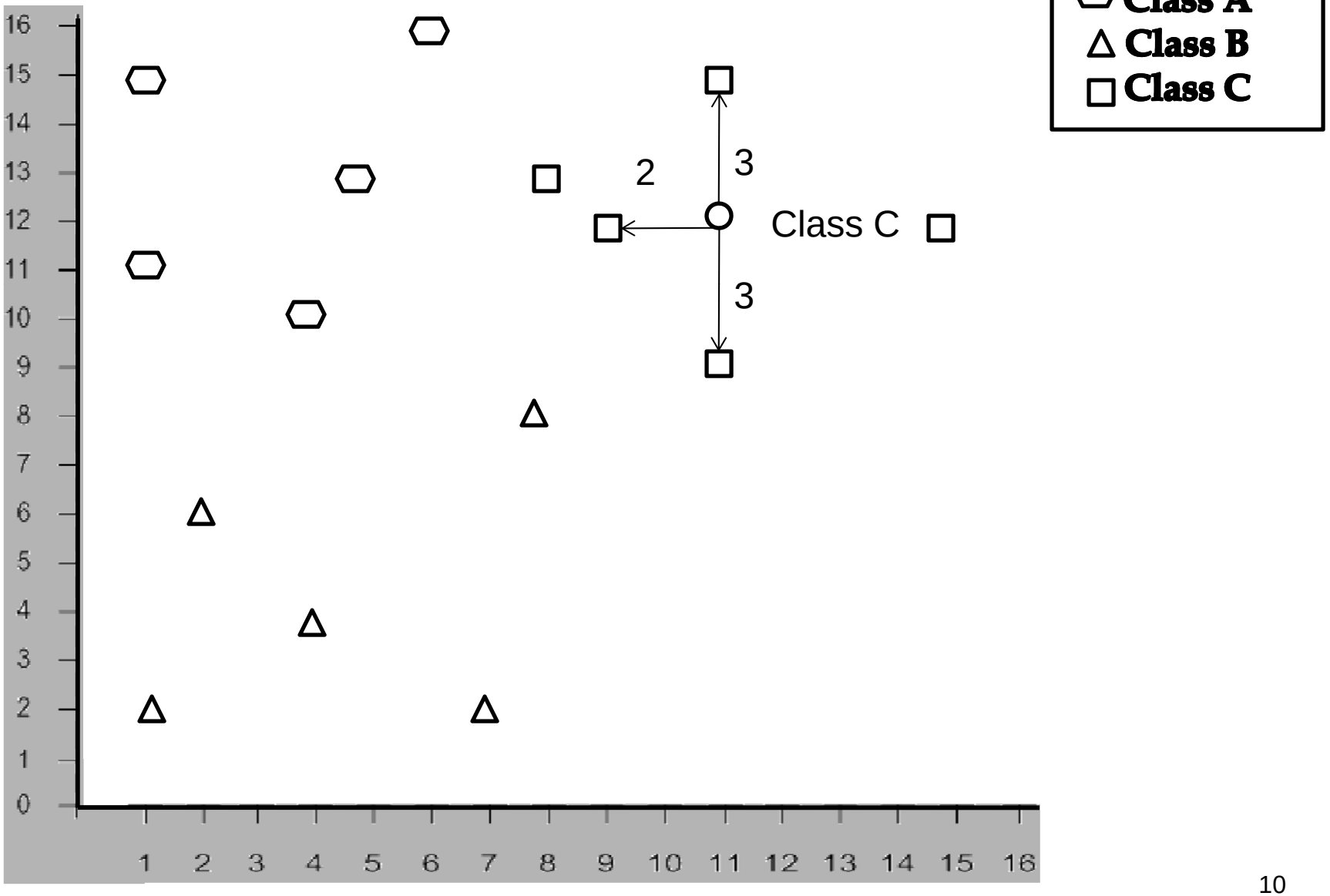
- Find 3 nearest neighbors (K = 3) by calculating Euclidean distance between point (4, 15) and each of the training samples.

- Since all 3 nearest neighbors are of class A, point (4, 15) is assigned to class A.

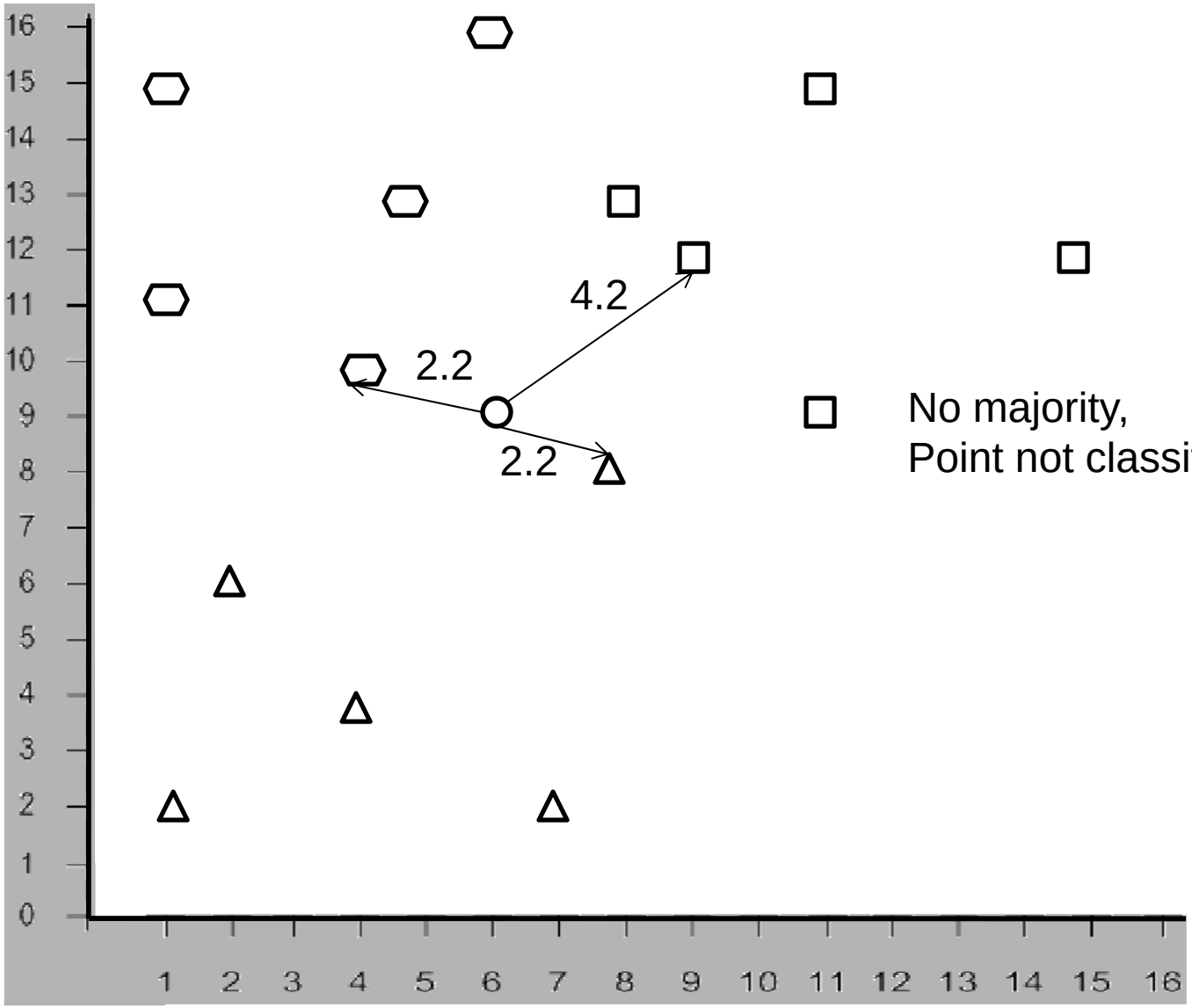
Classify (1,9)


Class A
Class B
Class C
2
3.1
3.1
Class A
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

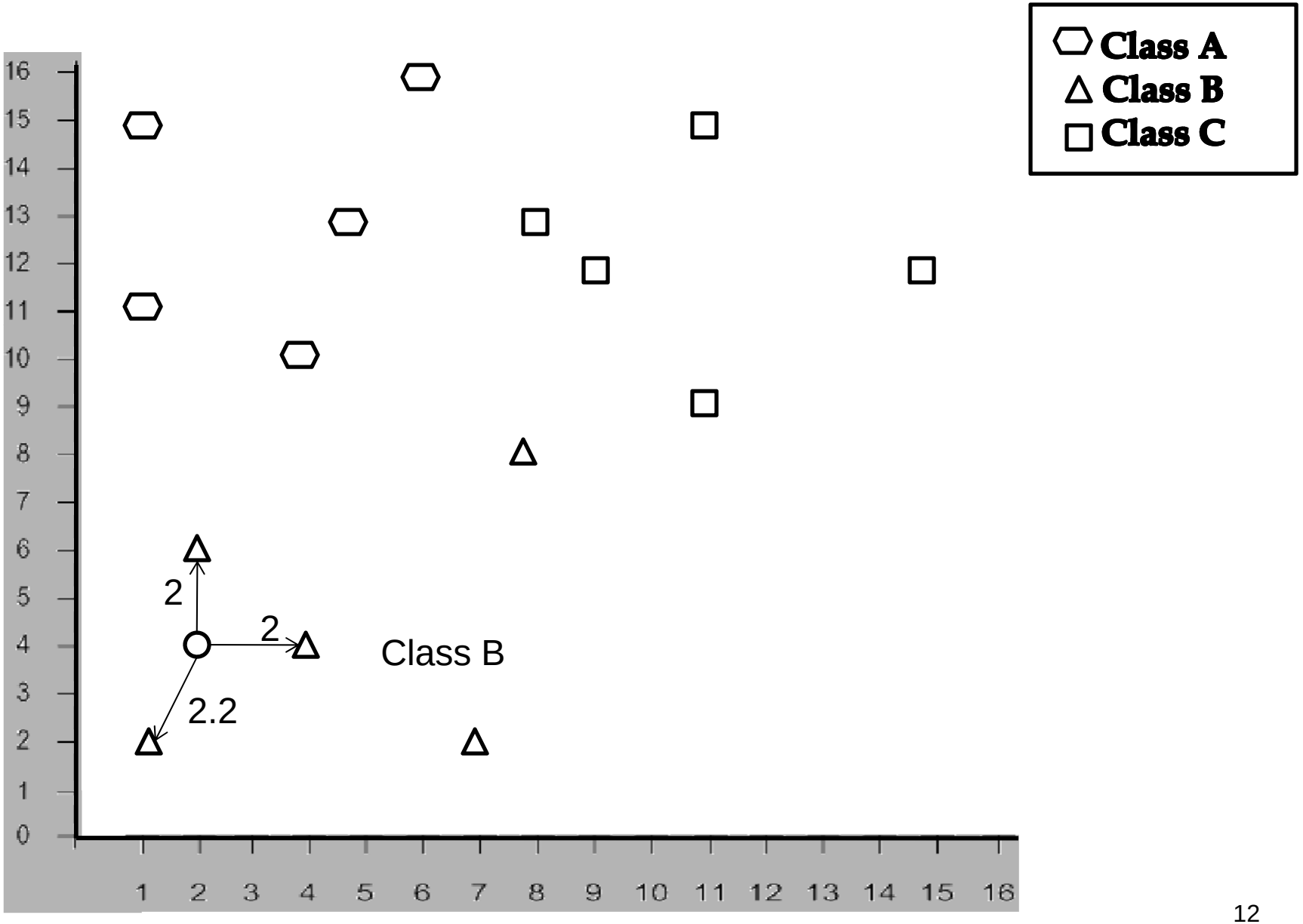
Classify (11,12)



Classify (6,9)


Class A
Class B
Class C
4.2
2.2
2.2
No majority,
Point not classified

Classify (2,4)



End of
Other Classification Methods Module (Part B)