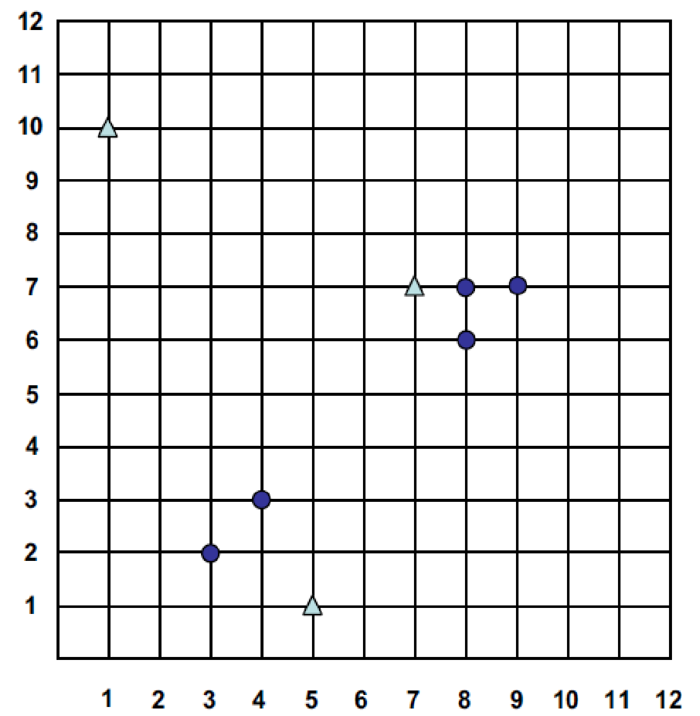
MMMO - Tutorial

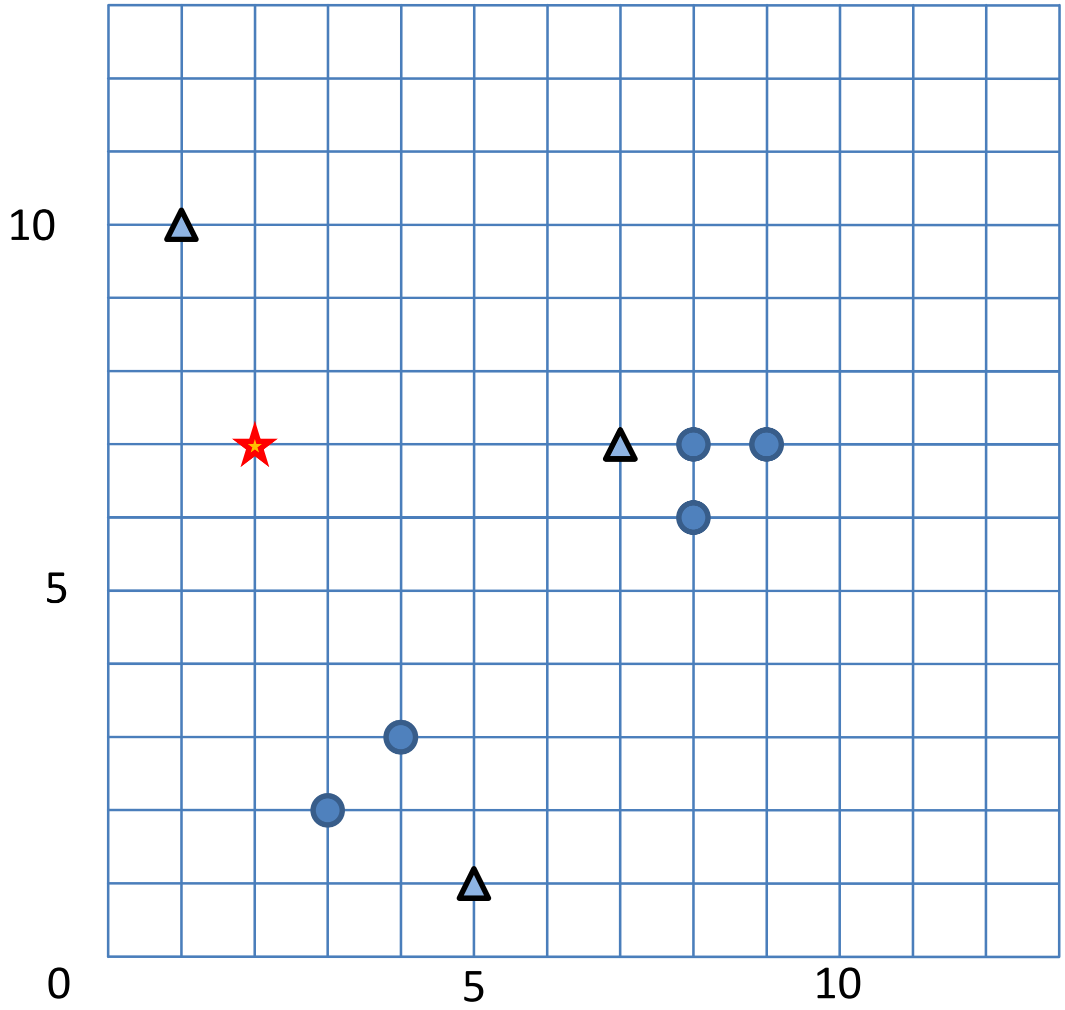
The following data set with 8 points (e.g. two-dimensionally feature vectors) is given. The triangles build one class and the circles build the other.

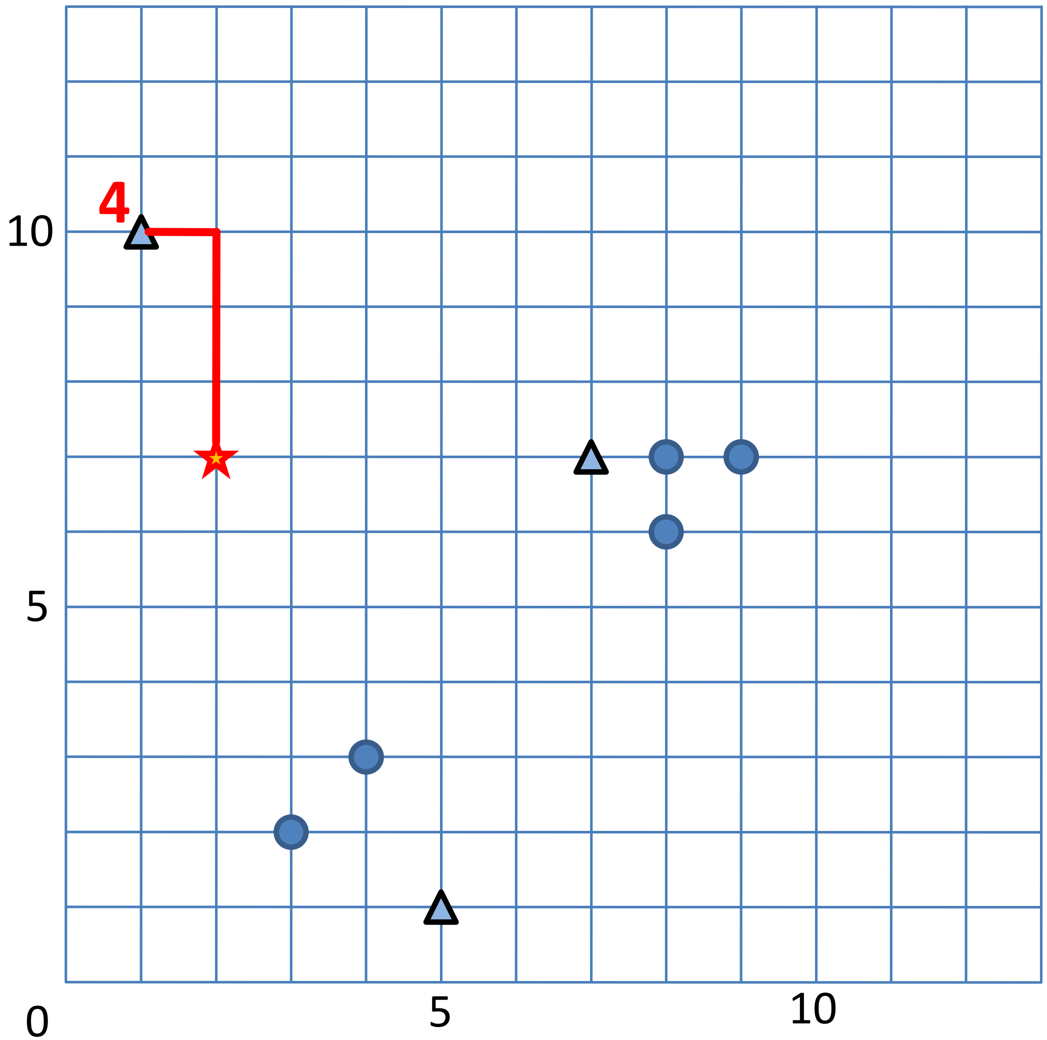


In the following the classes of data points should be determined with the k-nearest neighbors algorithm. As distance function between two points the Manhattan distance (l_1 norm) shall be used:

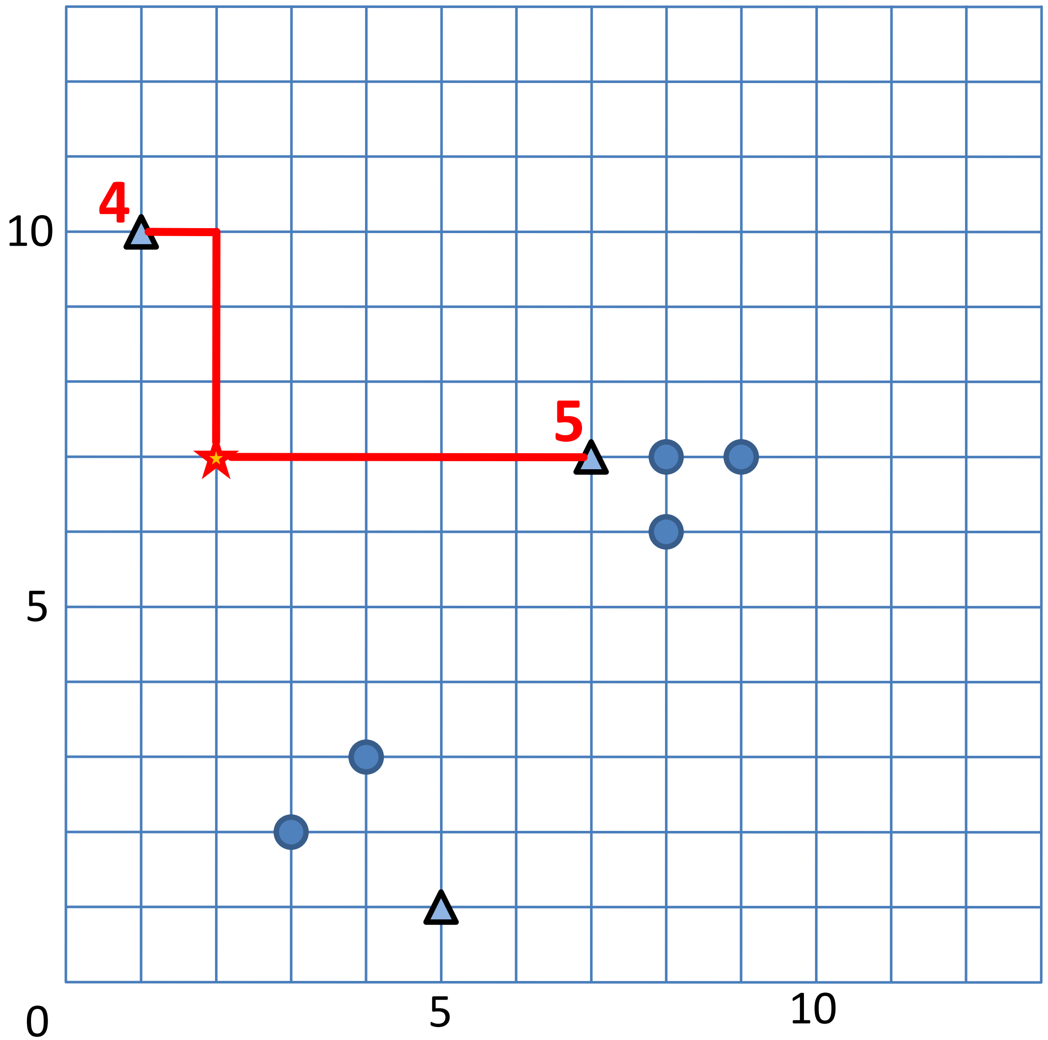
$$L_1(x, y) = \sum_{i=1}^{d} |x_i - y_i|$$

Determine the class of point (2,7) for k = 2 using the class of majority of its k-nearest neighbors, i.e. the point is assigned to the class which occurs most often among its k-nearest neighbors.

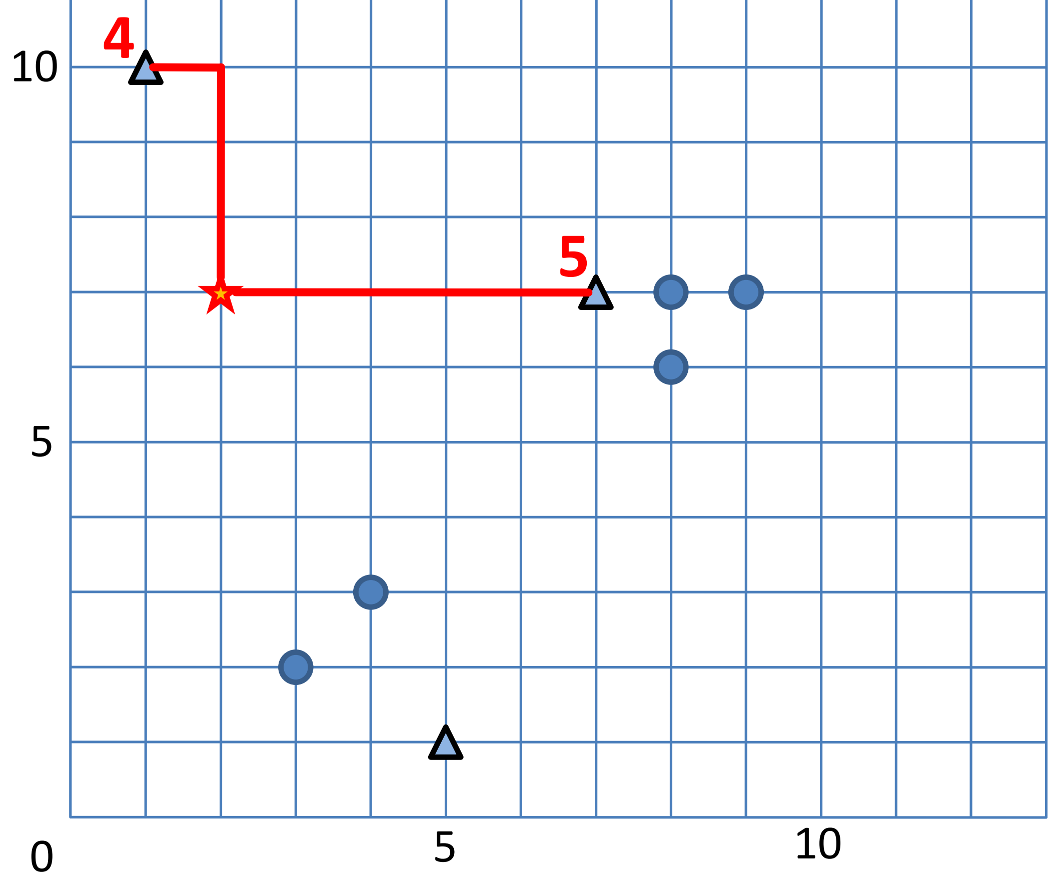




4
10
5
0
5
10

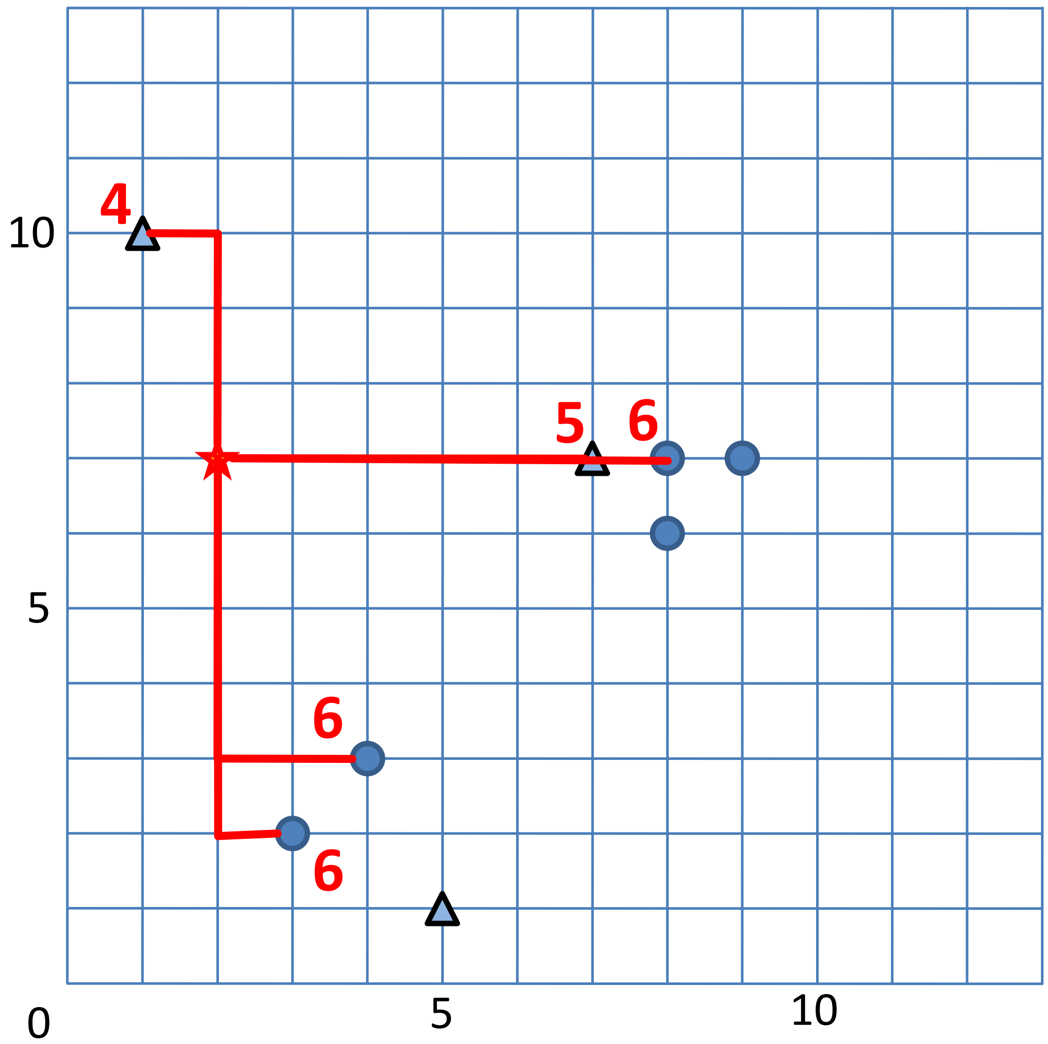

4
5
10
5
0
5
10

Thus, for k=2 ★ is classified as △



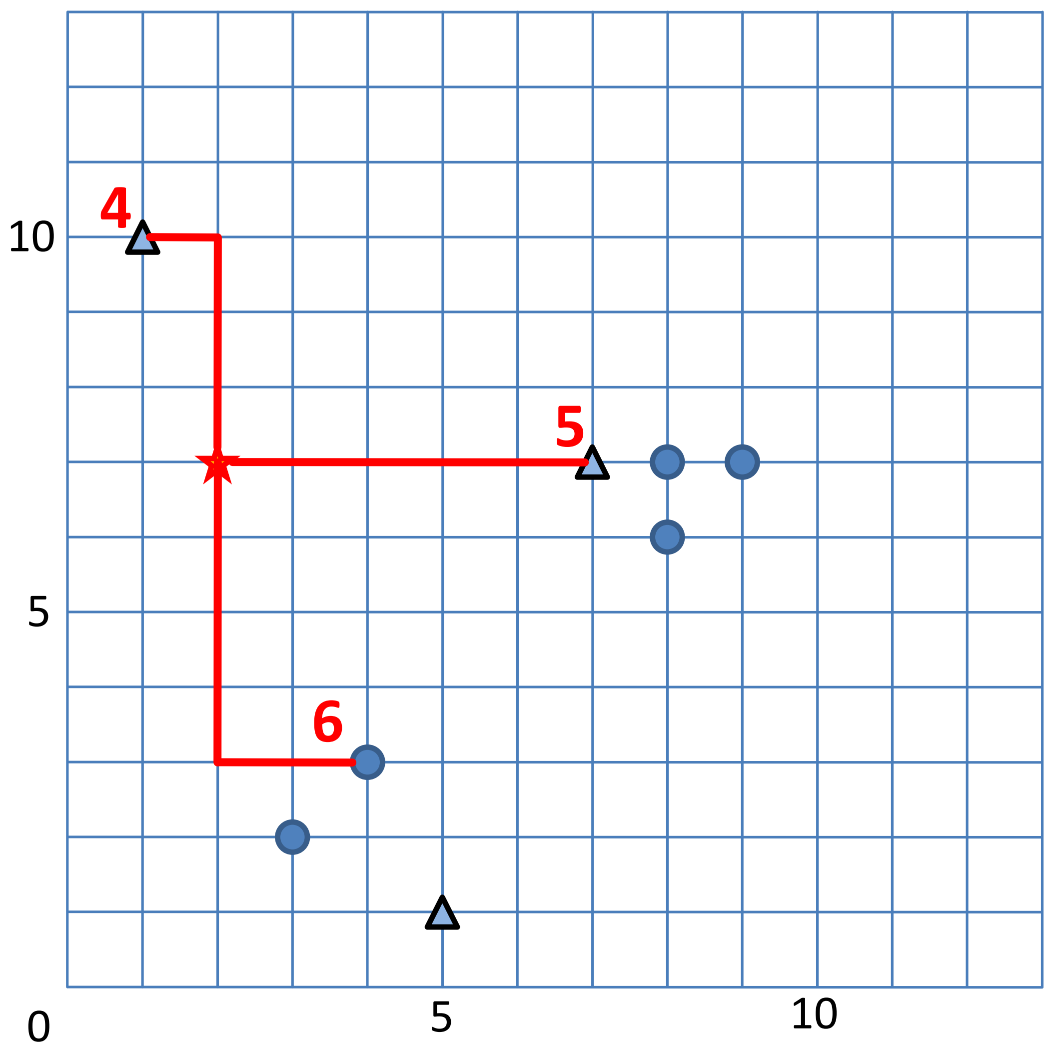
Determine the class of point (2,7) for k=3 using the class of majority of its k-nearest neighbors.

=> Special case: several points have the distance 6 to the query point



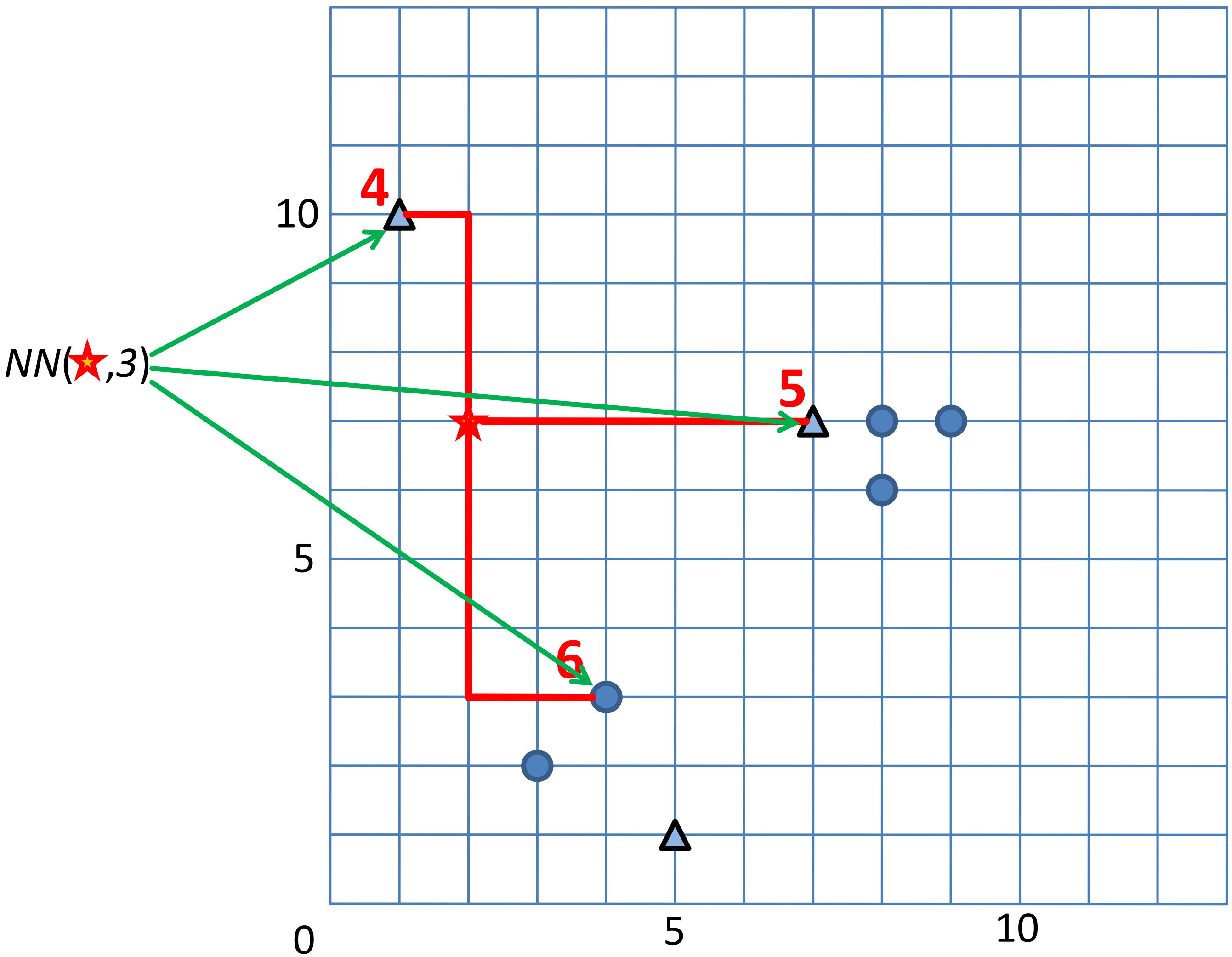
Alternative 1: Nondeterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with <u>exactly</u> k objects such that:

$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) \leq dist(q,o')$$



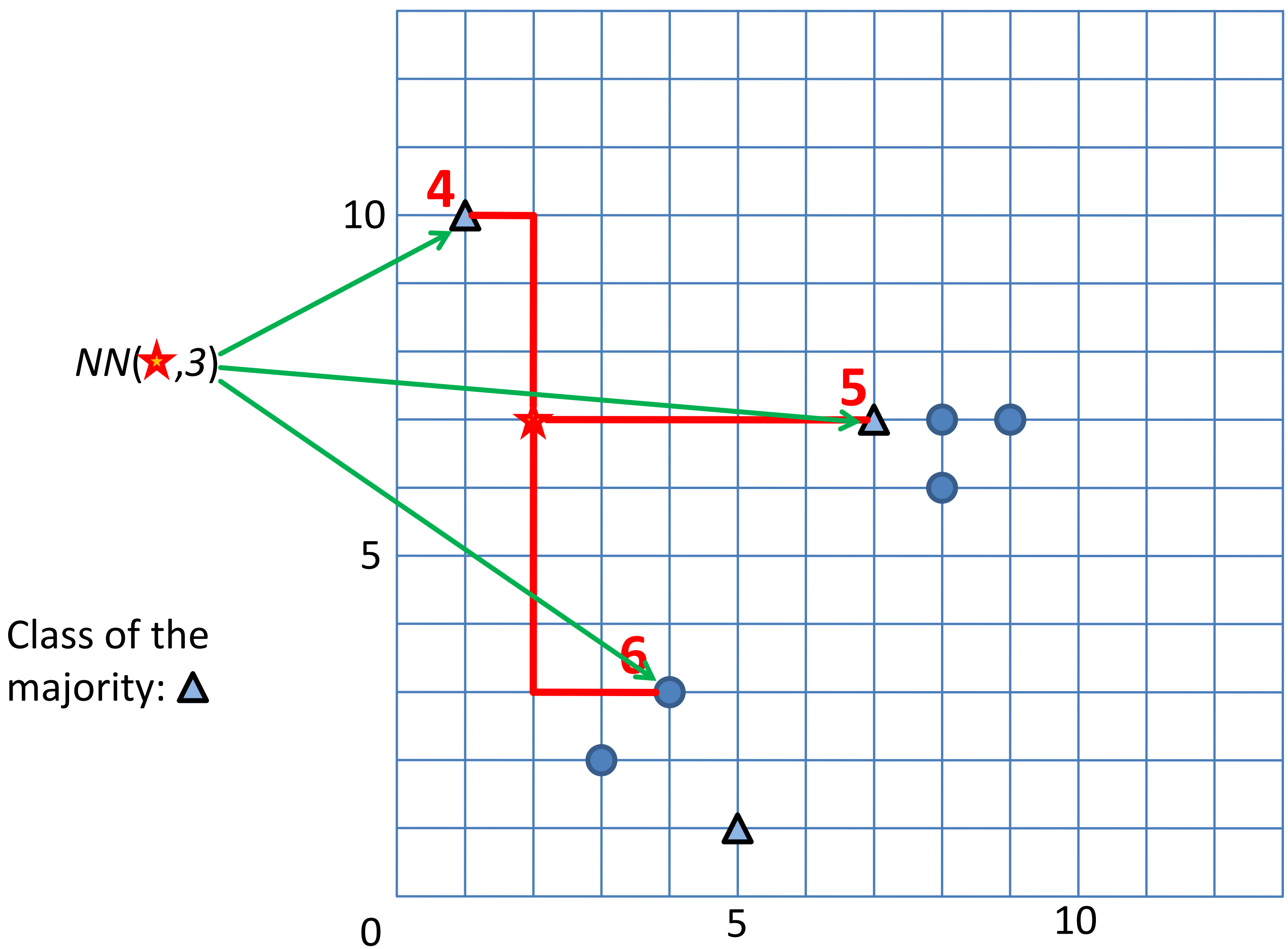
Nondeterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with <u>exactly</u> k objects such that:

$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) \leq dist(q,o')$$



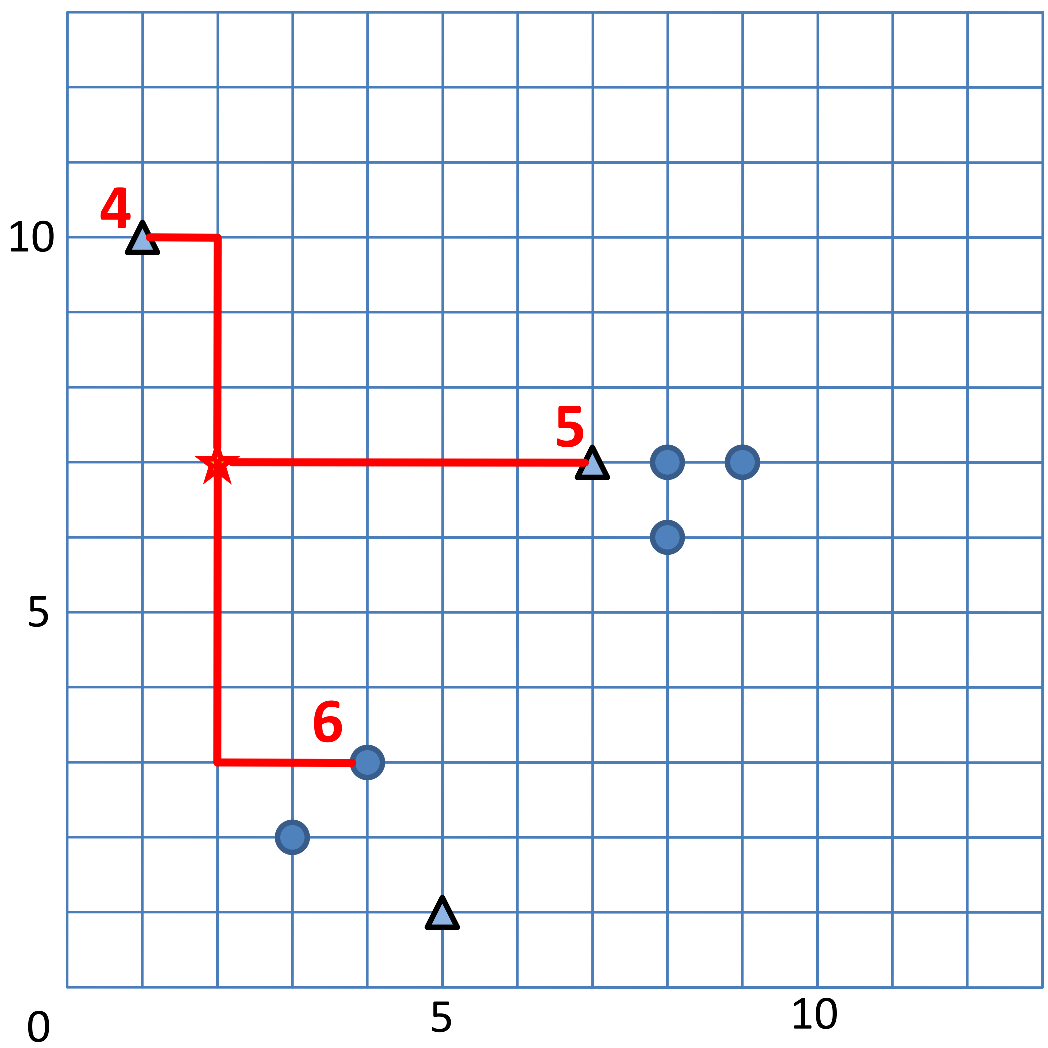
Nondeterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with <u>exactly</u> k objects such that:

$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) \leq dist(q,o')$$



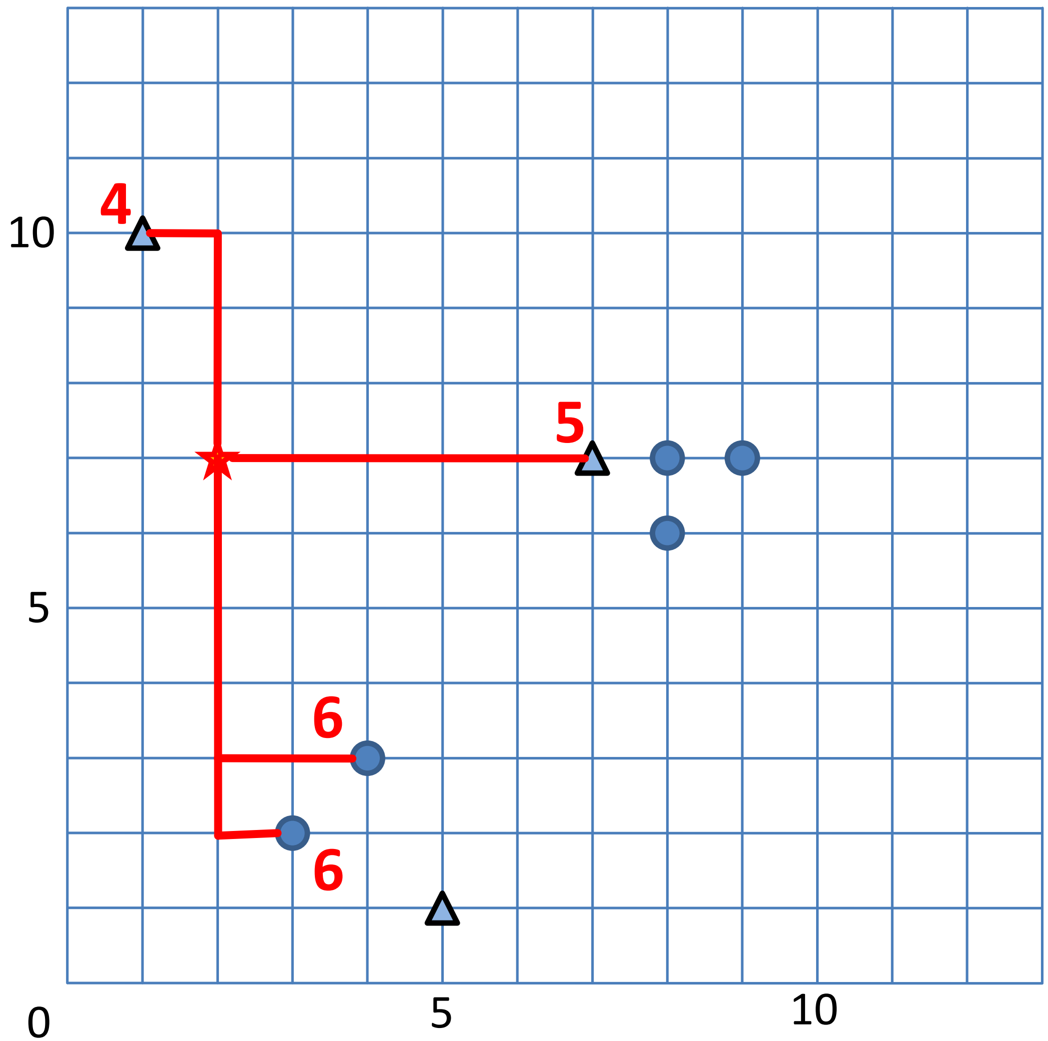
Deterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with at least k objects such that:

$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) < dist(q,o')$$



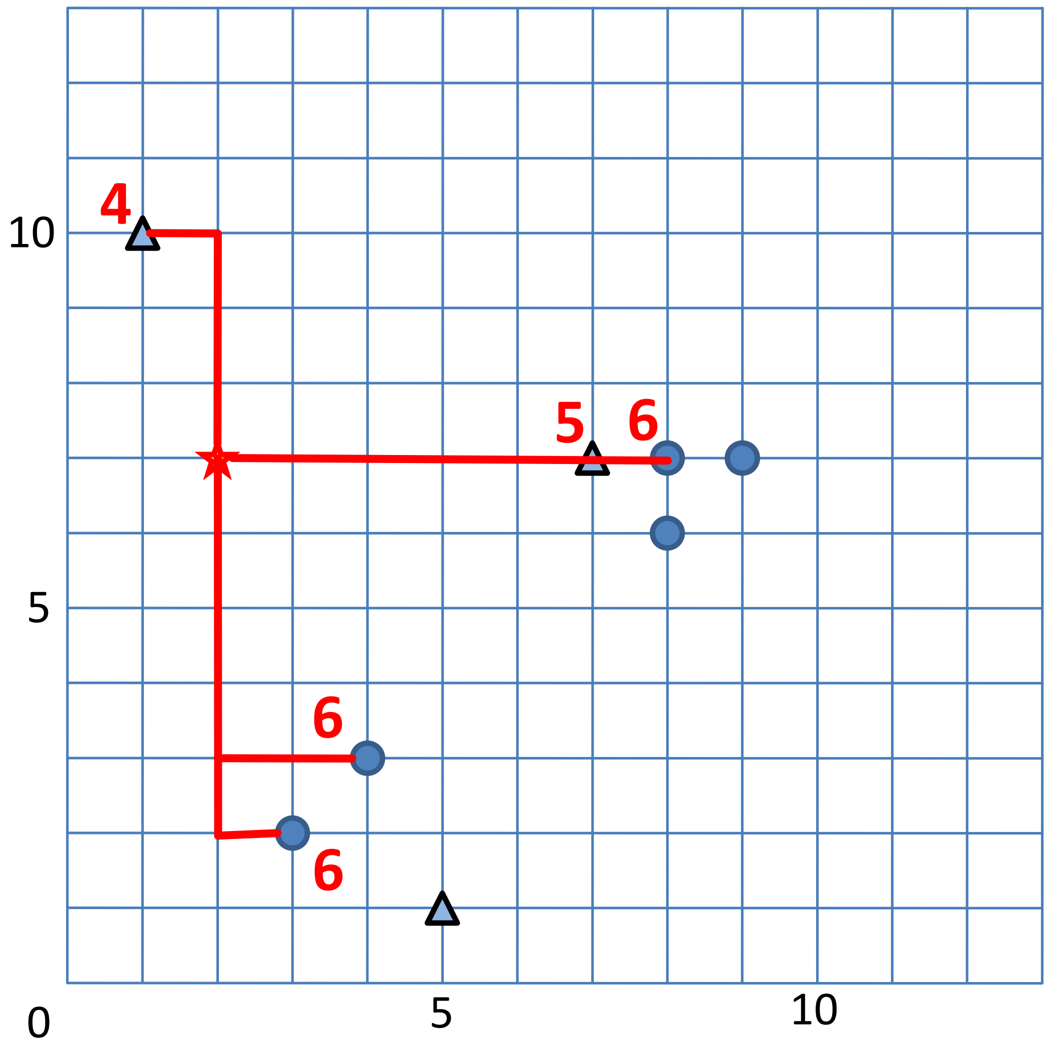
Deterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with <u>at least</u> k objects such that:

$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) < dist(q,o')$$



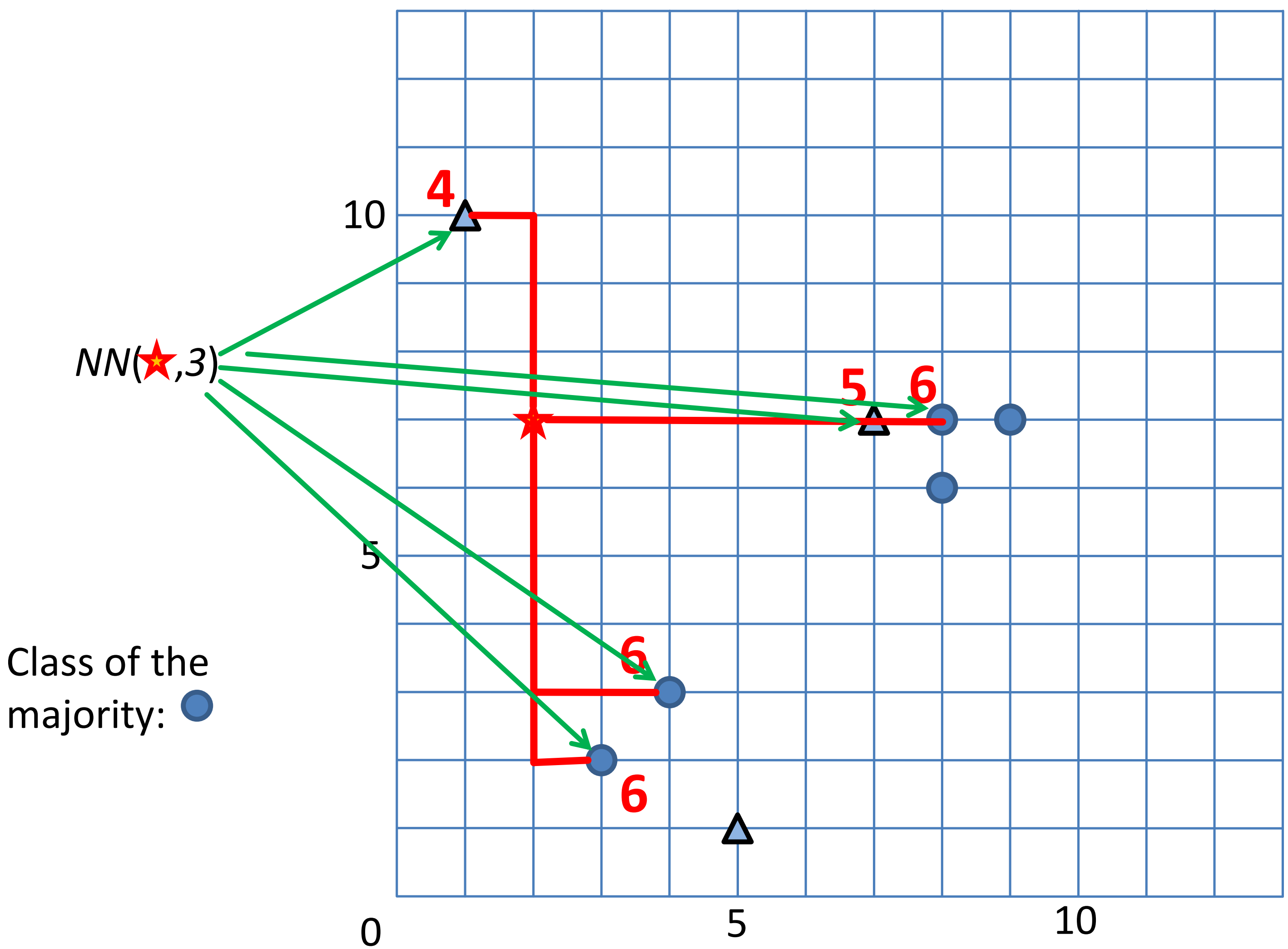
Deterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with at least k objects such that:

$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) < dist(q,o')$$



Deterministic definition of kNN: Set $NN(q,k) \subseteq$ DB with <u>at least</u> k objects such that:

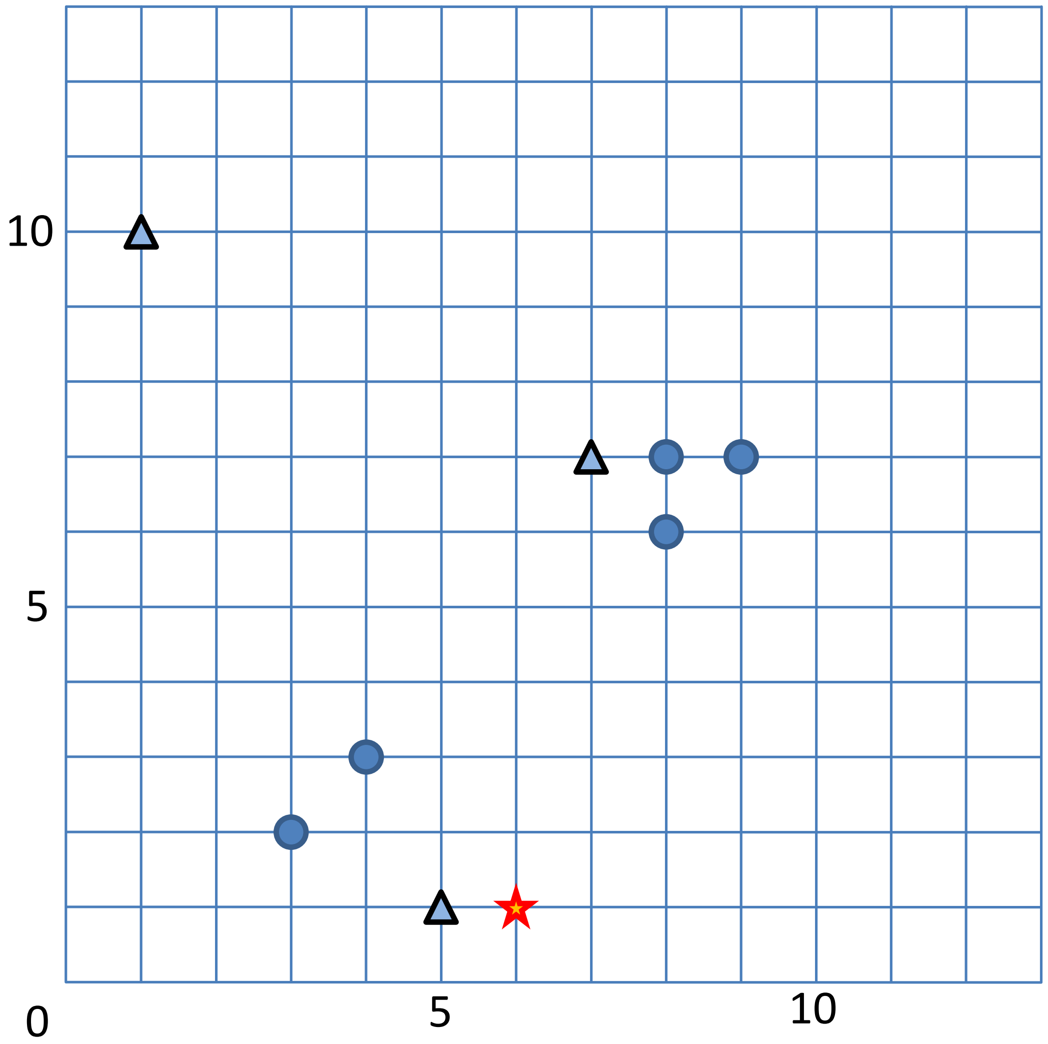
$$\forall o \in NN(q,k), \forall o' \in DB - NN(q,k) : dist(q,o) < dist(q,o')$$

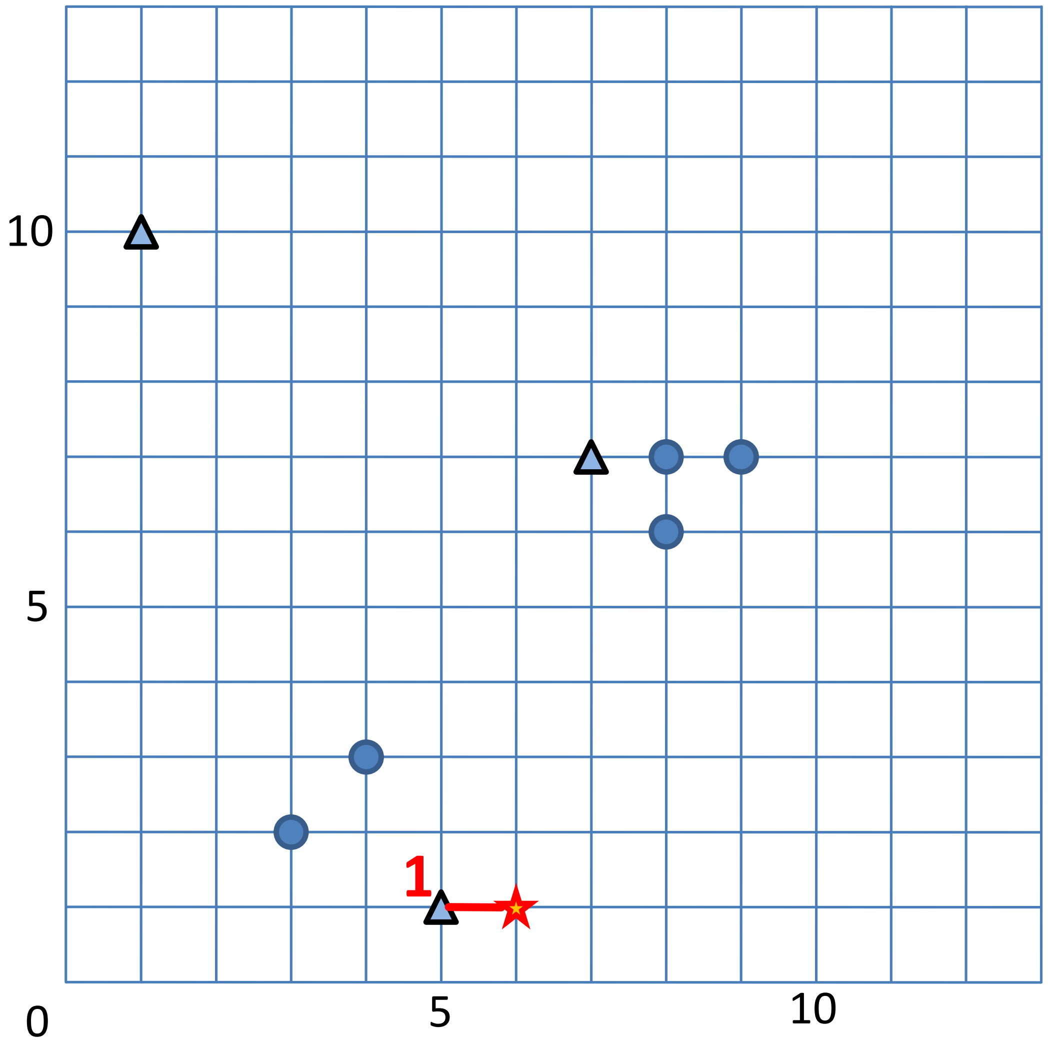


Determine the class of point (2,7) for k=5 using the class of majority of its k-nearest neighbors.

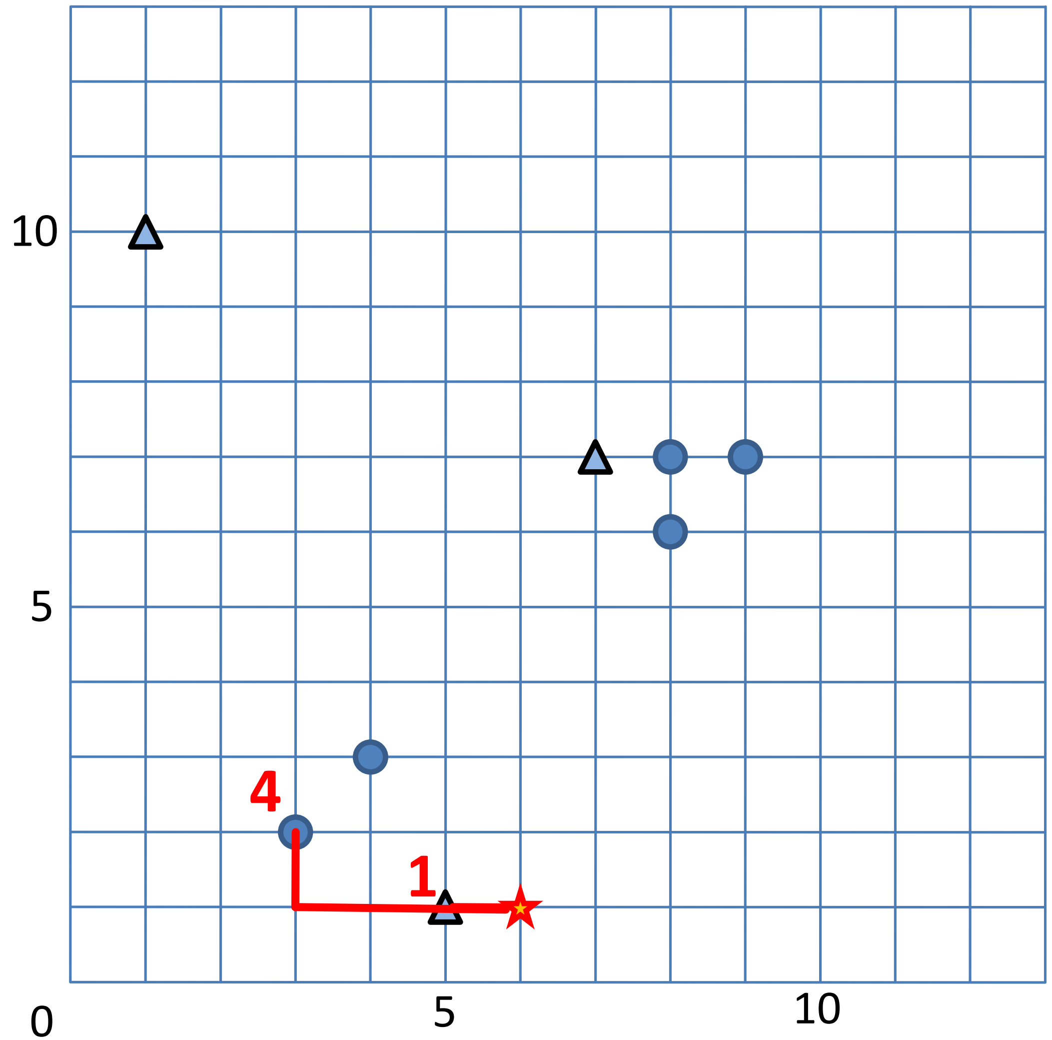
Analogously to the deterministic alternative of the exercise before

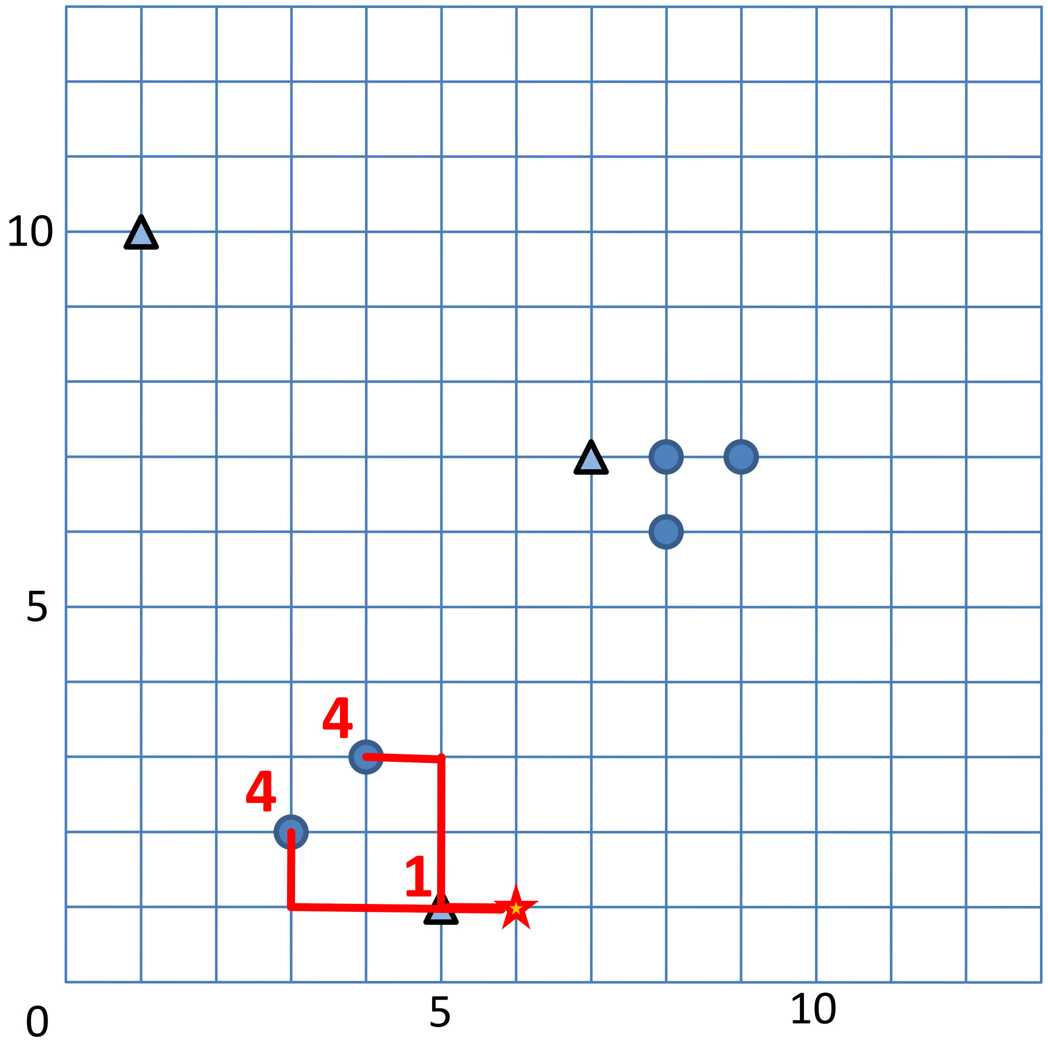
Determine the class of point (6,1) for k=3 using the class of majority of its k-nearest neighbors.

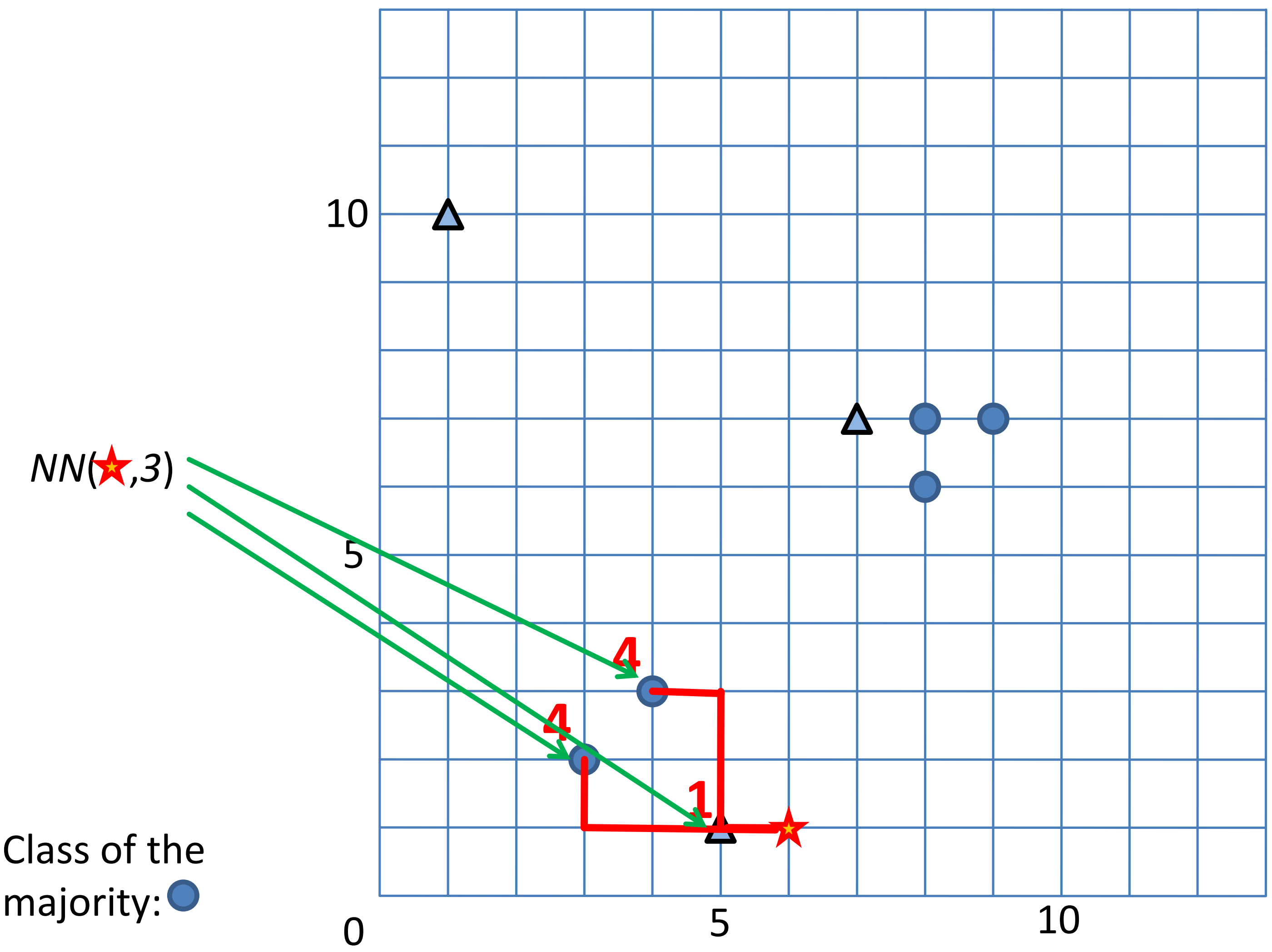




10
5
0
5
10
1

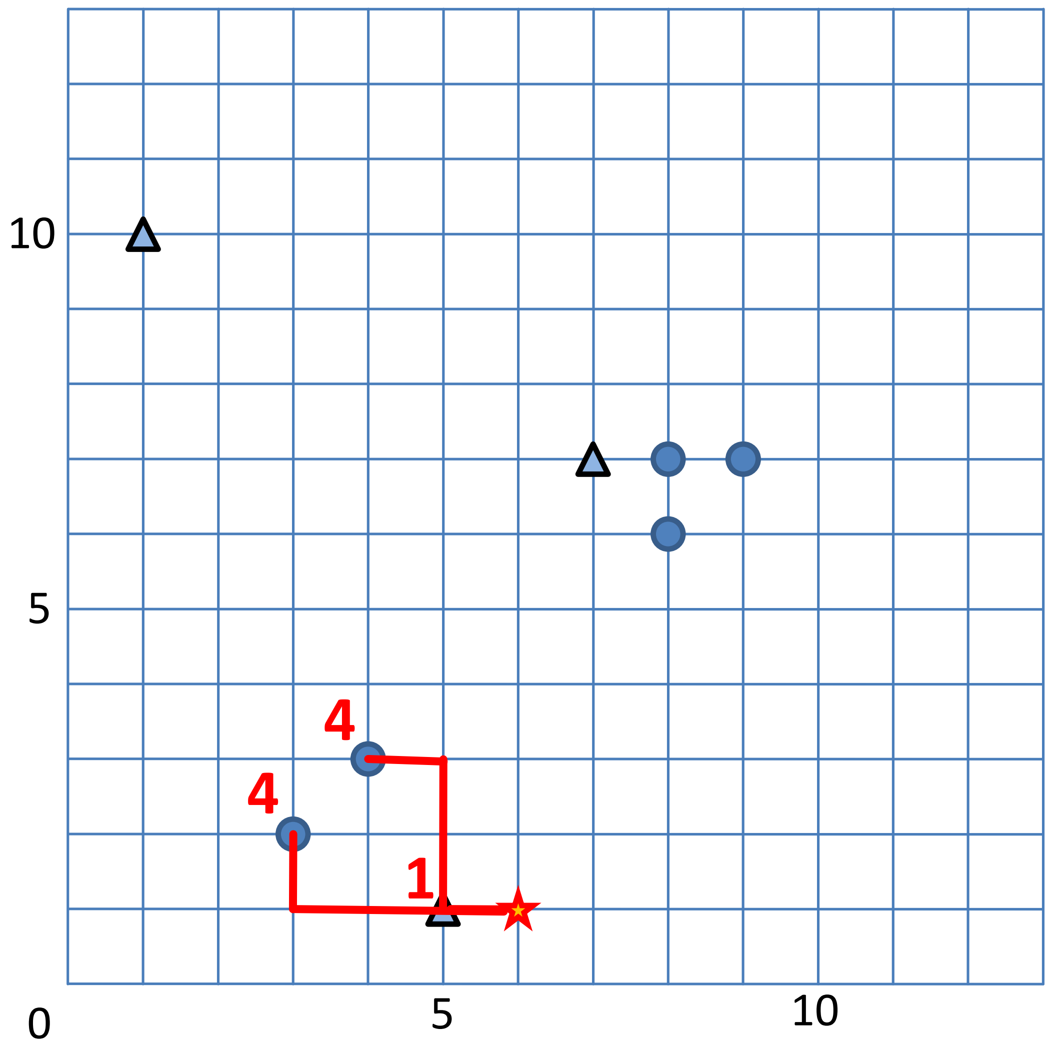

10
5
0
5
10
4
1

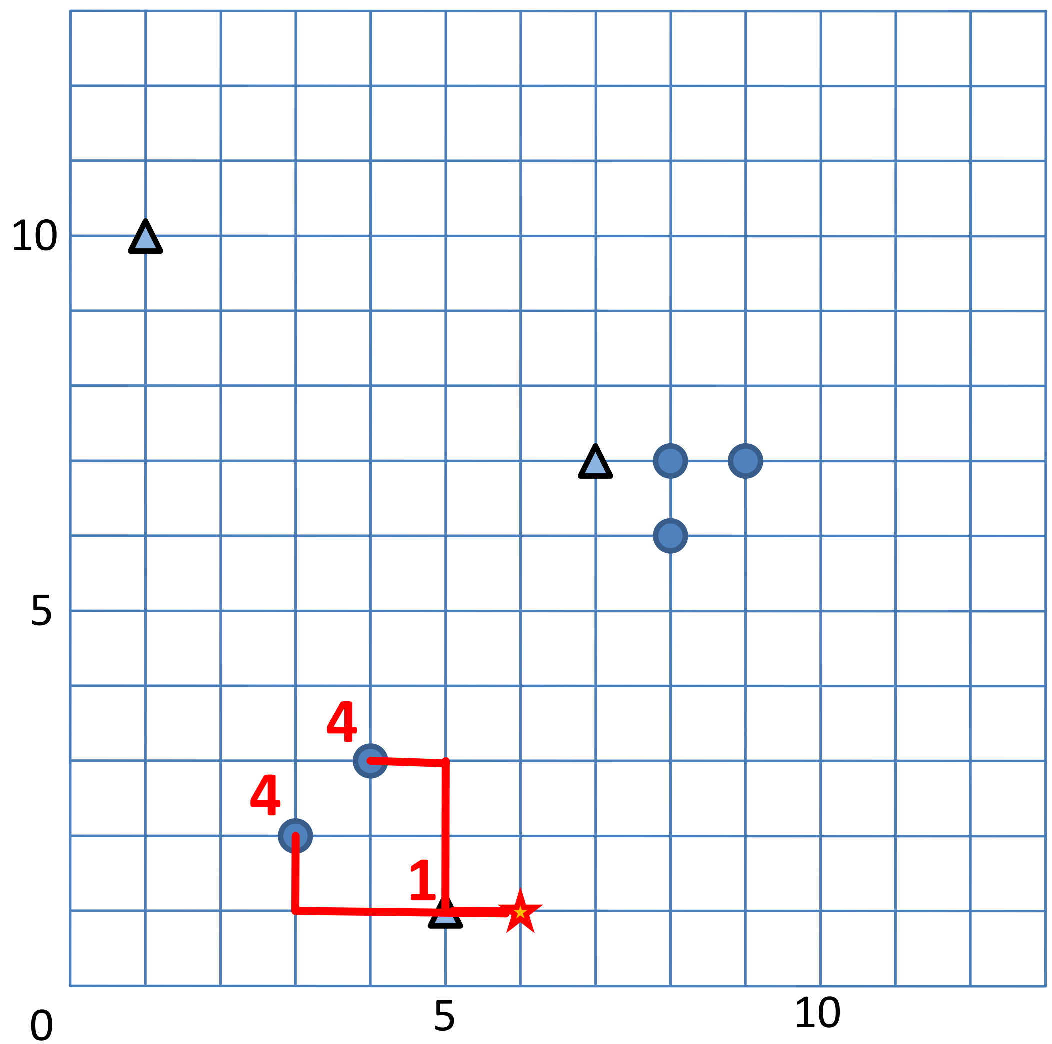

10
5
0
5
10
4
4
1



NN(,3)
10
5
0
5
10
4
4
1
Class of the
majority:

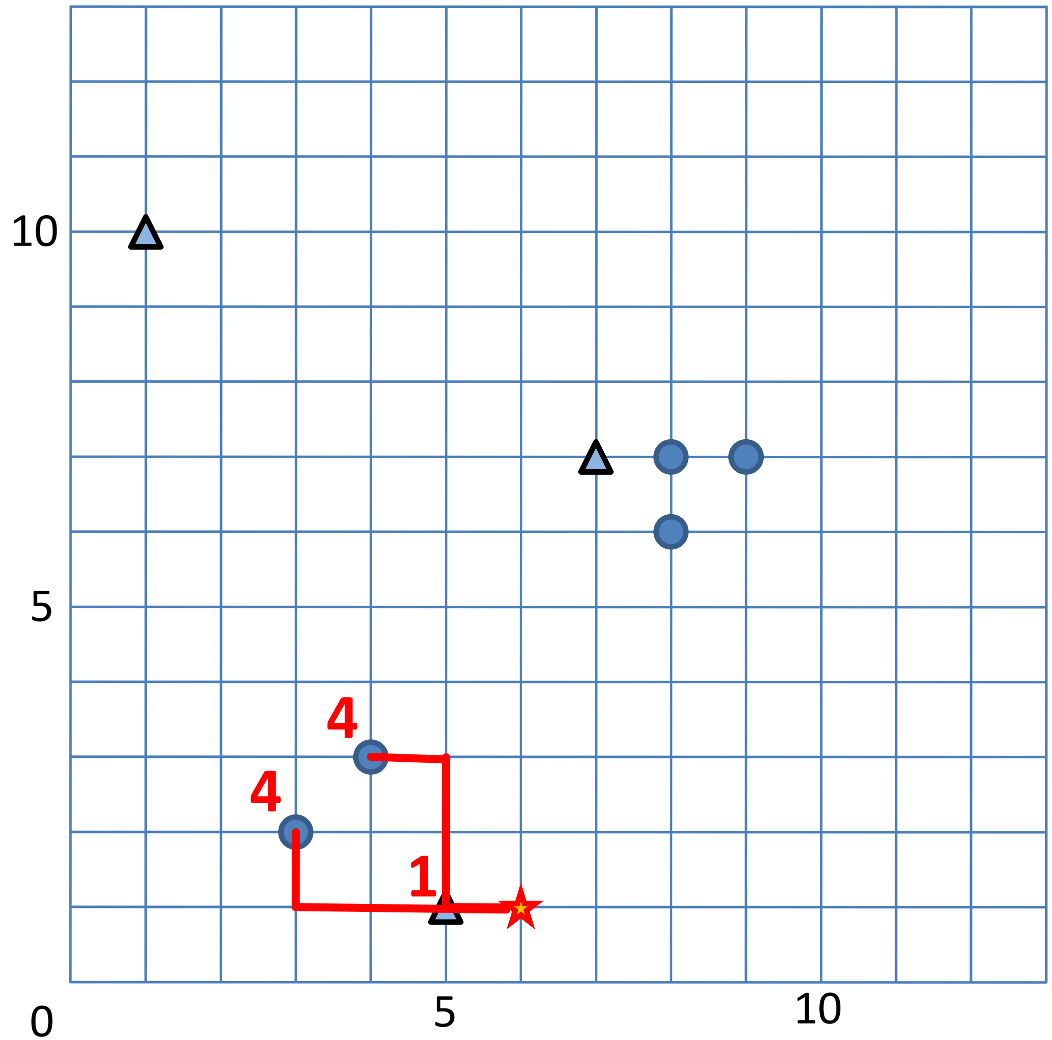
Determine the class of point (6,1) for k=3 using the class of majority of its k-nearest neighbors weighting the classes with inverse Manhattan distance.





Weighting(●) = ¼ + ¼ = ½

Weighting(▲) = 1/1 = 1



Weighting(●) = ¼ + ¼ = ½

Weighting(▲) = 1/1 = 1

Highest weight: ▲