Classification with *k* nearest neighbours

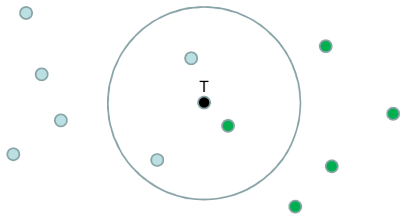
Marzena Kryszkiewicz

Classification with k nearest neighbours (kNN)

- Let T be a an object to be classified. Usually, its classification with kNN consists in:
  1. Finding *k nearest neighbours* of T (that is, *k* objects that are *least distant* from T or *most similar* to T) in decision table,
  2. Assigning T to the decision class containing the majority of the found k nearest neighbours.

2

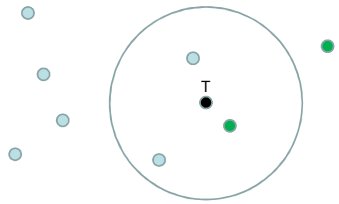
Example: Classification with kNN



Let k = 3. Then (usually) object T will be classified to the decision with **blue** objects.

3

Example: Classification with kNN when decision classes are imbalanced



Let k = 3. Taking into account that:
- 1/2 = 50% of green objects is among kNN,
- 2/6 ≈ 33% of blue objects is among kNN.

object T might be assigned to the decision class with **green** objects.

4

References

- Jiawei Han, Micheline Kamber, Jian Pei: Data Mining: Concept and Techniques, The Morgan Kaufmann Series in Data Management Systems, 2011

5