Nearest neighbor Classification…

- k-NN classifiers are lazy learners since they do not build models explicitly

- Classifying unknown records are relatively expensive

- Can produce arbitrarily shaped decision boundaries

- Easy to handle variable interactions since the decisions are based on local information

- Selection of right proximity measure is essential

- Superfluous or redundant attributes can create problems

- Missing attributes are hard to handle

Improving KNN Efficiency

- Avoid having to compute distance to all objects in the training set
    - Multi-dimensional access methods (k-d trees)
    - Fast approximate similarity search
    - Locality Sensitive Hashing (LSH)
- Condensing
    - Determine a smaller set of objects that give the same performance
- Editing
    - Remove objects to improve efficiency

KNN and Proximity Graphs

- Proximity graphs
  - a graph in which two vertices are connected by an edge if and only if the vertices satisfy particular geometric requirements
  - nearest neighbor graphs,
  - minimum spanning trees
  - Delaunay triangulations
  - relative neighborhood graphs
  - Gabriel graphs
- See recent papers by Toussaint
  - G. T. Toussaint. Proximity graphs for nearest neighbor decision rules: recent progress. In Interface-2002, 34th Symposium on Computing and Statistics, ontreal, Canada, April 17–20 2002.
  - G. T. Toussaint. Open problems in geometric methods for instance based learning. In Discrete and Computational Geometry, volume 2866 of Lecture Notes in Computer Science, pages 273–283, December 6-9, 2003.
  - G. T. Toussaint. Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining. Int. J. Comput. Geometry Appl., 15(2):101–150, 2005.