# Machine Learning

## Lecture 14: Nearest Neighbor Classifiers

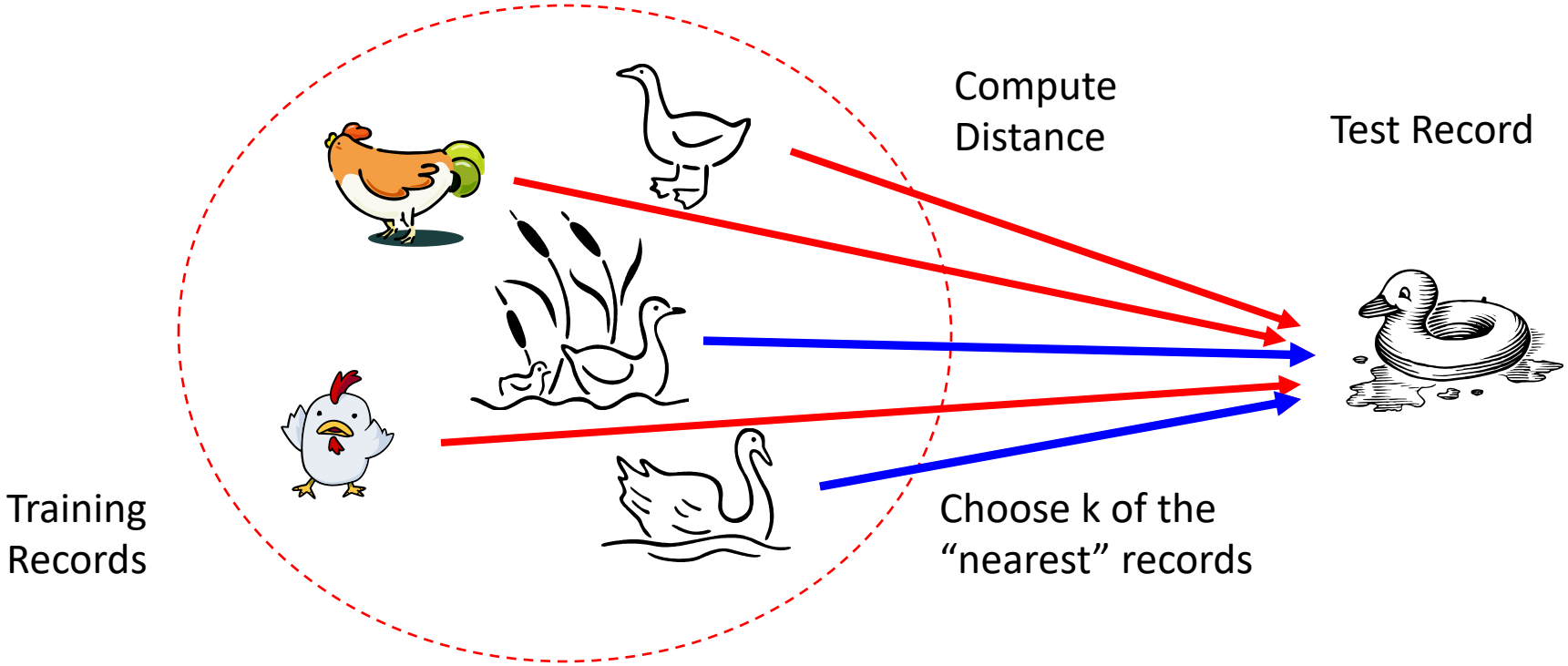
**杨思蓓**

**SIST**

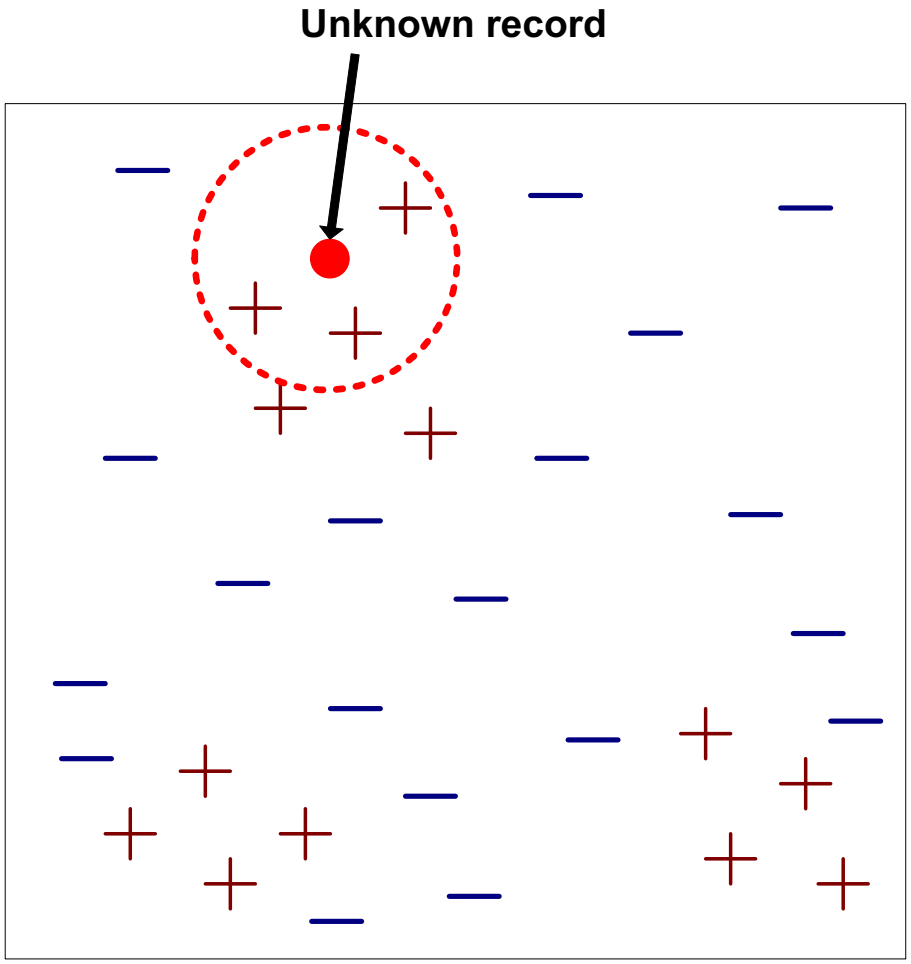
**Email: yangsb@shanghaitech.edu.cn**

# Nearest Neighbor Classifiers

- Basic idea:
  - If it walks like a duck, quacks like a duck, then it's probably a duck

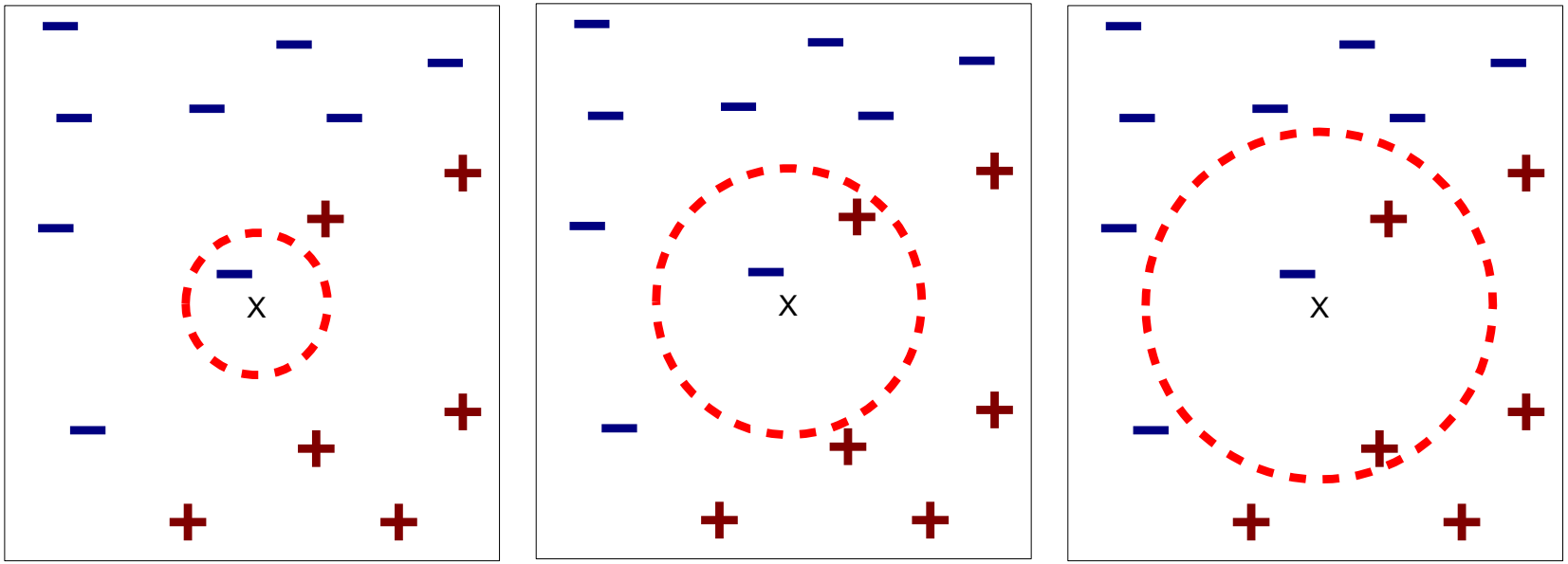


# Nearest-Neighbor Classifiers




- Requires three things
  - The set of stored records
  - Distance Metric to compute distance between records
  - The value of *k*, the number of nearest neighbors to retrieve
- To classify an unknown record:
  - Compute distance to other training records
  - Identify *k* nearest neighbors
  - Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

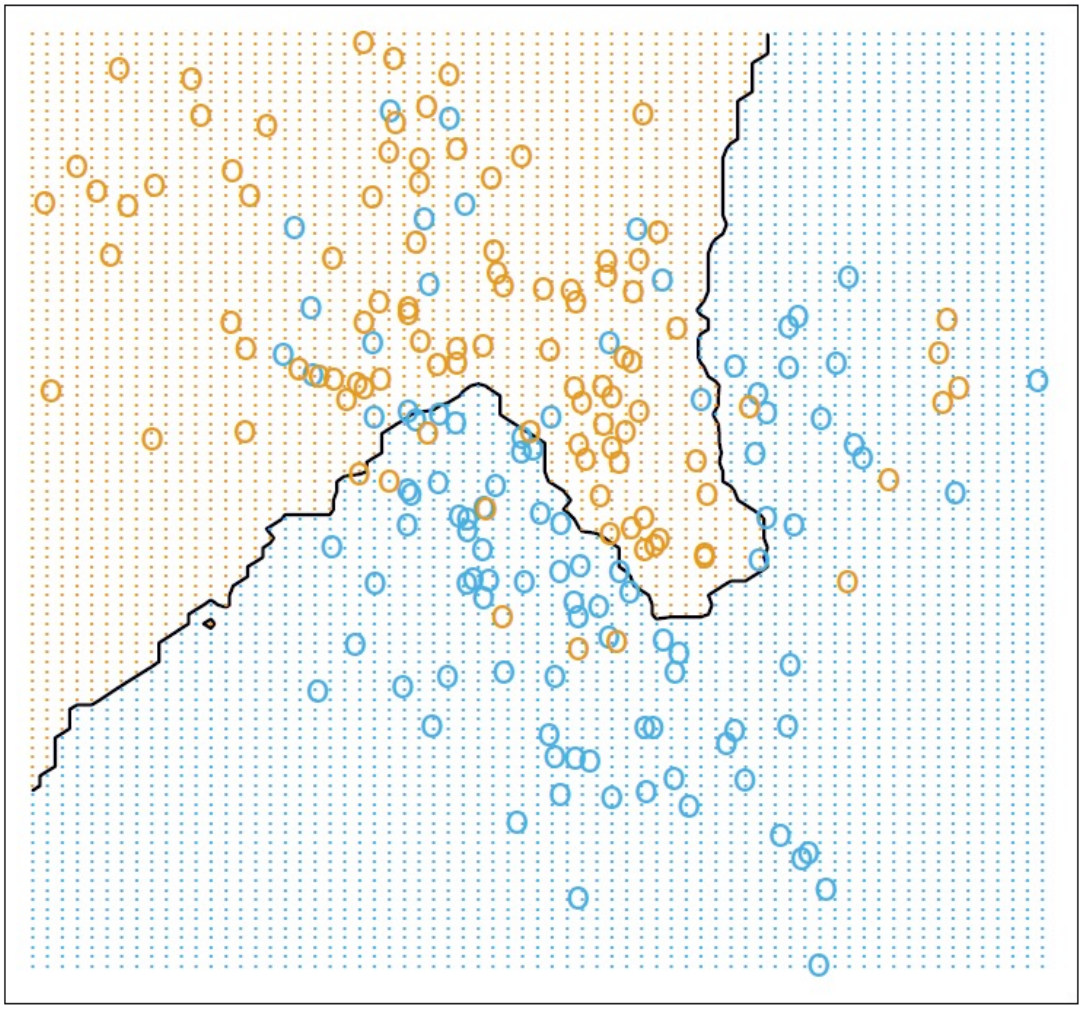
# Definition of Nearest Neighbor



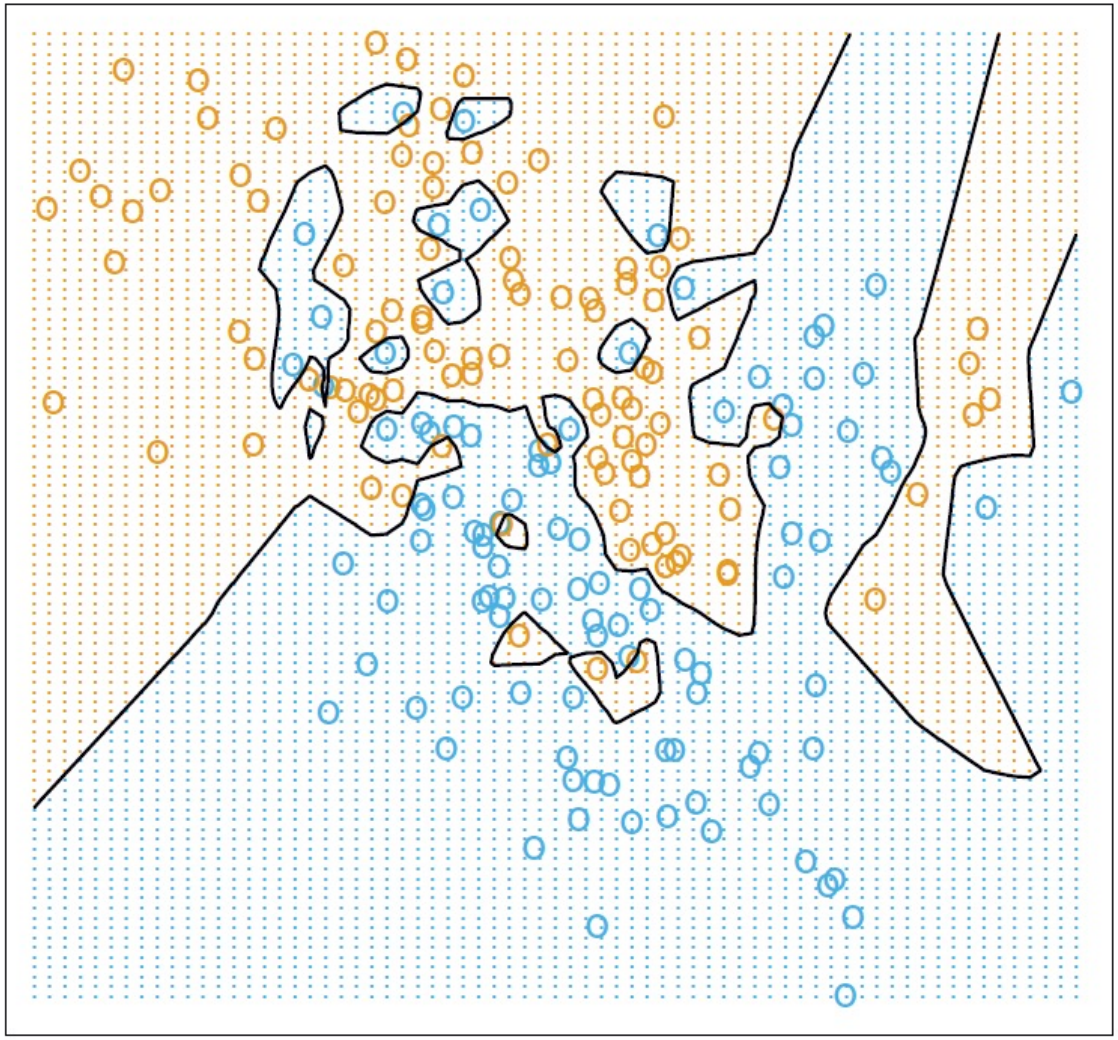

(a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

K-nearest neighbors of a record ***x*** are data points that have the *k* smallest distance to ***x***

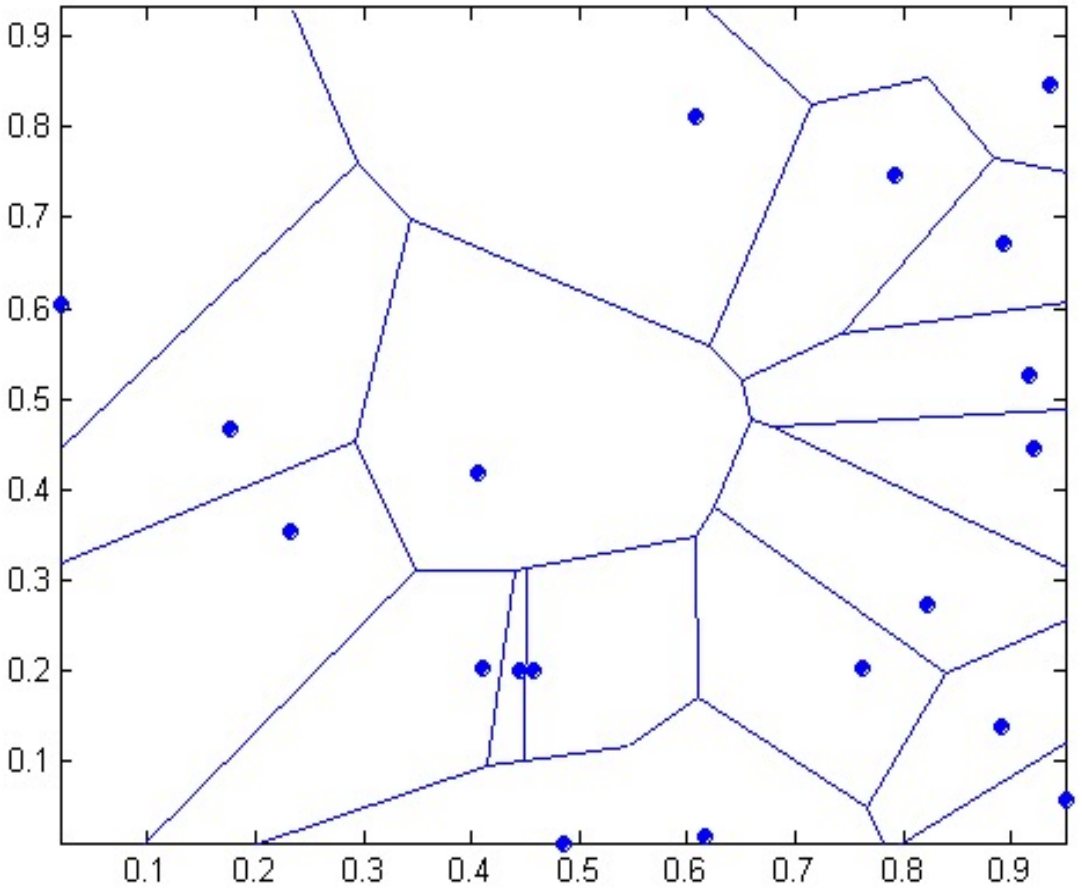
# 15-Nearest Neighbor Classifier



# 1-Nearest Neighbor Classifier

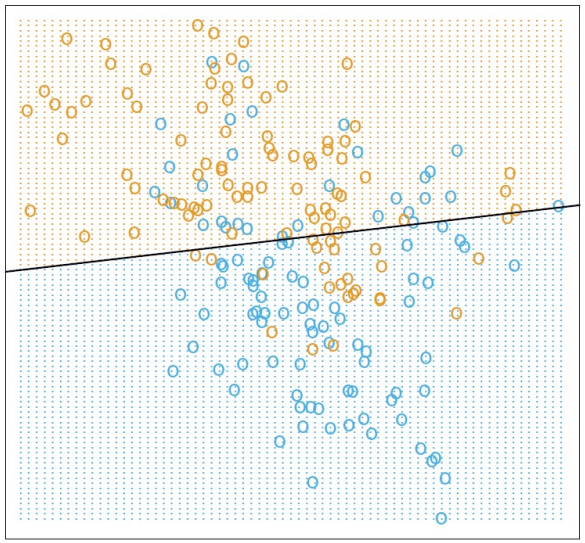
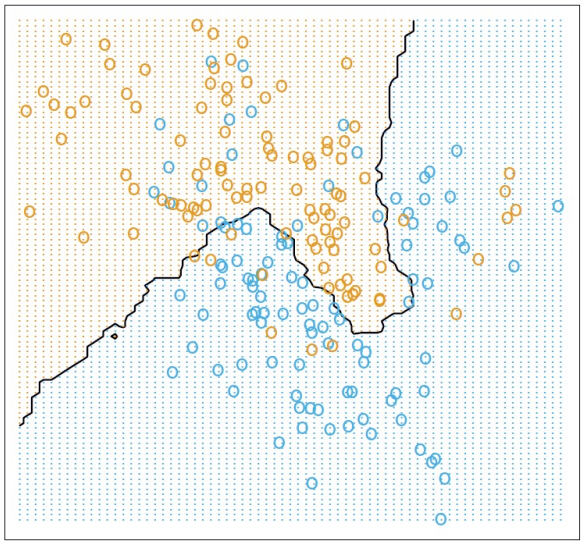
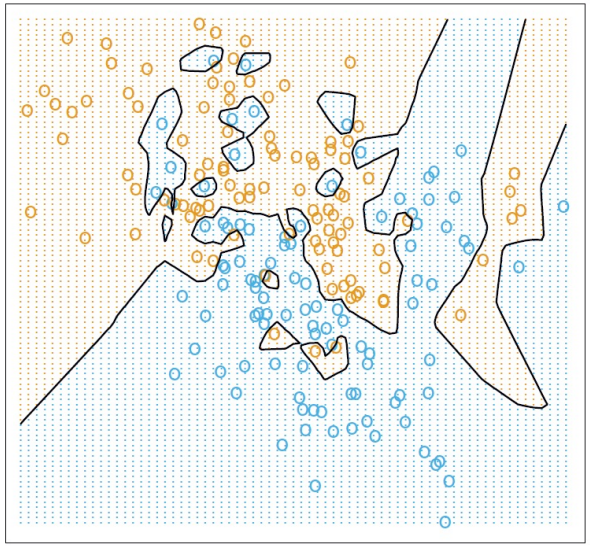


# 1 Nearest-Neighbor




Voronoi diagram (tessellation)

# Linear Separator vs. K Nearest Neighbor

- Which one is better?
- How many parameters?
  - Effective number of parameters