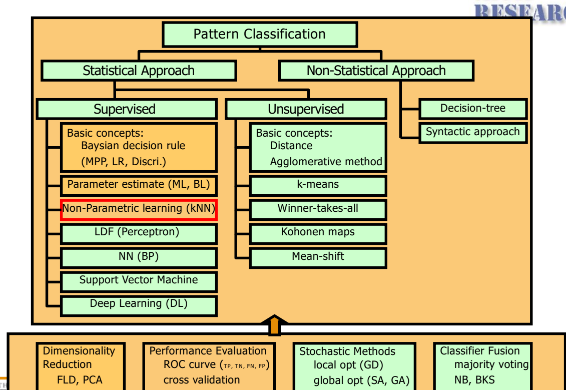
ECE471-571 – Pattern Recognition

Lecture 10 – Nonparametric Density Estimation – k-nearest-neighbor (kNN)

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu



Review - Bayes Decision Rule

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{p(x)}$$

Maximum Posterior Probability: For a given x, if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$, then x belongs to class 1, otherwise, 2.

Likelihood Ratio: If $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$, then x belongs to class 1, otherwise, 2.

Discriminant Function: The classifier will assign a feature vector x to class ω_i if $g_i(x) > g_j(x)$

Case 1: Minimum Euclidean Distance (Linear Machine), $\Sigma_i=\sigma^2 I$
Case 2: Minimum Mahalanobis Distance (Linear Machine), $\Sigma_i = \Sigma$
Case 3: Quadratic classifier, Σ_i = arbitrary

Estimate Gaussian, Two-modal Gaussian

Dimensionality reduction

Performance evaluation and ROC curve

3

Motivation

- Estimate the density functions without the assumption that the pdf has a particular form

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

4

*Problem with Parzen Windows

- Discontinuous window function -> Gaussian
- The choice of h
- Still another one: fixed volume

$$p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right)}{V_n}$$

5

kNN (k-Nearest Neighbor)

- To estimate $p(x)$ from n samples, we can center a cell at x and let it grow until it contains k_n samples, and k_n can be some function of n
- Normally, we let

$$k_n = \sqrt{n}$$

6

kNN in Classification

AICIP RESEARCH

$$p_n(x) = \frac{k_n / n}{V_n}$$

- Given c training sets from c classes, the total number of samples is $$n = \sum_{m=1}^{c} n_m$$
- *Given a point **x** at which we wish to determine the statistics, we find the hypersphere of volume ***V*** which just encloses k points from the combined set. If within that volume, k_m of those points belong to class m, then we estimate the density for class m by

$$p(x \mid \omega_m) = \frac{k_m}{n_m V} \qquad P(\omega_m) = \frac{n_m}{n} \qquad p(x) = \frac{k}{nV}$$

THE UNIVERSITY OF TENNESSEE KNOXVILLE

kNN Classification Rule

AICIP RESEARCH

$$P(\omega_m \mid x) = \frac{p(x \mid \omega_m) P(\omega_m)}{p(x)} = \frac{\frac{k_m}{n_m V} \frac{n_m}{n}}{\frac{k}{nV}} = \frac{k_m}{k}$$

- The decision rule tells us to look in a neighborhood of the unknown feature vector for k samples. If within that neighborhood, more samples lie in class i than any other class, we assign the unknown as belonging to class i.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Problems

AICIP RESEARCH

- What kind of distance should be used to measure “nearest”
  - Euclidean metric is a reasonable measurement
- Computation burden
  - Massive storage burden
  - Need to compute the distance from the unknown to all the neighbors

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Computational Complexity of kNN

AICIP RESEARCH

- In both space (storage space) and time (search time)
- Algorithms reducing the computational burden
  - Computing partial distances
  - Prestructuring
  - Editing the stored prototypes

THE UNIVERSITY OF TENNESSEE

10

Method 1 - Partial Distance

AICIP RESEARCH

- Calculate the distance using some subset r of the full d dimensions. If this partial distance is too large, we don't need to compute further
- What we know about the subspace is indicative of the full space

$$D_r(a,b) = \left(\sum_{k=1}^{r} (a_k - b_k)^2 \right)^{1/2}$$

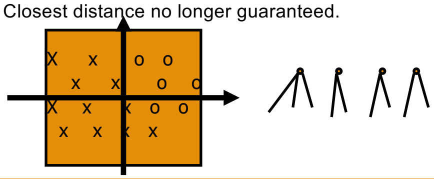
THE UNIVERSITY OF TENNESSEE

11

Method 2 - Prestructuring

AICIP RESEARCH

- Prestructure samples in the training set into a search tree where samples are selectively linked
- Distances are calculated between the testing sample and a few stored "root" samples, and then consider only the samples linked to it
- Closest distance no longer guaranteed.



THE UNIVERSITY OF TENNESSEE

12

AICIP RESEARCH

Method 3 – Editing/Pruning/Condensing

- Eliminate "useless" samples during training
  - For example, eliminate samples that are surrounded by training points of the same category label
- Leave the decision boundaries unchanged

THE UNIVERSITY OF TENNESSEE KNOXVILLE

13

AICIP RESEARCH

Discussion

- Combining three methods
- Other concerns
  - The choice of k
  - The choice of metrics

THE UNIVERSITY OF TENNESSEE KNOXVILLE

14

AICIP RESEARCH

Distance (Metrics) Used by kNN

- Properties of metrics
  - Nonnegativity (D(a,b) >=0)
  - Reflexivity (D(a,b) = 0 iff a=b)
  - Symmetry (D(a,b) = D(b,a))
  - Triangle inequality (D(a,b) + D(b,c) >= D(a,c))
- Different metrics
  - Minkowski metric

$$L_k(\mathbf{a},\mathbf{b}) = \left(\sum_{i=1}^{d} |a_i - b_i|^k\right)^{1/k}$$

    - Manhattan distance (city block distance)

$$L_1(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{d} |a_i - b_i|$$

    - Euclidean distance

$$L_2(\mathbf{a},\mathbf{b}) = \sqrt{\sum_{i=1}^{d} |a_i - b_i|^2}$$

    - When k is inf, maximum of the projected distances onto each of the d coordinate axes

THE UNIVERSITY OF TENNESSEE KNOXVILLE

15

Visualizing the Distances

AICIP RESEARCH

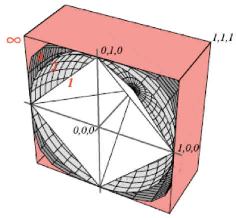


FIGURE 4.19. Each colored surface consists of points a distance 1.0 from the origin, measured using different values for k in the Minkowski metric (k is printed in red). Thus the white surfaces correspond to the L_1 norm (Manhattan distance), the light gray sphere corresponds to the L_2 norm (Euclidean distance), the dark gray ones correspond to the L_4 norm, and the pink box corresponds to the L_∞ norm. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

THE UNIVERSITY OF TENNESSEE 16