# Dimensionality reduction using margin subset of training data

Ilya B. Muchnik
DIMACS, Rutgers University
May 2003

# Contingency tables in epidemiology and pattern recognition: commonalities and differences

| Input \ Outcome | A | B |
|---|---|---|
| I | $n_{11}$ | $n_{12}$ |
| II | $n_{21}$ | $n_{22}$ |

**Input is a multi-factor condition.**

a. in epidemiology input classification is given
b. in pattern recognition the classification is constructed

# Influence estimation between two classifications in epidemiology and pattern recognition

a. testing statistical hypotheses on the whole contingency table in epidemiology
b. in pattern recognition:
   - divide data into two parts: training and testing
   - classifier design based on input training data
   - build contingency table only on testing data
   - statistical analysis of testing results

**Does epidemiology need pattern recognition to analyze the influence of input data on the classification output?**

Yes. When the number of observations is very small, but the number of input factors is large.

## Combinatorial model of the margin

Consider training data as a matrix

$$X = \|x_{ij}\|, i = 1, 2, \ldots, N; j = 1, 2, \ldots, n,$$

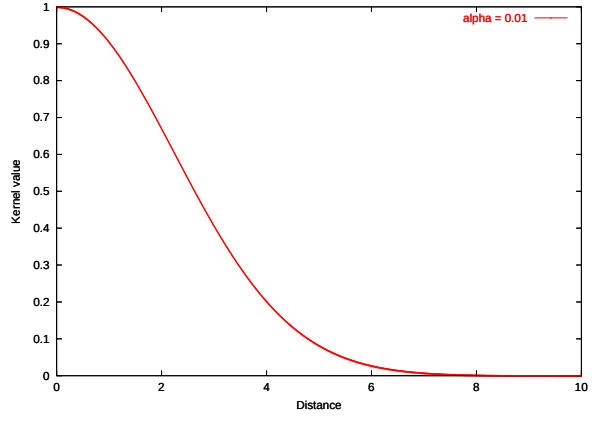
$N$ is the total number of objects, $N = N_A + N_B$ ($N_A$, $N_B$ are the numbers of objects in classes $A$ and $B$, respectively), $n$ is the dimension of the original Euclidean space $R^n$.

# Kernel similarity function

Use a Gaussian function $K(x_i, x_j)$ as similarity function of pairs of objects:

$$K(x_i, x_j) = e^{-\alpha R^2(x_i, x_j)},$$

where $\alpha$ is a free parameter designing the width of the function.



**Classifier rule:**

$$x \in A_s \Longleftrightarrow x = \arg\max_{s=1,k} f_s(x)$$

$$f_s(x) = \sum_{l \in I_s^m} \alpha_l K(x_l, x)$$

## Significance of original variables

$$S(x^j|x) = \frac{1}{k}\sum_{s=1}^{k}\left|\frac{\partial f_s}{\partial x^j}(x)\right|,$$

where $f_s(x) = \sum_{l\in I_s^m} \alpha_l K(x_l, x)$

$$\overline{S}(x^j) = \frac{1}{N}\sum_{l=1}^{N} S(x^j|x^l)$$

Build the matrix $M(W_A, W_B)$ of similarities between all objects from different classes:

$$M(W_A, W_B) = \|K(x_i, x_j)\|,\ \forall x_i \in A, x_j \in B,$$

where $W_A$, $W_B$ are the sets of indices of objects in classes $A$ and $B$, respectively. $|W_A| = N_A,\ |W_B| = N_B$.

**Modelling the margin model for a two-class problem as an optimization criterion**

For an arbitrary sub-matrix $M(H_A, H_B)$ defined by a pair of subsets $H_A \subseteq W_A$ and $H_B \subseteq W_B$, calculate the value

$$F(H_A, H_B) = \min \left( \min_{i \in H_A} \sum_{j \in H_B} K(x_i, x_j), \min_{j \in H_B} \sum_{i \in H_A} K(x_i, x_j), \right)$$

**The set of objects in the margin is the one which maximizes** $F$

## Dimensionality reduction using the proposed margin's set of objects

Denote the union of sets of the margin's vectors $X(H_A^*)$ and $X(H_B^*)$ as *support vectors* (we try to emphasize the analogy with the support vector machine method)

Using the vectors one can map an arbitrary vector $x$ into a new space $Z$ of vectors

$$z(x) =< K(x_i, x), \forall i \in H_A^* \cup H_B^* > .$$

## Modelling the margin model for a multi-class problem

In general, pattern recognition problem is formulated for many classes. In this case the matrix of training data

$$X = \|x_{ij}\|, i = 1, 2, \ldots, N; j = 1, 2, \ldots, n,$$

has to include examples of objects from all $k$ considered classes: $N = N_{A_1} + N_{A_2} + \cdots + N_{A_k}$, where $N_{A_s}$ is the number of objects in class $s$, $s = 1, 2, \ldots, k$.

## Optimization criterion

We want to define a model which gives a margin for the multi-class problem as an extension of the model of a two-class problem described above.
$W =< W_{A_s},\ s = 1, 2, \ldots, k >$ is a family of sets of indices of all $k$ classes and $H =< H_{A_s}, s = 1, 2, \ldots, k >$ is a sub-family of $W$.

The set of vectors in the margin is found as the solution of the problem:

$$\max_{H \subseteq W} F(H)$$

$$F(H) = \min_{s=\overline{1,k}} \min_{i \in H_{A_s}} \sum_{\substack{t=1 \\ t \neq s}}^{k} \sum_{j \in H_{A_t}} K(x_i, x_j)$$