# FINDING THE LARGEST HYPERCAVITY IN A LINEAR DATA SPACE

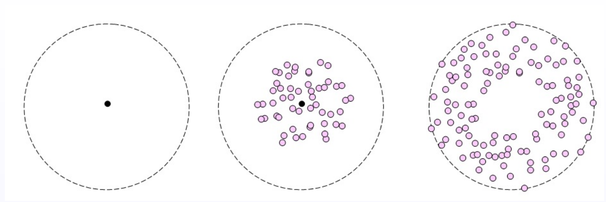
A.Gubareva, V.Sulimova, O.Seredin - Tula State University
A.Larin, V.Mottl - Moscow Institute of Physics and Technology
{a.a.gubareva, vsulimova, oseredin, vmottl}@yandex.ru, ekzebox@gmail.com

ICVSS 2014
Sicily ~ 13-19 July
International Computer Vision Summer School

## Abstract

This poster proposes the definition and the solution of the problem of finding a hypercavity as a data-free hypersphere with a maximal radius. This problem is formulated here as multiextremal problem with constraints in a linear feature space and in a linear space produced by a kernel function. In accordance with the proposed approach, which succeeds to the one-class SVM, a center of a hypersphere is found as a linear combination of some small quantity of so called "support" objects.

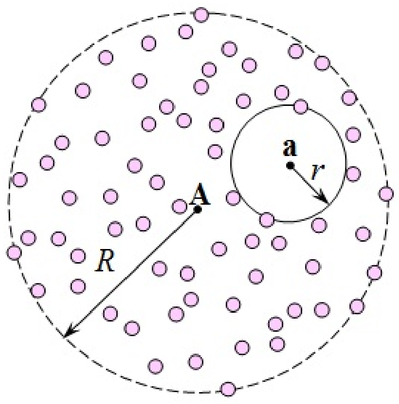
## The evolution of genes



The hypothetical progress of the gene evolution process, accompaned by formation of a hypercavity(according to All-Russia Research Institute of Agricultural Microbiology,St-Petersburg)

Finding a hypercavity would allow to determine possible location of the common ancestor of existing gene variants.

## Problem definition




*Let's we have:*

1) $N$ objects, which are represented by $n$ - dimensional vectors of their real features $\mathbf{x}_i \in \mathbb{R}^n, i = 1, ..., N$ and the Euclidean distance $d(\mathbf{x}', \mathbf{x}'')$ between points $\mathbf{x}', \mathbf{x}'' \in \mathbb{R}^n$ in the respective linear space.

2) a hypershere in this linear space, which is completely characterized by a center $\mathbf{A} \in \mathbb{R}^n$ and a radius $R$ and includes the assumed hypercavity localization region.

*We will seek:*

a hypersphere with a center $\mathbf{a}$ in the convex hull of the set of objects $Conv(X_N)$ a maximal radius $r$, which does not contain objects and is located inside this hypersphere with the center $\mathbf{A} \in \mathbb{R}^n$ and the radius $R$

## Kernel functions

Kernel function $K(\omega_i, \omega_j)$ is a symmetric real-valued function of two arguments, the matrix of which values is positive definite for any finite set of objects $\omega_i, \omega_j \in \Omega, \quad i, j = 1, \ldots, N$.

The kernel function $K(\omega_i, \omega_j)$ produces a Euclidian metric:

$$\rho(\omega_i, \omega_j) = [K(\omega_i, \omega_i) + K(\omega_j, \omega_j) - 2K(\omega_i, \omega_j)]^{1/2}$$

It embeds objects $\Omega$ into Euclidian real linear space $\tilde{\Omega} \supset \Omega$ and plays the role of inner product $(\omega_i, \omega_j) = K(\omega_i, \omega_j)$ in it.

*The kernel trick allows* for algorithms to be more flexible and to identify not only spherical-shaped hypercavities. Moreover, it allows to use the proposed approach in cases when it is problematic to form the useful for further analysis feature space of objects. It is typical, in particular, for analyzing the biomolecular sequences, signals of different nature and images.

## Finding a hypercavity

The problem is formulated as multiextremal problem and brings to necessety of finding a vector of nonnegative coefficients $\lambda = [\lambda_i, ..., \lambda_N]^T$, which defines a hypersphere's.

*In a linear feature space:*

$$\begin{cases} \min\limits_{i=1,\ldots,N} d\left(\sum\limits_{j=1}^{N} \lambda_j \mathbf{x}_j, \mathbf{x}_i\right) \to \max(\lambda), \\ d(\mathbf{A}, \sum\limits_{i=1}^{N} \lambda_i \mathbf{x}_i) \leqslant R - \min\limits_{i=1,\ldots,N} d\left(\sum\limits_{j=1}^{N} \lambda_j \mathbf{x}_j, \mathbf{x}_i\right). \\ \sum\limits_{i=1}^{N} \lambda_i = 1, \quad \lambda_i \geqslant 0, i = 1, ..., N \end{cases}$$

*In a linear space $\tilde{\Omega}$ produced by kernel function $K(\omega_i, \omega_j)$:*

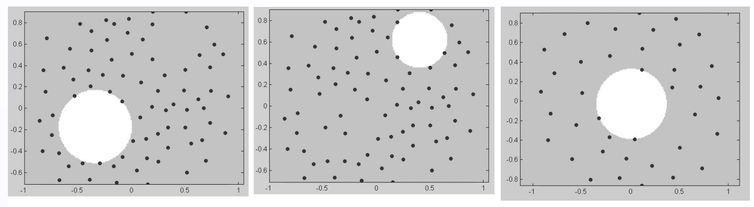
$$\begin{cases} r^2(\boldsymbol{\lambda}) \to \max(\boldsymbol{\lambda}), \\ r^2(\boldsymbol{\lambda}) = \sum\limits_{j=1}^{N} \sum\limits_{t=1}^{N} \lambda_j \lambda_t K(\omega_j, \omega_t) + \min\limits_{i=1\ldots N} \left(K(\omega_i, \omega_i) - 2 \sum\limits_{j=1}^{N} \lambda_j K(\omega_i, \omega_j)\right), \\ \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} K(\omega_i, \omega_j) (\mu_i \mu_j + \lambda_i \lambda_j - 2\lambda_i \mu_j) \leqslant (R - r(\boldsymbol{\lambda}))^2. \\ \sum\limits_{i=1}^{N} \lambda_i = 1, \quad \lambda_i \geqslant 0, \quad i = 1 \ldots N, \end{cases}$$

where $\mu_i, i = 1, ..., N$ are coefficients, which define the center of the outer hypersphere $\mathbf{A} = \sum\limits_{i=1}^{N} \mu_i \omega_i$.

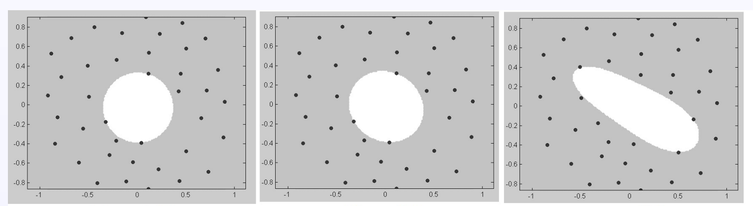
In practice, *only a small number of the coefficients are not equal to zero* and describe the hypersphere's center. These coefficients correspond to objects located on the hyperspheres boundary. These objects are naturally called support ones similar to support vectors defining the hypersphere's center in the Support Vector Data Description (SVDD).

## Results of experiments

Cavity search results in two-dimentional feature space:



Cavity search results in spaces produced by different kernel functions:



(а)α = 0.01 (б)α = 0.1 (в)α = 0.5

So, kernel functions allow to define the cavity boundaries more precisely.

*The cumulative result:* about 92% of 150 hypercavities were correctly found. At that all the detected errors were connected with unreached global maximum.

## Acknowledge

The authors are very appreciate
1) E.Andronov and other scientists of the All-Russia Research Institute of Agricultural Microbiology, St-Petersburg for the interesting biomolecular problem and consultations.
2) Russian Foundation For Basic Research for financial support (project No. 11-07-00728a)