The Metric-based Compactness Hypothesis and Method of Potential Functions for Machine Learning

Oleg Seredin

Tula State University
Tula, Russia

Vadim Mottl

Computing Centre of the Russian
Academy of Sciences
Moscow, Russia

A conceptual basis of dependency estimation: The compactness hypothesis

The set of real-world objects	$\omega \in \Omega$
The hidden characteristic of an object (goal characteristic)	$y \in \mathbb{Y}$
The sought-for decision rule	$\hat{y}(\omega): \Omega \to \mathbb{Y}$

The main idea:

A metric is to be chosen in the set of real-world objects

$\rho(\omega', \omega''): \Omega \times \Omega \to \mathbb{R}$

$\rho(\omega', \omega'') = \rho(\omega'', \omega') \geq 0$, $\rho(\omega', \omega'') > 0$ if $\omega' \neq \omega''$, $\rho(\omega', \omega'') + \rho(\omega'', \omega''') \geq \rho(\omega', \omega''')$

The choice is appropriate if the metric meets **the compactness hypothesis** (Emmanuel Braverman. Ph.D. Thesis: “Experiments on learning a machine to distinguish patterns. 1961, Institute of Automation and Remote Control, Moscow):

If two objects $\omega', \omega'' \in \Omega$ are similar to each other in the sense of the chosen metric $\rho(\omega', \omega'') \cong 0$, their values of the goal characteristic are mostly almost the same $y(\omega') \cong y(\omega'')$.

Hence, the decision rule follows:

For close objects $\rho(\omega', \omega'') \cong 0$, close decisions are to be made

$\hat{y}(\omega') = \hat{y}(\omega'')$	in the problem of pattern recognition $\mathbb{Y} = \{y_1, ..., y_m\}$
$\hat{y}(\omega') \cong \hat{y}(\omega'')$	in the problem of regression estimation $\mathbb{Y} = \mathbb{R}$

Dipole in a metric space

The metric space of real-world objects: $\omega\in\Omega$, $\rho(\omega',\omega'')$ – the metric

A dipole in the metric space – an ordered pair: $<\alpha_{-1},\alpha_1>\in\Omega\times\Omega$

A simplest instantiation of the compactness hypothesis

Membership of an arbitrary object $\omega\in\Omega$ in one of two classes

$$\hat{y}(\omega)=\begin{cases}1\,if\;\;\rho(\alpha_1,\omega)<\rho(\alpha_{-1},\omega),\\ -1\,if\;\;\rho(\alpha_1,\omega)>\rho(\alpha_{-1},\omega).\end{cases}$$

How should the dipole be chosen?

There are too few elements in the set of objects Ω. Besides, only a finite training set of objects is accessible to the observer $\left\{\omega_j,\ j=1,\ldots,N\right\}$

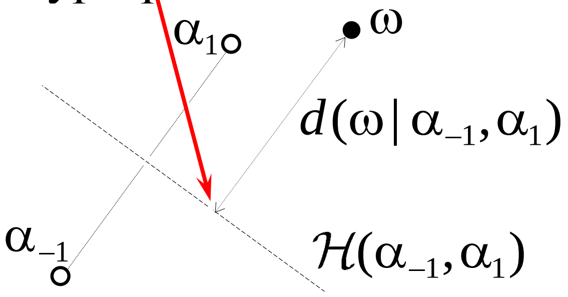
A more “delicate” realization of the compactness hypothesis in a dense hull of the initially sparse metric space

A hypothetical dense space, in which the set of real-world objects is subset of, maybe, isolated elements: $\tilde{\Omega}\supset\Omega$. The dipole is to be chosen in this dense hull.

$\alpha_{-1},\alpha_1\in\tilde{\Omega}$, $\mathcal{H}(\alpha_{-1},\alpha_1)=\left\{\vartheta\in\tilde{\Omega}:\rho(\alpha_{-1},\vartheta)=\rho(\alpha_1,\vartheta)\right\}$ – metric “hyperplane” in $\tilde{\Omega}$

$\underbrace{\omega_{\mathcal{H}}(\alpha_{-1},\alpha_1)}\in\mathcal{H}(\alpha_{-1},\alpha_1)$ – projection of $\omega\in\Omega$ onto the “hyperplane” in $\tilde{\Omega}$

Score function: sign-dependent distance of the point from the “hyperplane” in $\tilde{\Omega}$


α_1
ω
$d(\omega\,|\,\alpha_{-1},\alpha_1)$
α_{-1}
$\mathcal{H}(\alpha_{-1},\alpha_1)$

$$d(\omega\,|\,\alpha_{-1},\alpha_1)=\begin{cases}\rho(\omega_{\mathcal{H}},\omega)\;if\;\;\rho(\alpha_1,\omega)\le\rho(\alpha_{-1},\omega),\\ -\rho(\omega_{\mathcal{H}},\omega)\;if\;\;\rho(\alpha_1,\omega)>\rho(\alpha_{-1},\omega).\end{cases}$$

This is an analog of the linear approach to dependence estimation:
All the decisions on the hidden characteristic y of a real-world object ω are to be made only from the sign-dependent distance $d(\omega\,|\,\alpha_{-1},\alpha_1)$.

Dipole in a metric space

The metric space of real-world objects: $\omega \in \Omega$, $\rho(\omega', \omega'')$ – the metric

A dipole in the metric space – an ordered pair: $< \alpha_{-1}, \alpha_1 > \in \Omega \times \Omega$

A simplest instantiation of the compactness hypothesis

Membership of an arbitrary object $\omega \in \Omega$ in one of two classes

$$\hat{y}(\omega) = \begin{cases} 1\, if\ \rho(\alpha_1, \omega) < \rho(\alpha_{-1}, \omega), \\ -1\, if\ \rho(\alpha_1, \omega) > \rho(\alpha_{-1}, \omega). \end{cases}$$

How should the dipole be chosen?

There are too few elements in the set of objects Ω. Besides, only a finite training set of objects is accessible to the observer $\left\{ \omega_j,\ j = 1, ..., N \right\}$

A more “delicate” realization of the compactness hypothesis in a dense hull of the initially sparse metric space

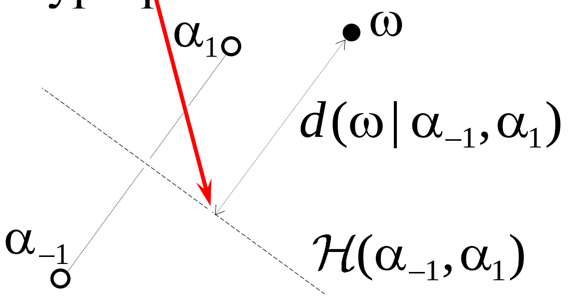
A hypothetical dense space, in which the set of real-world objects is subset of, maybe, isolated elements: $\tilde{\Omega} \supset \Omega$. The dipole is to be chosen in this dense hull.

$\alpha_{-1}, \alpha_1 \in \tilde{\Omega}$, $\mathcal{H}(\alpha_{-1}, \alpha_1) = \left\{ \vartheta \in \tilde{\Omega} : \rho(\alpha_{-1}, \vartheta) = \rho(\alpha_1, \vartheta) \right\}$ – metric “hyperplane” in $\tilde{\Omega}$

$\underbrace{\omega_{\mathcal{H}}(\alpha_{-1}, \alpha_1)} \in \mathcal{H}(\alpha_{-1}, \alpha_1)$ – projection of $\omega \in \Omega$ onto the “hyperplane” in $\tilde{\Omega}$

Score function: sign-dependent distance of the point from the “hyperplane” in $\tilde{\Omega}$

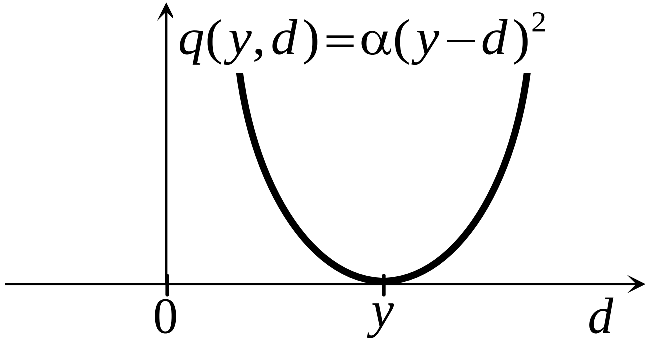
$$d(\omega | \alpha_{-1}, \alpha_1) = \begin{cases} \rho(\omega_{\mathcal{H}}, \omega)\ if\ \rho(\alpha_1, \omega) \le \rho(\alpha_{-1}, \omega), \\ -\rho(\omega_{\mathcal{H}}, \omega)\ if\ \rho(\alpha_1, \omega) > \rho(\alpha_{-1}, \omega). \end{cases}$$


α_1
ω
$d(\omega | \alpha_{-1}, \alpha_1)$
α_{-1}
$\mathcal{H}(\alpha_{-1}, \alpha_1)$

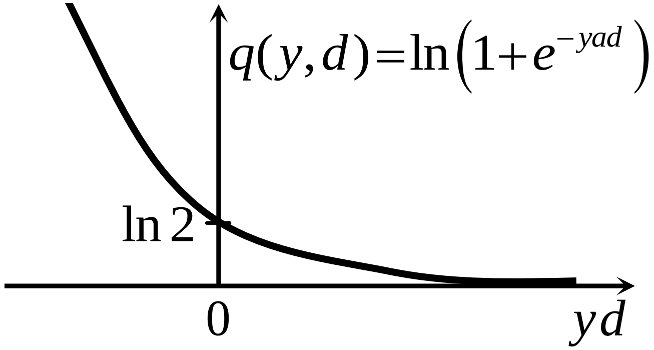
Classification: $\hat{y}(\omega) = \begin{cases} 1\, if\ d(\omega | \alpha_{-1}, \alpha_1)\ + b\ > 0, \\ -1\, if\ d(\omega | \alpha_{-1}, \alpha_1)\ + b\ < 0. \end{cases}$

Regression: $\hat{y}(\omega) = a\, d(\omega | \alpha_{-1}, \alpha_1)\ + b$

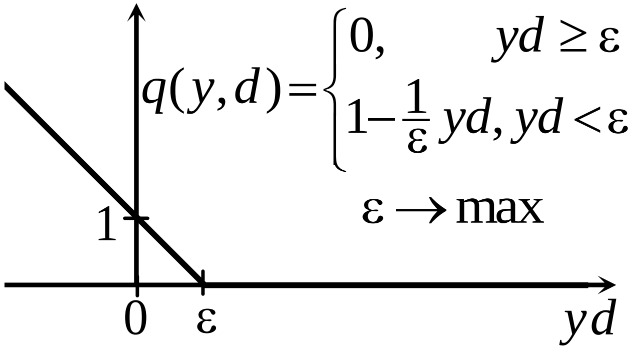
Hinge functions: A bridge to particular problems of dependence estimation



Linear regression with quadratic loss



Two-class pattern recognition: logistic regression



Two-class pattern recognition: maximum margin

But first, the initially sparse metric space of real-world objects is to be embedded into a dense hull $\tilde{\Omega} \supseteq \Omega$, in which the dipole could be continuously chosen.

A way:
Embedding an arbitrary metric space into a linear space.
However, it will not be a proper metric space.
The metric will be defined in it not for all pairs of elements.

But the initial metric space will be embedded isometrically.

An instrument of embedding a metric space into a linear space: Commonality of two elements of a metric space

Metric space Ω with metric $\rho(\omega',\omega'')$:

$\rho(\omega,\omega)=0,\ \rho(\omega',\omega'')>0$ if $\omega'\neq\omega''$; $\rho(\omega',\omega'')=\rho(\omega'',\omega')$;

$\rho(\omega',\omega'')+\rho(\omega'',\omega''')\geq\rho(\omega',\omega''')$ – triangle inequality

Let us choose an element of the metric space $\phi\in\Omega$ as its “center”

Two-argument commonality function $K_{\phi}(\omega',\omega''):\Omega\times\Omega\rightarrow\mathbb{R}$

$$K_{\phi}(\omega',\omega'')=\frac{1}{2}\Big[\,\rho^{2}(\omega',\phi)+\rho^{2}(\omega'',\phi)-\rho^{2}(\omega',\omega'')\Big]$$

Theorem 1. Properties of the commonality function:

$K_{\phi}(\omega',\omega'')=K_{\phi}(\omega'',\omega')$ – symmetry

$K_{\phi}(\omega,\omega)=\rho^{2}(\omega,\phi)\geq 0$ – non-negativity for $\omega'=\omega''$

$|\,K_{\phi}(\omega',\omega'')\,|\leq\sqrt{K_{\phi}(\omega',\omega')}\sqrt{K_{\phi}(\omega'',\omega'')}$ – inequality of Cauchy-Bunyakowsky type

$K_{\tilde{\phi}}(\omega',\omega'')=K_{\phi}(\omega',\omega'')-K_{\phi}(\omega',\tilde{\phi})-K_{\phi}(\omega'',\tilde{\phi})+K_{\phi}(\tilde{\phi},\tilde{\phi})$ – the rule of center translation

$\rho^{2}(\omega',\omega'')=K_{\phi}(\omega',\omega')+K_{\phi}(\omega'',\omega'')-2K_{\phi}(\omega',\omega'')$ – return to the metric

It resembles very much the inner product!
But there are no linear operations as yet!?

Commonality matrix for a finite metric space

Let us assume, for simplicity sake, that the set of real-world objects with metric $\rho(\omega', \omega'')$ is finite $|\Omega| = M$, $\Omega = \{\omega_1, ..., \omega_M\}$.

$\mathbf{K}_\phi = \left[K_\phi(\omega_i, \omega_j), i, j = 1, ..., M \right]$ – the symmetric commonality matrix for some center $\phi \in \Omega$

$\xi_{\phi,1} \in \mathbb{R}, ..., \xi_{\phi,N} \in \mathbb{R}$ eigen values are real numbers

$\underbrace{\xi_{\phi,1} \geq 0, ..., \xi_{\phi,p_\phi} \geq 0}_{p_\phi}, \underbrace{\xi_{\phi,p_\phi+1} < 0, ..., \xi_{\phi,M} < 0}_{q_\phi}$ for an arbitrary metric, the matrix may be not positive definite

$p_\phi + q_\phi = M$ – signature of the commonality

Theorem 2. *The signature of matrix* $\mathbf{K}_\phi$ *does not depend on the choice of the center in the metric space* $\phi \in \Omega$.

Thus, any metric on a finite metric space $|\Omega| = M$ is characterized by its signature $p + q = M$.

Definition. *Metric is said to be proto-Euclidean one*, *if for any subset of objects* $\{\omega_1, ..., \omega_N\}$ *matrix* $\mathbf{P}_N = \left[-\rho^2(\omega_j, \omega_l), \ j, l = 1, ..., N \right]$ *is conditionally positive semidefinite:*

$$\mathbf{c}^T \mathbf{P}_N \mathbf{c} \geq 0, \ \mathbf{1}^T \mathbf{c} = \sum_{j=1}^N c_j = 0.$$

Theorem 3. *The signature of a proto-Euclidean metric* $p = M, q = 0$.

In other words, the commonality matrix of a finite metric space with a proto-Euclidean metric (finite proto-Euclidean metric space) is positive semidefinite for any choice of the center.

Indefinite inner product and pseudo-Euclidean linear space

Let metric $\rho(\omega', \omega'')$ on the set of objects $|\Omega| = M$ be of signature $p + q = M$.

Let the following two-argument real-valued function be defined over the entire linear space $\mathbb{R}^M$:

$$K(\mathbf{x}', \mathbf{x}'') = \mathbf{x}'^T \mathbf{J}_{p,q} \mathbf{x}'' : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R} \qquad \mathbf{J}_{p,q} = \begin{pmatrix} \mathbf{I}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & -\mathbf{I}_{q\times q} \end{pmatrix} (M \times M)$$

signature depends only on the metric

Properties:

Symmetricity $\qquad K(\mathbf{x}', \mathbf{x}'') = K(\mathbf{x}'', \mathbf{x}')$

Bilineariry $\qquad K(c'\mathbf{x}' + c''\mathbf{x}'', \mathbf{x}''') = c'K(\mathbf{x}', \mathbf{x}''') + c''K(\mathbf{x}'', \mathbf{x}''')$

The property of non-negativity when the arguments coincide is absent $\qquad K(\mathbf{x}, \mathbf{x}) < 0$ for some $\mathbf{x} \in \mathbb{R}^M$

$$K(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T \mathbf{J}_{p,q} \mathbf{x} = (\mathbf{u}^T \mathbf{v}^T) \begin{pmatrix} \mathbf{I}_{p\times p} & \mathbf{0}_{p\times q} \\ \mathbf{0}_{q\times p} & -\mathbf{I}_{q\times q} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \mathbf{u}^T \mathbf{u} - \mathbf{v}^T \mathbf{v} < 0 \text{ if } \mathbf{u}^T \mathbf{u} < \mathbf{v}^T \mathbf{v}$$

This is no inner product!

Such a function is said to be indefinite inner product,
and the linear space $\mathbb{R}^M$ – pseudo-Euclidean linear space.

A finite-dimensional linear apace with indefinite inner product is called the Krein space.

Mark Krein (1907-1989),
Professor of Odessa Civil-Engineering Institute.

Vector length and distance between vectors in the pseudo-Euclidean linear space

The pseudo-Euclidean linear space is neither normed (not all elements have norm) nor metric (metric is defined not for all pairs of elements).

Nevertheless, it is possible to isometrically embed an arbitrary finite metric space of signature $p+q=M$ into a pseudo-Euclidean linear space of the same signature.

Isometric embedding of an arbitrary metric space into a pseudo-Euclidean linear space

Let $\Omega=\{\omega_1,...,\omega_M\}$ be a finite metric space with metric $\rho(\omega',\omega'')$, $p+q=M$ be its signature, and $\phi\in\Omega$ be an arbitrary element assigned as the center.

Theorem 4. *There exist* M *real-valued vectors* $\mathbf{x}_{\phi,j}\in\mathbb{R}^M$, $j=1,...,N$, *in the pseudo-Euclidean linear space of signature* $p+q=M$, *for which are defined:*

the norm $r(\mathbf{x}_{\phi,j},\mathbf{0})=(\underbrace{\mathbf{x}_{\phi,j}^T\mathbf{J}_{p,q}\mathbf{x}_{\phi,j}}_{\geq 0})^{1/2}=\rho(\omega_j,\phi)$ *distance to the zero point*

the metric $r(\mathbf{x}_{\phi,j},\mathbf{x}_{\phi,l})=[\underbrace{(\mathbf{x}_{\phi,j}-\mathbf{x}_{\phi,l})^T\mathbf{J}_{p,q}(\mathbf{x}_{\phi,j}-\mathbf{x}_{\phi,l})}_{\geq 0}]^{1/2}=\rho(\omega_j,\omega_l)$ *distance between pairs of vectors*

Discriminant (score) functions in a pseudo-Euclidean linear space

Discriminant dipole:

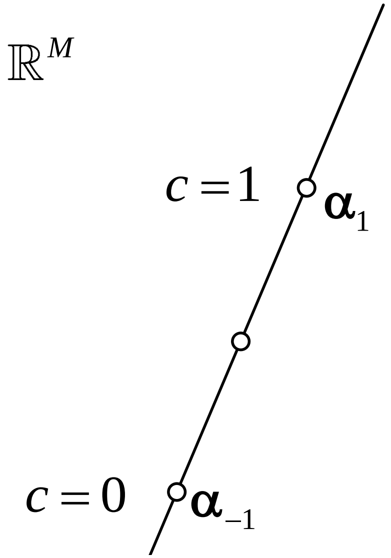
An ordered pair of vectors $<\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1>\in\mathbb{R}^M\times\mathbb{R}^M$, $\quad\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1\in\mathbb{R}^M$ – the nodes of the dipole.

In what follows, we shall consider only dipoles, for which the metric distance between the nodes is defined, i.e., the squared distance between the nodes is positive

$$r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)^T\,\mathbf{J}_p(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)>0$$

The axis defined by a dipole:

$$\left\{\mathbf{x}_c\in\mathbb{R}^M:\ \mathbf{x}_c=(1-c)\boldsymbol{\alpha}_{-1}+c\boldsymbol{\alpha}_1\in\mathbb{R}^M,\ c\in\mathbb{R}\right\}\subset\mathbb{R}^M$$



Discriminant functions in a pseudo-Euclidean linear space

Discriminant dipole:

An ordered pair of vectors $<\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1>\in\mathbb{R}^M\times\mathbb{R}^M$, $\quad\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1\in\mathbb{R}^M$ – the nodes of the dipole.

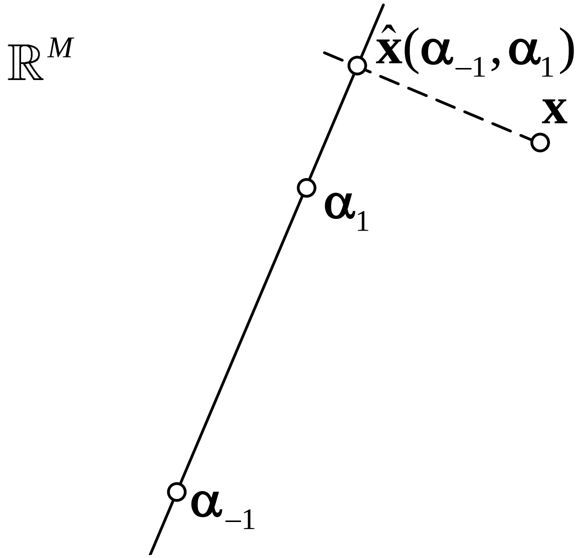
In what follows, we shall consider only dipoles, for which the metric distance between the nodes is defined, i.e., the squared distance between the nodes is positive

$$r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)^T\,\mathbf{J}_p(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)>0$$

The axis defined by a dipole:

$$\left\{\mathbf{x}_c\in\mathbb{R}^M:\ \mathbf{x}_c=(1-c)\boldsymbol{\alpha}_{-1}+c\boldsymbol{\alpha}_1\in\mathbb{R}^M,\ c\in\mathbb{R}\right\}\subset\mathbb{R}^M$$

$\hat{\mathbf{x}}(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$ – projection of a vector $\mathbf{x}\in\mathbb{R}^M$ on the axis defined by the dipole


$\mathbb{R}^M$
$\hat{\mathbf{x}}(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$
$\mathbf{x}$
$\boldsymbol{\alpha}_1$
$\boldsymbol{\alpha}_{-1}$

Discriminant functions in a pseudo-Euclidean linear space

Discriminant dipole:

An ordered pair of vectors $<\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1>\in\mathbb{R}^M\times\mathbb{R}^M$, $\quad\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1\in\mathbb{R}^M$ – the nodes of the dipole.

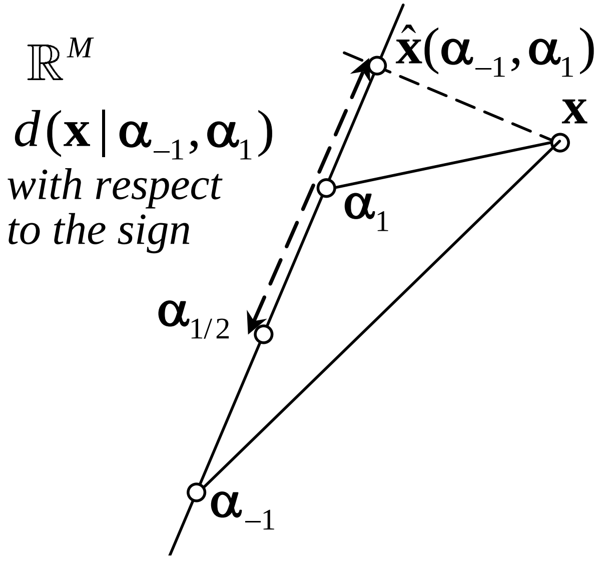
In what follows, we shall consider only dipoles, for which the metric distance between the nodes is defined, i.e., the squared distance between the nodes is positive

$$r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)^T\mathbf{J}_p(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)>0$$

The axis defined by a dipole:

$$\left\{\mathbf{x}_c\in\mathbb{R}^M:\ \mathbf{x}_c=(1-c)\boldsymbol{\alpha}_{-1}+c\boldsymbol{\alpha}_1\in\mathbb{R}^M,\ c\in\mathbb{R}\right\}\subset\mathbb{R}^M$$

$\hat{\mathbf{x}}(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$ – projection of a vector $\mathbf{x}\in\mathbb{R}^M$ on the axis defined by the dipole


$\mathbb{R}^M$
$d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$
with respect
to the sign
$\hat{\mathbf{x}}(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$
$\mathbf{x}$
$\boldsymbol{\alpha}_1$
$\boldsymbol{\alpha}_{1/2}$
$\boldsymbol{\alpha}_{-1}$

Theorem 5. *If* $r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)>0$ *then* $d^2(\mathbf{x}|\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)\geq 0$ *for any vector* $\mathbf{x}\in\mathbb{R}^M$*, and, with respect to the sign*

$$d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=\frac{1}{2}\Big(r^2(\boldsymbol{\alpha}_{-1},\mathbf{x})-r^2(\boldsymbol{\alpha}_1,\mathbf{x})\Big)\frac{1}{r(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)}.$$

Discriminant functions in a pseudo-Euclidean linear space

Discriminant dipole:

An ordered pair of vectors $<\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1>\in\mathbb{R}^M\times\mathbb{R}^M$, $\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1\in\mathbb{R}^M$ – the nodes of the dipole.

In what follows, we shall consider only dipoles, for which the metric distance between the nodes is defined, i.e., the squared distance between the nodes is positive

$$r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)^T\,\mathbf{J}_p(\boldsymbol{\alpha}_{-1}-\boldsymbol{\alpha}_1)>0$$

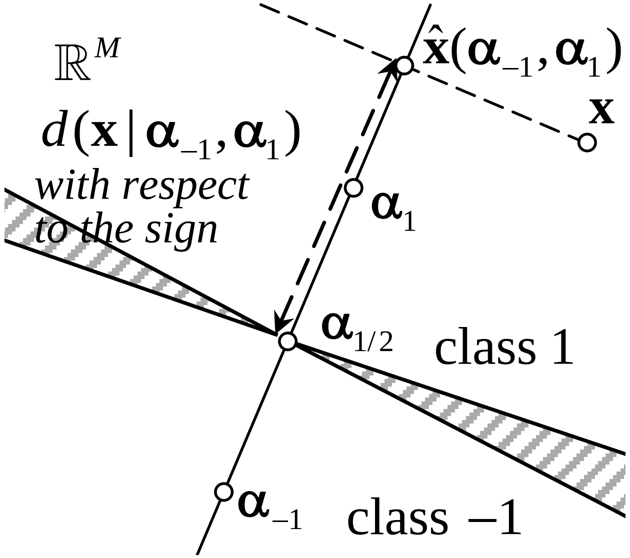
The axis defined by a dipole:

$$\left\{\mathbf{x}_c\in\mathbb{R}^M\!:\ \mathbf{x}_c=(1-c)\boldsymbol{\alpha}_{-1}+c\boldsymbol{\alpha}_1\in\mathbb{R}^M,\ c\in\mathbb{R}\right\}\subset\mathbb{R}^M$$

$\hat{\mathbf{x}}(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$ – projection of a vector $\mathbf{x}\in\mathbb{R}^M$ on the axis defined by the dipole

Theorem 5. *If* $r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)>0$ *then* $d^2(\mathbf{x}|\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)\geq 0$ *for any vector* $\mathbf{x}\in\mathbb{R}^M$*, and, with respect to the sign*

$$d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=\frac{1}{2}\Big(r^2(\boldsymbol{\alpha}_{-1},\mathbf{x})-r^2(\boldsymbol{\alpha}_1,\mathbf{x})\Big)\frac{1}{r(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)}.$$


$\mathbb{R}^M$
$d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$
with respect to the sign
$\hat{\mathbf{x}}(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)$
$\mathbf{x}$
$\boldsymbol{\alpha}_1$
$\boldsymbol{\alpha}_{1/2}$
class 1
$\boldsymbol{\alpha}_{-1}$
class –1

Parametric family of discriminant functions:

$$d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)\begin{cases}>0\Rightarrow \textit{the positive part},\\ =0\Rightarrow \textit{the neutral part},\\ <0\Rightarrow \textit{the negative part}.\end{cases}$$

Training: Choice of the discriminant dipole

A training set in the metric space of real-world objects $\omega\in\Omega$:

$\left\{(\omega_j, y_j),\ j=1,...,N\right\}$, distances $\left[\rho(\omega_j,\omega_l),\ j,l=1,...,N\right]$, classes $y_j = y(\omega_j)$.

The center $\phi\in\Omega$ + a real-world object $\omega\in\Omega$ $\Rightarrow$ $\mathbf{0}\in\mathbb{R}^M$ + vector $\mathbf{x}_{\phi,\omega}\in\mathbb{R}^M$ (in the pseudo-Euclidean space).

The training process boils down to the choice of the dipole $<\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1>\in\mathbb{R}^M\times\mathbb{R}^M$:

We shall consider the dipole's nodes as linear (affine) combinations of the vectors $\{\mathbf{x}_{\phi,\omega_1},...,\mathbf{x}_{\phi,\omega_N}\}$, into which the objects of the training set are mapped:

$$\boldsymbol{\alpha}_{-1} = \sum\nolimits_{j=1}^{N} c_{-1,j}\mathbf{x}_{\phi,\omega_j},\ \sum\nolimits_{j=1}^{N} c_{-1,j} = 1,$$
$$\boldsymbol{\alpha}_{1} = \sum\nolimits_{j=1}^{N} c_{1,j}\mathbf{x}_{\phi,\omega_j},\quad \sum\nolimits_{j=1}^{N} c_{1,j} = 1,$$
$$a_j = c_{1,j} - c_{-1,j},\quad \sum\nolimits_{j=1}^{N} a_j = 0.$$

Score function:
$$d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1) = \frac{1}{2}\Big(r^2(\boldsymbol{\alpha}_{-1},\mathbf{x}) - r^2(\boldsymbol{\alpha}_1,\mathbf{x})\Big)\frac{1}{r(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)}$$

Theorem 6: *For any real-world object* $\omega\in\Omega$, *any training set* $\left\{\omega_j,\ j=1,...,N\right\}\subset\Omega$, *and any* $\phi\in\Omega$:

$$\boxed{r^2(\boldsymbol{\alpha}_{-1},\mathbf{x}_{\phi,\omega}) - r^2(\boldsymbol{\alpha}_1,\mathbf{x}_{\phi,\omega}) = \sum_{l=1}^{N}\Big(-\rho^2(\omega_j,\omega_l)\Big)a_j + b},\ \boxed{r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1) = \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\Big(-\rho^2(\omega_j,\omega_l)\Big)a_j a_l}.$$

Training: Choice of the discriminant dipole

A training set in the metric space of real-world objects $\omega\in\Omega$:

$\left\{(\omega_j, y_j),\ j=1,\dots,N\right\}$, distances $\left[\rho(\omega_j,\omega_l),\ j,l=1,\dots,N\right]$, classes $y_j = y(\omega_j)$.

The center $\phi\in\Omega$ + a real-world object $\omega\in\Omega \ \Rightarrow\ \mathbf{0}\in\mathbb{R}^M$ + vector $\mathbf{x}_{\phi,\omega}\in\mathbb{R}^M$ (in the pseudo-Euclidean space).

The training process boils down to the choice of the dipole $<\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1>\in\mathbb{R}^M\times\mathbb{R}^M$:

We shall consider the dipole’s nodes as linear (affine) combinations of the vectors $\{\mathbf{x}_{\phi,\omega_1},\dots,\mathbf{x}_{\phi,\omega_N}\}$, into which the objects of the training set are mapped:

$$\begin{aligned}\boldsymbol{\alpha}_{-1} &= \sum\nolimits_{j=1}^{N} c_{-1,j}\mathbf{x}_{\phi,\omega_j}, \ \sum\nolimits_{j=1}^{N} c_{-1,j}=1,\\ \boldsymbol{\alpha}_{1} &= \sum\nolimits_{j=1}^{N} c_{1,j}\mathbf{x}_{\phi,\omega_j}, \ \sum\nolimits_{j=1}^{N} c_{1,j}=1,\end{aligned}\qquad a_j = c_{1,j}-c_{-1,j},\quad \sum\nolimits_{j=1}^{N} a_j = 0.$$

Scaled score function: $$r(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)d(\mathbf{x}\,|\,\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=\frac{1}{2}\Big(r^2(\boldsymbol{\alpha}_{-1},\mathbf{x})-r^2(\boldsymbol{\alpha}_1,\mathbf{x})\Big)$$

For any real-world object $\omega\in\Omega$, any training set $\left\{\omega_j,\ j=1,\dots,N\right\}\subset\Omega$, and any $\phi\in\Omega$:

$$r^2(\boldsymbol{\alpha}_{-1},\mathbf{x}_{\phi,\omega})-r^2(\boldsymbol{\alpha}_1,\mathbf{x}_{\phi,\omega})=\sum_{l=1}^{N}\Big(-\rho^2(\omega_l,\omega)\Big)a_l+b,$$

$$r^2(\boldsymbol{\alpha}_{-1},\boldsymbol{\alpha}_1)=\frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\Big(-\rho^2(\omega_j,\omega_l)\Big)a_j a_l\,.$$

The general training problem

Training set: $\left\{(\omega_j, y_j),\ j=1,...,N\right\},\quad \rho(\omega_j,\omega_l),\ j,l=1,...,N$

Unified items of the training criterion	Score function	$d(\omega \mid a_1,...,a_N,b)=\sum_{j=1}^{N}\left(-\rho^2(\omega_j,\omega)\right)a_j+b$
	Regularization function: squared dipole length	$\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l>0,\ \sum_{l=1}^{N}a_l=0$
The only problem-specific item	Hinge function	$q\left(y,d(\omega \mid a_1,...,a_N,b)\right)=q(y,\omega,a_1,...,a_N,b)$

The unified training criterion

$$\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l+C\underbrace{\sum_{j=1}^{N}q(y_j,\omega_j,a_1,...,a_N,b)}\to\min(a_1,...,a_n,b),\ \sum_{l=1}^{N}a_l=0.$$

Particular versions of the problem-specific hinge function

Regression estimation	$\sum_{j=1}^{N}\left\{y-\left(\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_l+b\right)\right\}^2$
Logistic regression	$\sum_{j=1}^{N}\ln\left\{1+\exp\left[-y\left(\sum_{l=1}^{N}\left(-\rho^2(\omega_l,\omega)\right)a_l\right)+b\right]\right\}$
Maximum margin	$\sum_{j=1}^{N}\max\left\{0,\ 1-y\left(\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_l\right)+b\right\}$

All the particular hinge-function summands are convex.

The general training problem

Training set: $\left\{(\omega_j, y_j),\ j=1,...,N\right\},\quad \rho(\omega_j,\omega_l),\ j,l=1,...,N$

Unified items of the training criterion	Score function	$d(\omega \mid a_1,...,a_N,b)=\sum_{j=1}^{N}\left(-\rho^2(\omega_j,\omega)\right)a_j+b$
	Regularization function: squared dipole length	$\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l>0,\ \sum_{l=1}^{N}a_l=0$
The only problem-specific item	Hinge function	$q\left(y,d(\omega \mid a_1,...,a_N,b)\right)=q(y,\omega,a_1,...,a_N,b)$

The unified training criterion

$$\underbrace{\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l}_{}+C\sum_{j=1}^{N}q(y_j,\omega_j,a_1,...,a_N,b)\rightarrow\min(a_1,...,a_n,b),\ \underbrace{\sum_{l=1}^{N}a_l=0}_{}.$$

Particular versions of the problem-specific hinge function

Regression estimation	$\sum_{j=1}^{N}\left\{y-\left(\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_l+b\right)\right\}^2$
Logistic regression	$\sum_{j=1}^{N}\ln\left\{1+\exp\left[-y\left(\sum_{l=1}^{N}\left(-\rho^2(\omega_l,\omega)\right)a_l\right)+b\right]\right\}$
Maximum margin	$\sum_{j=1}^{N}\max\left\{0,\ 1-y\left(\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_l\right)+b\right\}$

All the particular hinge-function summands are convex.
But the unified regularization term may be nonconvex in the case of an arbitrary metric.

Nonconvex maximum margin classifier

The training criterion

$$\begin{cases} \sum_{j=1}^{N}\sum_{l=1}^{N}\big(-\rho^2(\omega_j,\omega_l)\big)a_j a_l + C\sum_{j=1}^{N}\xi_j \to \min(a_1,\dots,a_N,b,\delta_1,\dots,\delta_N), \\ y_j\Big[\sum_{l=1}^{N}\big(-\rho^2(\omega_j,\omega_l)\big)a_l + b\Big] \geqslant 1-\xi_j, \quad \xi_j \geqslant 0,\ j=1,\dots,N. \end{cases}$$

Restriction $\sum_{j=1}^{N} a_j = 0$ is met automatically.

In the general case, this is a nonconvex quadratic programming problem.
Function $\sum_{j=1}^{N}\sum_{l=1}^{N}\big(-\rho^2(\omega_j,\omega_l)\big)a_j a_l$ under equality restriction $\sum_{j=1}^{N} a_j = 0$ is convex only for a proto-Euclidean metric.

The necessary condition of the minimum: The dual problem

$$\begin{cases} \sum_{j=1}^{N}\lambda_j - (1/4)\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l\big(-\rho^2(\omega_j,\omega_l)\big)\lambda_j\lambda_l \to \max(\lambda_1,\dots,\lambda_N), \\ \sum_{j=1}^{N} y_j\lambda_j = 0,\ 0 \leqslant \lambda_j \leqslant C,\ j=1,\dots,N. \end{cases}$$

$$a_j = (1/2)y_j\lambda_j,\ \sum_{j=1}^{N} a_j = 0,\ b = \frac{(1/2)\sum_{j:0<\lambda_j<\bar{C}}\lambda_j\sum_{l:\lambda_l>0}\rho^2(\omega_j,\omega_l)y_l\lambda_l - C\sum_{j:\lambda_j=\bar{C}} y_j}{\sum_{j:0<\lambda_j<C}\lambda_j},$$

For an arbitrary metric, the solution of the non-concave dual problem may result in the negative value of the squared maximum margin:

$$1/\varepsilon^2 = (1/2)\sum_{j=1}^{N}\sum_{l=1}^{N} y_j y_l\big(-\rho^2(\omega_j,\omega_l)\big)\lambda_j\lambda_l \gtrless 0.$$

Nonconvex maximum margin classifier

The training criterion

$$\begin{cases}\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l + C\sum_{j=1}^{N}\xi_j \to \min(a_1,\dots,a_N,b,\delta_1,\dots,\delta_N),\\ y_j\left[\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_l + b\right]\geqslant 1-\xi_j, \quad \xi_j\geqslant 0,\ j=1,\dots,N.\end{cases}$$

Restriction $\sum_{j=1}^{N}a_j = 0$ is met automatically.

In the general case, this is a nonconvex quadratic programming problem.
Function $\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l$ under equality restriction $\sum_{j=1}^{N}a_j = 0$ is convex only for a proto-Euclidean metric.

The necessary condition of the minimum: The dual problem

$$\begin{cases}\sum_{j=1}^{N}\lambda_j - (1/4)\sum_{j=1}^{N}\sum_{l=1}^{N}y_j y_l\left(-\rho^2(\omega_j,\omega_l)\right)\lambda_j\lambda_l \to \max(\lambda_1,\dots,\lambda_N),\\ \sum_{j=1}^{N}y_j\lambda_j = 0,\ 0\leqslant\lambda_i\leqslant C, j=1,\dots,N.\end{cases}$$

$$\sum_{j=1}^{N}\sum_{l=1}^{N}y_j y_l\left(-\rho^2(\omega_j,\omega_l)\right)\lambda_j\lambda_l \geq \delta > 0$$ quadratically constrained quadratic optimization

$$a_j = (1/2)y_j\lambda_j,\ \sum_{j=1}^{N}a_j = 0,\ b = \frac{(1/2)\sum_{j:0<\lambda_j<\bar{C}}\lambda_j\sum_{l:\lambda_l>0}\rho^2(\omega_j,\omega_l)y_l\lambda_l - C\sum_{j:\lambda_j=\bar{C}}y_j}{\sum_{j:0<\lambda_j<C}\lambda_j},$$

In this case, the squared maximum margin remains always positive for an arbitrary metric:

$$1/\varepsilon^2 = (1/2)\sum_{j=1}^{N}\sum_{l=1}^{N}y_j y_l\left(-\rho^2(\omega_j,\omega_l)\right)\lambda_j\lambda_l > 0.$$

Decision rule (score function) in an arbitrary metric space

The equivalent form of the discriminant function, which is completely determined by $N+1$ real numbers $(a_1,...,a_N,b)$:

new object

new object

$$d(\omega \mid \underbrace{a_1,...,a_N,b}_{\textit{Parameters to be estimated}}) = \frac{1}{2}\left[\sum_{j=1}^{N} a_j\left(-\rho^2(\omega_j,\,\omega)\right)+b\right] \gtrless 0,$$

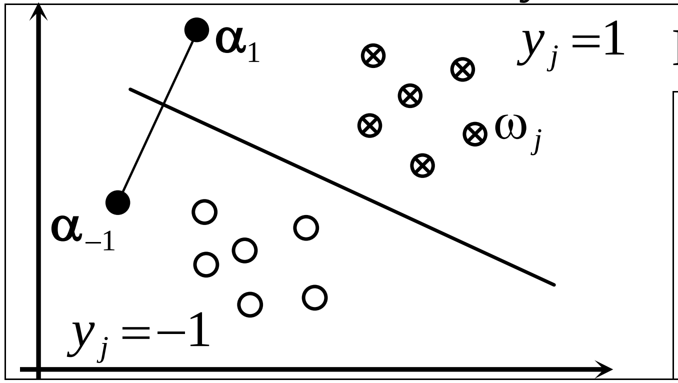
objects of the training set

$$\sum_{j=1}^{N} a_j = 0, \quad \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l > 0.$$

This is a linear decision rule in the pseudo-Euclidean space spanned over the given metric space of real-world objects

Eliminating the difference between linear and nonlinear decision rules in a pseudo-Euclidean space

Despite the fact that neither a metric space nor its pseudo-Euclidean embedding do not lend themselves to easy geometrical interpretation, we can conventionally demonstrate the essence of this boundary in an appropriate Euclidean space:



Linearly separable classes

Decision rule in an arbitrary metric space

The equivalent form of the discriminant function, which is completely determined by $N+1$ real numbers $(a_1,...,a_N,b)$:

new object *new object*

$$d(\omega \mid \underbrace{a_1,...,a_N,b}_{\textit{Parameters to be estimated}}) = \frac{1}{2}\left[\sum_{j=1}^{N} a_j\left(-\rho^2(\omega_j,\omega)\right)+b\right] \gtrless 0,$$

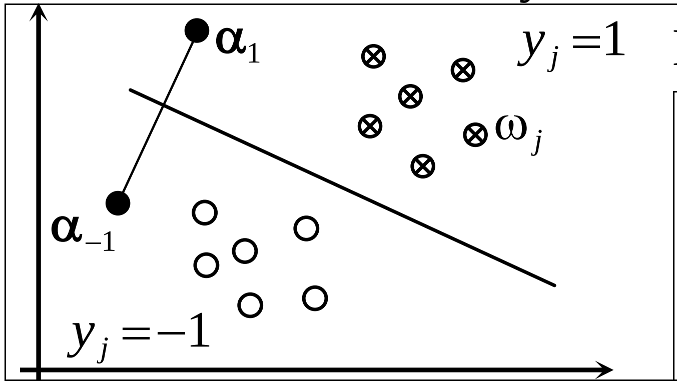
objects of the training set

$$\sum_{j=1}^{N} a_j = 0, \quad \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l > 0.$$

This is a linear decision rule in the pseudo-Euclidean space spanned over the given metric space of real-world objects

Eliminating the difference between linear and nonlinear decision rules in a pseudo-Euclidean space

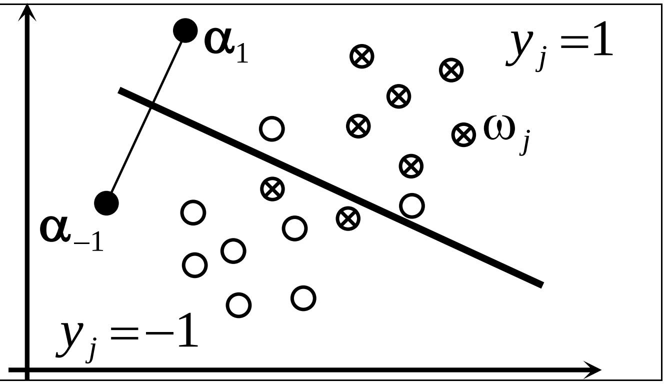
Despite the fact that neither a metric space nor its pseudo-Euclidean embedding do not lend themselves to easy geometrical interpretation, we can conventionally demonstrate the essence of this boundary in an appropriate Euclidean space:



Linearly separable classes

Linearly inseparable classes

The separating ability depends on the accepted metric



Decision rule in an arbitrary metric space

The equivalent form of the discriminant function, which is completely determined by $N+1$ real numbers $(a_1,...,a_N,b)$:

new object

new object

$$d(\omega \mid \underbrace{a_1,...,a_N,b}_{\textit{Parameters to be estimated}}) = \frac{1}{2}\left[\sum\nolimits_{j=1}^{N} a_j\big(-\rho^2(\omega_j,\omega)\big)+b\right] \gtrless 0,$$

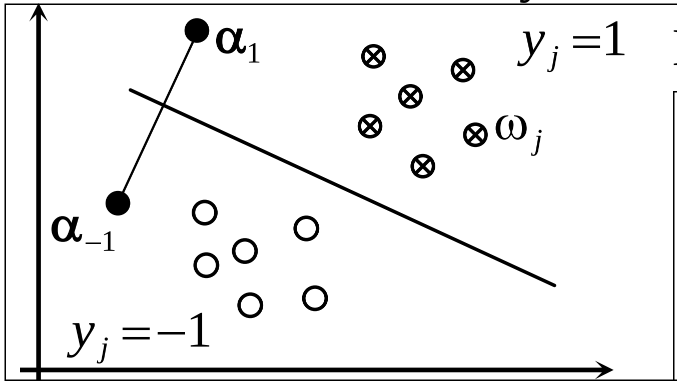
objects of the training set

$$\sum\nolimits_{j=1}^{N} a_j = 0, \quad \frac{1}{2}\sum\nolimits_{j=1}^{N}\sum\nolimits_{l=1}^{N}\big(-\rho^2(\omega_j,\omega_l)\big)a_j a_l > 0.$$

This is a linear decision rule in the pseudo-Euclidean space spanned over the given metric space of real-world objects

Eliminating the difference between linear and nonlinear decision rules in a pseudo-Euclidean space

Despite the fact that neither a metric space nor its pseudo-Euclidean embedding do not lend themselves to easy geometrical interpretation, we can conventionally demonstrate the essence of this boundary in an appropriate Euclidean space:

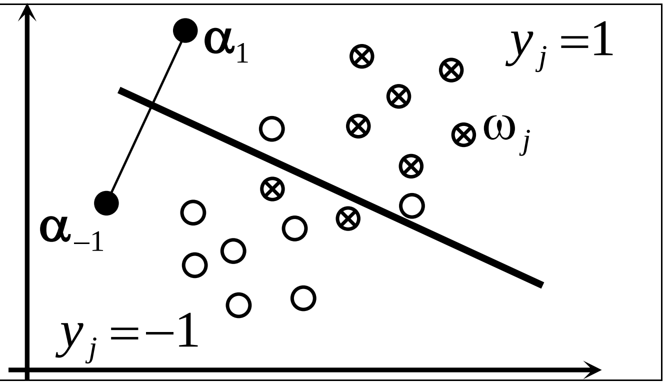


Linearly separable classes

Linearly inseparable classes

An urgent desire:

to eliminate the difference between separable and inseparable clases



Decision rule in an arbitrary metric space

The equivalent form of the discriminant function, which is completely determined by $N+1$ real numbers $(a_1,\dots,a_N,b)$:

new object

new object

$$d(\omega\,|\,\underbrace{a_1,\dots,a_N,b}_{\textit{Parameters to be estimated}}) = \frac{1}{2}\left[\sum_{j=1}^{N} a_j\left(-\rho^2(\omega_j\,,\,\omega)\right)+b\right] \gtrless 0,$$

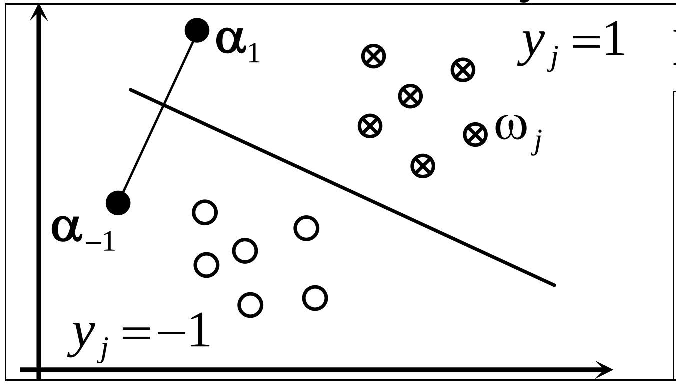
objects of the training set

$$\sum_{j=1}^{N} a_j = 0, \quad \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l > 0.$$

This is a linear decision rule in the pseudo-Euclidean space spanned over the given metric space of real-world objects

Eliminating the difference between linear and nonlinear decision rules in a pseudo-Euclidean space

Despite the fact that neither a metric space nor its pseudo-Euclidean embedding do not lend themselves to easy geometrical interpretation, we can conventionally demonstrate the essence of this boundary in an appropriate Euclidean space:

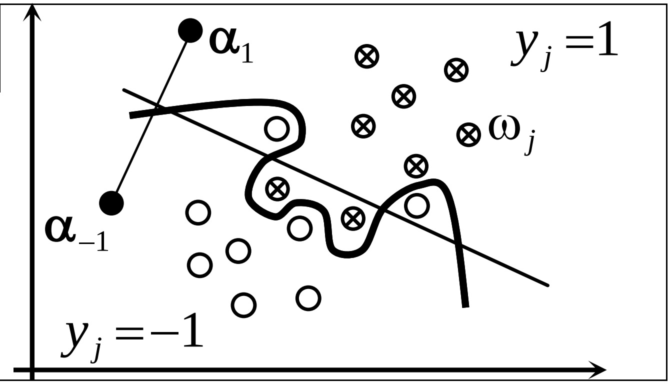


Linearly separable classes

Linearly inseparable classes

An urgent desire:

to eliminate the difference between separable and inseparable classes



Decision rule in an arbitrary metric space

The equivalent form of the discriminant function, which is completely determined by $N+1$ real numbers $(a_1,...,a_N,b)$:

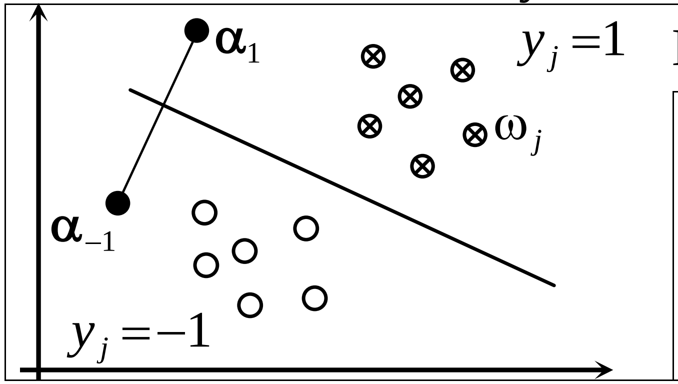
$$d(\overset{\text{new object}}{\omega} \mid \underbrace{a_1,...,a_N,b}_{\text{Parameters to be estimated}}) = \frac{1}{2}\left[\sum_{j=1}^{N} a_j\left(-\rho^2(\underset{\text{objects of the training set}}{\omega_j}, \overset{\text{new object}}{\omega})\right)+b\right] \gtreqless 0,$$

$$\sum_{j=1}^{N} a_j = 0,\quad \frac{1}{2}\sum_{j=1}^{N}\sum_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l > 0.$$

This is a linear decision rule in the pseudo-Euclidean space spanned over the given metric space of real-world objects

Eliminating the difference between linear and nonlinear decision rules in a pseudo-Euclidean space

Despite the fact that neither a metric space nor its pseudo-Euclidean embedding do not lend themselves to easy geometrical interpretation, we can conventionally demonstrate the essence of this boundary in an appropriate Euclidean space:

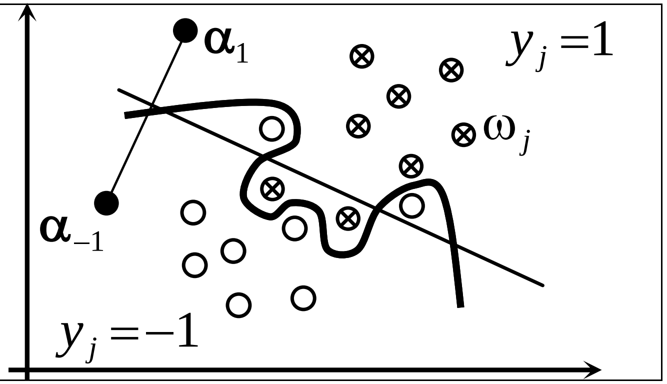


Linearly separable classes

Linearly inseparable classes

An urgent desire:

to eliminate the difference between separable and inseparable classes



Way-out: An extension of the well-known notion of potential functions from finite-dimensional linear spaces onto metric spaces.

Decision rule in an arbitrary metric space

The equivalent form of the discriminant function, which is completely determined by $N+1$ real numbers $(a_1,\dots,a_N,b)$:

new object *new object*

$$d(\omega \mid \underbrace{a_1,\dots,a_N,b}_{\text{Parameters to be estimated}}) = \frac{1}{2}\left[\sum\nolimits_{j=1}^{N} a_j\left(-\rho^2(\omega_j\,,\,\omega)\right)+b\right] \gtrless 0,$$

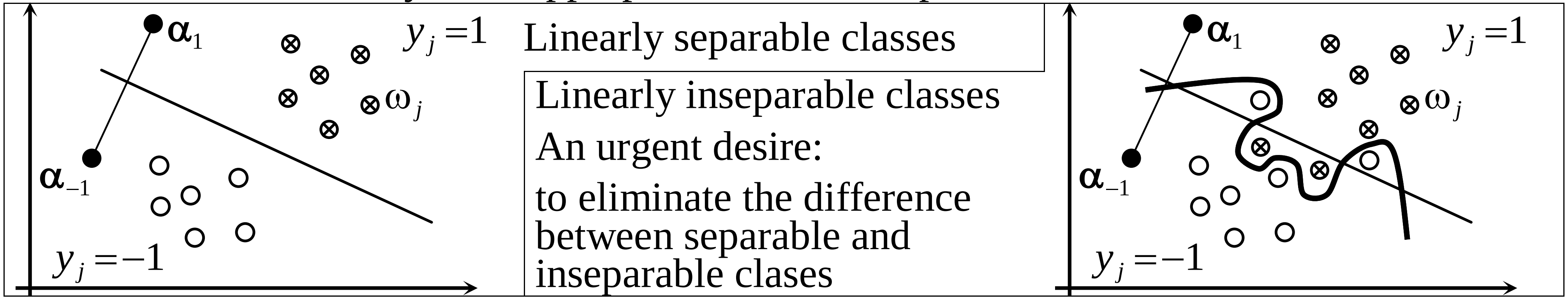
objects of the training set

$$\sum\nolimits_{j=1}^{N} a_j = 0,\quad \frac{1}{2}\sum\nolimits_{j=1}^{N}\sum\nolimits_{l=1}^{N}\left(-\rho^2(\omega_j,\omega_l)\right)a_j a_l > 0.$$

This is a linear decision rule in the pseudo-Euclidean space spanned over the given metric space of real-world objects

Eliminating the difference between linear and nonlinear decision rules in a pseudo-Euclidean space

Despite the fact that neither a metric space nor its pseudo-Euclidean embedding do not lend themselves to easy geometrical interpretation, we can conventionally demonstrate the essence of this boundary in an appropriate Euclidean space:



Linearly separable classes

Linearly inseparable classes

An urgent desire:

to eliminate the difference between separable and inseparable clases

M. Aizerman, E. Braverman, L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control, 1964, Vol. 25.

Potential function (kernel) in a finite-dimensional linear space

The initial inner product: $K(\mathbf{x}',\mathbf{x}'') = \mathbf{x}'^T\mathbf{x}'', \quad \mathbf{x}',\mathbf{x}'' \in \mathbb{R}^n$

The initial Euclidean metric: $\rho(\mathbf{x}',\mathbf{x}'') = \left[K(\mathbf{x}',\mathbf{x}') + K(\mathbf{x}'',\mathbf{x}'') - 2K(\mathbf{x}',\mathbf{x}'')\right]^{1/2} = \left[(\mathbf{x}'-\mathbf{x}'')^T(\mathbf{x}'-\mathbf{x}'')\right]^{1/2}$

Hilbert-space embedding –
a new inner product: $\tilde{K}(\mathbf{x}',\mathbf{x}'') = \exp\left(-\beta\rho^2(\mathbf{x}',\mathbf{x}'')\right)$ – radial potential function

The new metric
in the Hilbert space:

$$\tilde{\rho}(\mathbf{x}',\mathbf{x}'') = \left[\tilde{K}(\mathbf{x}',\mathbf{x}' \mid \beta) + \tilde{K}(\mathbf{x}'',\mathbf{x}'') - 2\tilde{K}(\mathbf{x}',\mathbf{x}'')\right]^{1/2} = \sqrt{2}\left[1 - \exp\left(-\beta\rho^2(\mathbf{x}',\mathbf{x}'')\right)\right]^{1/2}$$

However: $\tilde{\rho}(\mathbf{x}',\mathbf{x}'') \to 0$ when $\beta \to 0$

Normalization:

$$\boxed{\tilde{\rho}(\mathbf{x}',\mathbf{x}'') = \frac{1}{\sqrt{\beta}}\left[1 - \exp\left(-\beta\rho^2(\mathbf{x}',\mathbf{x}'')\right)\right]^{1/2}}$$

The kernel trick:
A linear function in the Hilbert space will produce a nonlinear one in the original space.

The idea: To apply such a transformation to an arbitrary metric.

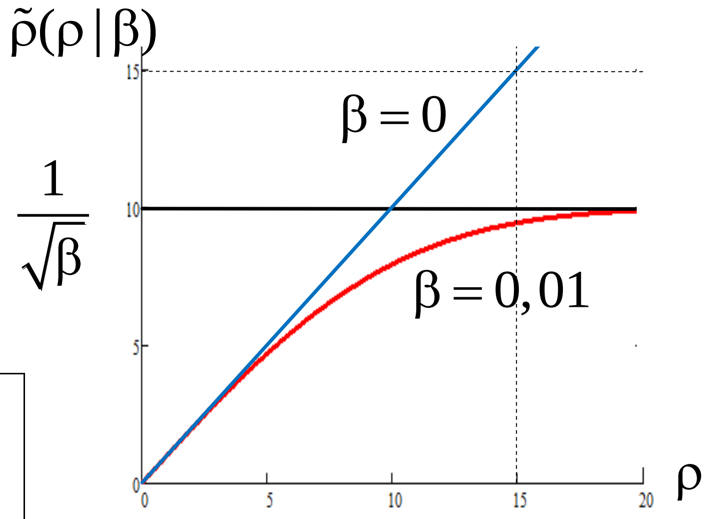
Metric decision functions of growing complexity

The separating ability depends on the accepted metric $\rho(\omega',\omega'')$

$$d(\omega\,|\,a_1,\dots,a_N,b)=\sum\nolimits_{j=1}^{N}\big(-\rho^2(\omega_j,\omega)\big)a_j+b\gtrless 0,$$
$$\sum\nolimits_{j=1}^{N}a_j=0$$

A parametric family of metric transformations, which guarantees improving the separating ability of the initial metric (family of saturable metrics):

$$\tilde\rho(\omega',\omega''\,|\,\beta)=\frac{1}{\sqrt{\beta}}\Big[1-\exp\big(-\beta\rho^2(\omega',\omega'')\big)\Big]^{1/2}$$



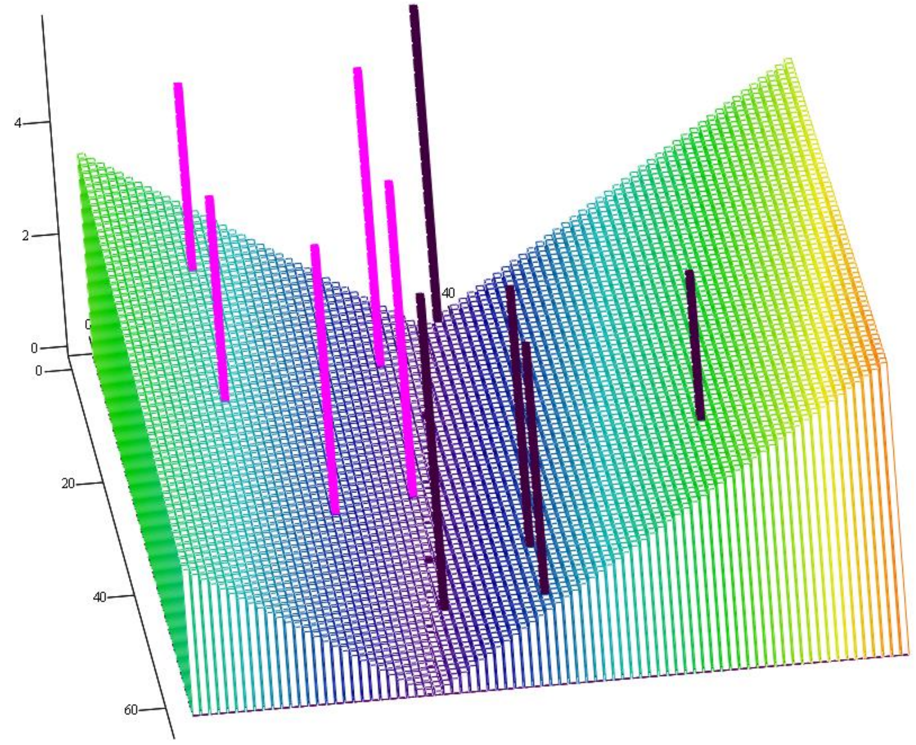
Theorem 7. *If* $\rho(\omega',\omega'')$ *is a metric on* Ω *then* $\tilde\rho(\omega',\omega''\,|\,\beta)$ *is a metric, too, with any value of the parameter* $\beta\ge 0$, *and* $\tilde\rho(\omega',\omega''\,|\,\beta)\to\rho(\omega',\omega'')$ *when* $\beta\to 0$.

Theorem 8. *Any training set* $\{(\omega_j,y_j=\pm1),\ j=1,\dots,N\}$ *with any metric* $\rho(\omega',\omega'')$ *is separable in the metric space* $\tilde\rho(\omega',\omega''\,|\,\beta)$ *if the parameter* $\beta>0$ *is large enough.*

The choice of the metric transformation parameter β is analogous to the choice of the parameter of a radial potential function in a finite-dimensional feature space:

Metric decision functions of growing complexity. Illustration

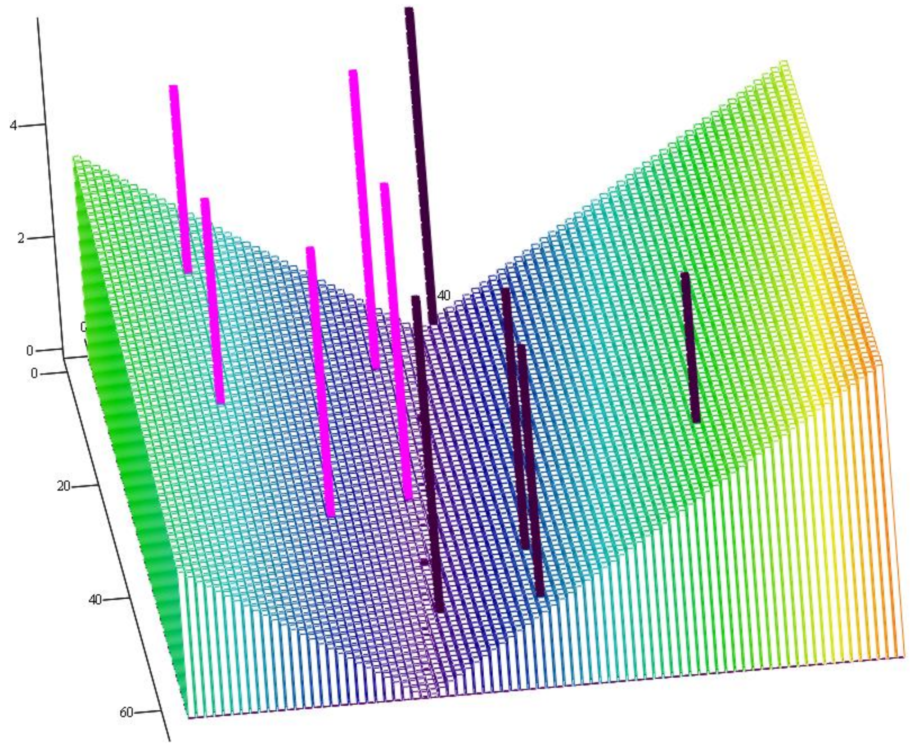
Separable case:



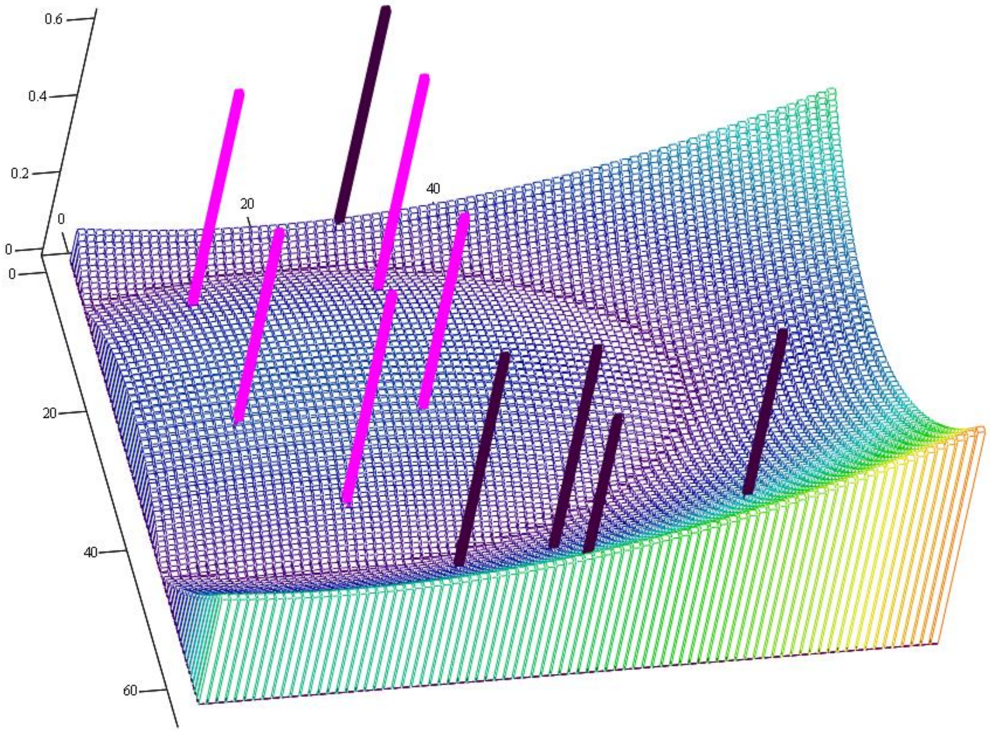
Non-Separable case:
Using of metric transformations

Metric decision functions of growing complexity. Illustration

Separable case:



Non-Separable case:
Using of metric transformations



Thank you for your attention!