THEORETICAL CONSIDERATIONS REGARDING THE USE OF „MACHINE LEARNING". CONCEPTS AND PARADIGMS FOR ADVANCED TECHNOLOGY ORGANIZATIONS

Marian Sorin IONESCU
Politehnica University of Bucharest
313 Splaiul Independenţei, Bucharest, Romania
marian.ionescu@man.ase.ro

Abstract. *The highly complex phenomenology approached by modern business models makes them difficult to understand, model and solve, without interdisciplinary tools, taken from the adjacent scientific field of competence, such as „Artificial Intelligence". The elaboration of mathematical models for the study, understanding and solving of the theoretical issues of operational-strategic applied business models is the most important paradigm of the scientific thinking that makes the transition to the economic performance generated based on the rational decisions with an optimal degree of efficiency developed by the top organizational management. Artificial Intelligence is the scientific branch that provides an infinite horizon for building algorithms, procedures and mathematical models of computing, with wide applicability in business and economic issues.*

Keywords: *Artificial Intelligence; Machine Learning; Data Mining;, Support Vector Machine.*

Introduction

The choice of this study is the "Machine Learning" area applied to the operational-strategic business models, with the notions, concepts and paradigms specific to the field. This entity of Artificial Intelligence knows the greatest development in computer sciences applicable to a variety of challenges coming from the economic macroenvironment. Along with "Machine Learning", the study of economic processes specific to advanced technology organizations has seen a growing development with the help of disciplines underlying Artificial Intelligence, such as „Data Mining". This has stimulated the development of related fields such as "Pattern Recognition" (recognition of forms, patterns) and "Mathematical Statistics", modern, flexible, easily transposable, software-deployable tools used to develop successful business models involved in the. processes of generating superior organizational added value.

The fundamental „target" of DATA MINING, inductive inference, „forecasting" based on operational examples is approachable through a multitude of different paradigms and techniques. Some of these raise major operational difficulties within the implementation processes, others are systemic processes structured on the classical cybernetic "black box" concept. A punctual but generic, powerful and with a high-efficiency level case is that of the "Support Vector Machine" (S.V.M.) which has proven to be relatively simple to implement within the advanced economic technology organizations applications.

The results of the activity of the economic organizations who have chosen this approach being achievements from good to excellent, compared to other approaches of the same type. The concept is introduced by Cortes & Vapnik, 1995, Boser, Guyon and Vapnik

1992. Support Vector Machine is a well-argued mathematically and complexly classification technology, applications to advanced economic technology organizations are favored by numerical data to be processed, interpreted and analyzed from the organizational economic performance perspective.

The technique underlying this type of approach is to identify a separating line, called the "maximum margin hyperplane", this leads to a process of separation between current data. The process does not always have a high degree of relevance, the fundamental use in artificial intelligence processes subsequently translated for economics models, actively and directly participate in the transformation of data into a different space where the dividing line is rigorously defined.

Designed for operationality in classification issues that occur in high technology process management, characterized by a high degree of complexity, subsequently paradigms of approach are extrapolated to regression tasks, the classification process is based on identifying the location to the division line where the new data is recorded.

Support Vector Machines are, for example, usable in data with a high degree of complexity, recognition of internal economic organization documents, facial recognition for the Human Resources Department, image classification. Precision level in such applications exceeding that of classical neural networks used in processes modeled with the help of Artificial Intelligence. S.V.M. presents economically measurable technological performances at the level of complexity of choosing features, algorithms for the extraction of specific features are identified in their system structure, a technology for identifying the proportion of white pixels in each internal document of an advanced economic technology organization is developed.

Among the advantages of the operational applications of S.V.M. within the used business models, it is necessary to specify the higher speed of classification of the new data, therefore the approach is operational with performant results even for a mix of data categories, the size of S.V.M. is from robust to very high. Comparing Support Vector Machines with Bayesian classifiers and decision trees, S.V.M. raises the issue of implementation complexity. It is possible that the classification process conducted with a S.V.M to contain a relevant degree of performance, but simultaneously it has a transparent degree of data for the low human decident factor. The amount of information provided to the economic organization in these cases is large enough in order for the approach to be operational-functional. Other types of approaches used in the specialized scientific literature are graphs and decision trees, competitors of the Support Vector Machines that provide much more efficient results in operational cases with low data volume.

Another vulnerability identified in the present study to address the issue of advanced technology organizations with S.V.M. is the lack of on-line access, the update process is re-initialized each time new training data appears. However, addressing processes specific to business models requires a high degree of understanding of algorithms and mathematical methods from top management decision makers. By identifying the business processes specific to the technological economic organizations as having the availability implementation and transposition into operation of some learning predictors in large spaces generated by the economic macroenvironment. We consider that it is necessary the position within the support vector machine paradigm (S.V.M.)

(see several studies such as Shalev-Shwartz, 2007; Shalev-Shwartz, Shamir & Sridharan, 2010; Shalev-Shwartz & Ben-David, 2017).

In the "Machine Learning" approach to business models, support vector machines, S.V.M., are supervised learning models with associated learning algorithms, thus analyzing the data used in the analysis and classification processes of the involved economic entities. An S.V.M. trainer algorithm type is built based on a model that assigns new operational case studies to a particular type of technology economic organization, transforming it into a non-probabilistic linear classifier. S.V.M. model types are a formalism as marks (points) in space (the external business environment), the examples of the categories are separated through as large as mathematically possible spread.

<table>
<tr><td>Image,
Identification</td><td rowspan="4">THE EFFECTIVENESS
OF APPLYING
THE SUPPORT
VECTOR
MACHINE</td><td>Analyzing Economic
Processes Making
Them More Efficient</td></tr>
<tr><td>Classic Text
Hyper Text
Categorization</td><td>Digital
Signature
Recogntion</td></tr>
<tr><td>Image Hierarchy</td><td>Systemic Generalized
System Prognosis</td></tr>
<tr><td>Computerized
Operational
Processes</td><td>Macro
Business
Environment</td></tr>
</table>

Extrapolating the linear classification phenomenon, S.V.M., is the non-linear classifier of phenomenology of business processes specific to the technological economic organizations, the use of algorithms and kernel procedures mapping the inputs into large, possibly infinite spaces.

In the operational practice of the technological economic organizations there is the possibility that the data is not labeled, the supervised learning process becomes very difficult to implement, the correct approach for these cases is unsupervised learning, the used cognitive technology searches for the natural grouping of the data to the groups, the new economic data used are grouped into the already formed entities.

For this type of algorithm, we have to mention the Siegelmann and Vapnik support vector grouping algorithm, it applies the support vector statistics in the support vector machine algorithm for the classification of the non-labelable data used by the technological economic organization, the algorithm is one of grouping with the widest spread within the organizations economic actors in the IT environment and the modern industrial one.

The dimensionality of the characteristic (proper) business processes space is determined both by the complexity of the economic entities used (data, samples), but also by the complexity challenges of specific computational processes, specific to the S.V.M. algorithm. is to address the complexity of economic evidence by identifying as large as possible margin separators.

The macroeconomic space in which the technological business phenomenon is operational is identifiable by about half of the surface as a possible set of training.

Learning and testing with a larger dimensional margin, the concrete cases of used technological business models being positioned not only on the right side of the hyperplane, but on the other one too, we observe that in a marginalization process, limiting the algorithm generating a large margin separator.

It is possible to develop a representative sample of relatively small complexity, spatial characteristic being large, with the possibility of development to infinity. In the operational applications of business models used by technology economic organizations, even in the case of function space with increased dimensionality. The complexity is diminishable by limiting, restricting the outputs generating algorithm for larger margin dividers. Thus, we introduce the concept of margin, which it is reported to the regularized paradigm that aims to minimize the losses, but also to the convergence rate of the Perceptron algorithm for such economic processes. Challenges coming from specialized markets, components of a macroeconomic structure that is undergoing globalization and a growing dynamic development, demand a computing power and data storage within the economic business technology organizations.

The fundamental idea pursued by top decision makers in close collaboration with the managers of research development departments, is that of transforming algorithms and data use technologies into a direct economic advantage, quantified in higher rates of earned profits.

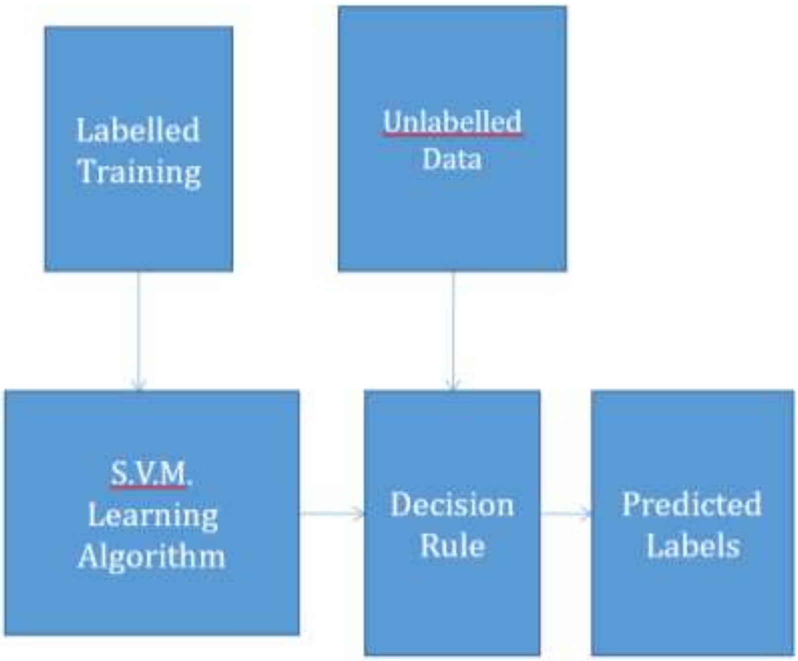

Figure 1. Transforming algorithms, use technologies into economic advantage

The formally chosen operational developed model has the following structure:

Domain Multitude: Let an arbitrary set, X, this is a collection of objects that are desired for a labeling process, the range of points is represented by an image vector, otherwise it is represented by value points, and X is a space of these values.

Label set: Labels set is restricted to two numeric values, {± 1}, and Y is the set of possible labels.

The learning data: $S = ((x_1, y_1), \dots, (x_m, y_m))$ is a finite sequence of pairs of $X \times Y$, a sequence of the labeled point range, is the input level at which the person that learns, inures, has access.

Learning outcomes: In the end of the learning process, a "predictive rule" must be issued, $h: X \rightarrow Y$, also known as a "predictor" or "hypothesis,"this is used for predicting point labels in the new defined domain.

Simplified Data Generation Model: The values generated with a certain distribution probability are acquired (assumed), representing the business macroenviroment. We introduce the probability distribution on the X set of D, within the learning process everything is known about the distribution, we consider that there is a correct labeling function $f: X \rightarrow Y$, with the property $y_i = f(x_i), \forall i$, each pair used in the learning process from the data set S is generated through a sampling process started in x_i in accordance with D labeled with the help of f.

Quantification of success: We introduce the notion of " classification error" as the probability of not correctly predicting the label for points representing random data generated by the aforementioned underlying distribution.
Having a subset of domain, A included in X, the condition is that A belongs to a σ- sub-algebra of X on the D -domain of definition, with the probability of distribution, assigning a numerical value $D(A)$, representing the probability of observing a point $x \in A$. A is perceived as an event and expressed using a function $\pi: X \rightarrow \{0, 1\}$, with $A = \{x \in X: \pi(x) = 1\}$, using the notation $P_{X \sim D}[\pi(x)]$ to express $D(A)$.
The predictive rule of error, $h: X \rightarrow Y$,, is thus defined by the formalism

$$L_{D,f}(h) = P_{x \sim D}[h(x) <> f(x)] = D(\{x: h(x) <> f(x)\}).$$

This type of formalism is also identifiable with the notion of "generalized error", the risk or "true error of h".

Information valid for the learning process: Within the learning process, there is a null perceptive about the distribution of D throughout its entire course and with the labeling function f. We identify the only way through which the “learner” can interact with the environment is to observe the multitude of forming, learning, training.

The Hard - S.V.M. concept, its application in business models

We consider that at the level of the operational activity of the technological economic organization, $S = (x_1, y_1), \dots, (x_m, y_m)$ is a set of concrete cases in the business environment, for each $x_i \in R^d$ and $y_i \in \{\pm 1\}$.
We note that this multitude of concrete cases of training, learning are linearly separable under the condition that there is half the space, (w, b) such that $y_i = sign((w, x_i) + b$, for any i the condition can be presented otherwise $\forall i \in [m], y_i\big((w, x_i) + b\big) > 0$.
The totality of the space halves (w, b) satisfying these conditions are submissive to E.R.M. hypotheses, if "$0 - 1$”, the error is zero, this represents the minimum possible error. The concept of „Hard - S.V.M.” is the learning rule in which a hyperplan of E.R.M. which separates the learning set of the operationalized business process with the greatest possible margin, for a structuring and defining process it is necessary to express

the distance between an arbitrary x point of the hyperplan using the semispace definition parameters.
Thus, a distance between a point x and the hyperplane defined by $(w, b)$where $||w||$ is $|(w, x) + b|$.
HARD - S.V.M. Rule is formulated as a square optimization problem. We identify the objective as the elaboration of an optimization process, the objective is a square convex function subject to constraints in the form of linear inequalities. We propose below an operational software structure for the conceptual HARD - S.V.M., previously presented.

$$HARD - SVM$$
$$input: (x_1, y_1), \dots, (x_m, y_m)$$
$$calculated: (w_{0,} b_0) = argmin_{(w,b)}||w||^2, s.t. \forall i, y_i((w, x_i) + b) \geq 1$$
(1)
$$output: w^{\wedge} = \frac{w_0}{||w_0||}, b^{\wedge} = \frac{b_0}{||w_0||}$$

(Shai Shalev-Shwartz and Shai Ben-David, Understanding MACHINE LERARNING)
It is shown that the output of HARD - S.V.M. is the separated hyperplan with the largest margin, we notice that the outputs of HARD - S.V.M. are solutions of the equation: $argmax_{(w,b):||w|=1} min_{i\in[m]} y_i((w, x_i) + b)$, which under the constraints$\forall i, y_i((w, x_i) + b) > 0$, represents the HARD - SVM Rule.
It should be remembered for the accuracy of the discussion the homogeneous case and the complexity of the pattern, sample, HARD - S.V.M.
In the case of homogeneity, the choice made reflects more flexibility in the case of homogeneous halves, the spatial halves that contain the origin, by definition the term with $sign((w, x))$, the term "bias" b is set to (chosen) with a "0 " value , and the HARD – S.V.M. structure for half of homogeneous spaces satisfies the relationship:

$$min_w||w||^2, s.c. y_i(w, x_i) \geq 1 \qquad (2)$$

In the case of regularizing the bias term, the resulting equation is a possible optimization of the economic processes specific to the studied advanced technology organization. If the bias term is not regularized, the optimization problem, (1) is not solved even if half of homogeneous learning space, R^{d+1}, is used, in the case of the complexity of the sample HARD - S.V.M., it should be recalled that the V.C. - dimensionality within the space half belonging to $R^d$is $(d + 1)$.

We observe that the fundamental learning theorem specifies if there is a significant number of examples smaller than $\frac{d}{\epsilon}$, in this case there is no ϵ accuracy learning algorithm in the half of the allocated space, we introduce the following definition:

Let D be a distribution process placed on $R^d \times \{\pm 1\}$, we assert that$D$ admits a margin separation process (γ, ρ) if there exists (w^*, b^*), astfel incat $||w^x|| = 1$, such that with probability 1, along the choice $(x, y) \sim D$, there is the relationship:
$y((w^*, x) + b^*) \geq \gamma$ si $||x|| \leq \gamma$.

Analogously, we assert that D is separable by the margin (γ, ρ) using half the homogeneous space having the form $(w^*, 0)$.

Observation: Considering D the distribution for $R^d \times \{\pm 1\}$ which is satisfying the separability condition (γ, ρ) with the marginal hypothesis using half of the homogeneous space. With a lower probability of $(1-\delta)$ for choosing the learning set, training, with size m, the output quantifiable error for HARD - S.V.M. is at most, $\sqrt{\frac{4(\rho/\gamma)^2}{m}} + \sqrt{\frac{2\log\left(\frac{2}{\delta}\right)}{m}}$.

We observe that for HARD – S.V.M. expression we assume that the set of learning-training is linearly separable. We present below a simplification of the HARD – S.V.M. exigencies that are applicable also in cases where the learning-training set is not linearly separable, thus, it is possible an exhaustive coverage of the totality of cases.Both cases, linearly separable or non-linearly separable, are identifiable within the operationalities and strategic developments specific to advanced economic technology organizations.

$$\begin{gathered} SOFT - SVM \\ input\ (x_1, y_1), \ldots, (x_m, y_m) \\ parameter\ \lambda > 0 \\ solution: min_{w,b,\xi} \lambda ||w||^2 + \frac{1}{m} \sum_{i=1}^{m} \xi_i) \\ s.c. \forall i, y_i((w, x_i) + b \geq 1 - \xi_{i,}\ \xi_i \geq 0 \\ output: w, b \end{gathered} \tag{3}$$

(Shalev-Shwartz & Ben-David, 2011, p.171)

Taking into account the regularized problem of minimizing losses, we assert that (3) is equivalent to the mathematical formalism:

$$min_{w,b}(\lambda ||w||^2 + L_S^{hinge\ (pivoting)}(w, b)) \tag{4}$$

Where $l^{hinge\ (pivotant)}\big((w, b), (x, y)\big) = \max\{0, 1 - y((w, x) + b\}$, represents by definition the losses generated by "hinge (pivoting)".

It is more convenient to consider SOFT - S.V.M. specific to the learning processes for half space, where the bias term b is set to zero, the optimization problem is formulated as follows:

$$min_w(\lambda ||w||^2 + L_s^{hinge\ (pivot)}(w)) \tag{5}$$

where:

$$L_s^{hinge(pivot)}(w)) = \frac{1}{m} \sum_{i=1}^{m} max\{0, 1 - y(w, x_i)\}$$

Support vectors, obtaining conditions of optimality

The name of S.V.M., "Support Vector Machine", is induced by the HARD - S.V.M. solution, w_0 has sustainability generated by operational cases located at a distance of $\frac{1}{||w_0|}$ from the separation hyperplan, the vectors generated are referred to as " Support Vectors ". Economic applications specific to advanced technology economic organizations are structured on Fritz John's optimality conditions. The conditions are presented below.

Sentence: We consider w_0 defined as in relation (2), let $I = \{i: |w_0, x_i| = 1\}$, there are coefficients $\alpha_1, \dots, \alpha_m$, such that that $w_0 = \sum_{i \in I} \alpha_i x_i$, Support vectors are identifiable in structures $\{x_i: i \in I\}$.

Fritz John's Observation: Suppose $w^* \in argmin_w f(w), s.c. \forall i \in [m], g_i(w) \leq 0$, where $f, g_1, \dots, g_m$ are differentiable, then there exists $\alpha \in R^m$, such that $\nabla f(w^*) + \sum_{i \in I} \alpha_i \nabla g_i(w^*) = 0$, where $I = \{i: g_i(w^*) = 0\}$.

The Stochastic Gradient Descendent Concept (S.G.D.) for minimizing the risk function

We define the operational-functional risk function that occurs during learning processes specific to economic processes, as being expressed by mathematical formalism:
$L_D(w) = E_{z \sim D}[l(w, z)]$,
We observe that the empirical risk minimization method, $L_s(w)$ as an estimate of the minimization $L_D(w)$. We propose the following operational software structure:

$$\begin{gathered}
The\ Stochastic\ Gradient\ Descendent\ , S.G.D., for\ minimizing\ L_D(w) \\
parameters: scalar\ \eta > 0, intreg\ T > 0 \\
initialization: w^{(1)} = 0 \\
for\ t = 1, 2, \dots, T \\
(sample) z \sim D \\
(pick)\ v_t \in \delta l(w^t, z) \\
(update)\ w^{(t+1)} = w^t - \eta v_t \\
output w^{-} = \frac{1}{T} \sum_{t=1}^{T} w^t
\end{gathered}$$

(Shalev-Shwartz & Ben-David, 2011, p.177)

Implementation of SOFT - S.V.M. using the S.G.D.

In the operational practice of the technological economic organizations solving the optimization problem for SOFT-S.V.M becomes a central problem., this admits a representation in the form of the mathematical expression:

$$\boldsymbol{min_w\left(\frac{\lambda}{2}||w||^2 + \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y(w, x_i)\}\right)} \qquad \textbf{(6)}$$

We propose for the regularized problems of minimizing losses a S.G.D. frame implemented software as below.
For loss of "hinge" applications we choose v_j as being 0 if $y(w^{(j)}, x) \geq 1$ and $v_j = --yx$ otherwise, noting $\theta^{(t)} = \sum_{j<t} v_j$.

$$\begin{gathered}
S.G.D. for\ solving\ SOFT - SVM \\
goal: resolving\ the\ optimization\ equation\ SOFT - SVM \\
parameter: T \\
initialize: \theta^{(1)} = 0 \\
for\ t = 1, \dots, T \\
let\ w^t = \frac{1}{\lambda t} \theta^{(t)} \\
we\ choose\ i\ randomly\ uniformly\ from\ [m] \\
if\ (y_i \left(w^{(t)}, x_i\right) < 1 \\
set\ \theta^{(t+1)} = \theta^{(t)} + y_i x_i \\
else \\
set\ \theta^{(t+1)} = \theta^{(t)}
\end{gathered}$$

$$output: w^{-} = \frac{1}{T}\sum_{t=1}^{T} w^{(t)}$$

Approach with "Kernel Methods (cores)"

In computer-specific learning processes, "Kernel methods" are a set of algorithms for model analysis, we identify the S.V.M support vector machine. The model has the fundamental objective to identify and analyze the general typology of relationships, clusters, hierarchies, correlations, transformation into series of data. The algorithms of this type are explicitly transformed into vector representations that provide a presentation of the user-specific features. In the case of Kernel's Methods, it is required only one user-specified kernel, a similarity function on pairs of rough representation points. Under Kernel Methods, operation is allowed in a characteristic dimensional implied space, without developing a process of calculating data coordinates in that space. The operation is thus in terms of financially acceptable allocation, costs are greatly reduced, in scientific literature this approach is called the "kernel trick", kernel functions are introduced for sequential data, graphics, text, images, vectors.

Among the algorithms used in analyzing and solving business issues specific to advanced technology economic organizations, we recall S.V.M., GAUSS processes, the processes of the Principal Component Analysis (P.C.A.), the canonical correlation analysis, ridge regression, spectral grouping, linear adaptive filters. The totality of linear patterns is transformable into nonlinear models by applying the concepts and paradigms specific to the "kernel trick", replacing its features, predictors, with the kernel function.

We observe that the totality of kernel algorithms is based on convex optimization, statically well argued, their static properties are studied through static learning theory, a way of approach is the RADEMACHER complexity. The fundamentals of this type of approach are structured on the idea of incorporating data into a high dimensional characteristic space. A core („kernel") is a measuring type of the similarity between sequences, they are perceived as HILBERT's inner space products, the instant space is virtually incorporated.

We consider the field of data used by the advanced technology economic organization as being the real number axis, limited to a certain set of numeric values, $\{-10, \ldots, 0, \ldots, 10\}$, the labels for the total $|x| > 2$ are $+1$ and -1 otherwise. A learning process for half space, in its initial representation, is replaceable by using a function $\psi: R \to R^2$, having the formalism, $\psi(x) = (x, x^2)$.

We call the characteristic space of the function set ψ, half of the space is characterized by the relationality, $h(x) = sign\left(\left(w, \psi(x)\right) - b\right)$, where the corresponding numerical values are $w = (0, 1), b = 5$.

The approach model used, so the thinking paradigm within the "Kernel Methods" is developed on the following hubs:
-considering a lot of X domains, a learning task (objective), choosing mapping, $\psi: X \to F$, for a characteristic space F, operational from R^n, for certain n values.
-we consider a lot of labels, $S = (x_1, y_1), \ldots, (x_m, y_m)$, the images sequence is generated, $S^{\wedge} = (\psi(x_1), y_1), \ldots, (\psi(x_m), y_m)$
- is prepared, trained a linear type predictor, h during $S^{\wedge}$.
- the label of a test point, x, is predicted to be expressed by $h\left(\psi(x)\right)$.

Several studies address these aspects (Shalev-Shwartz, Shamir & Sridharan, 2010; Japkowicz & Shah, 2011).

The Kernel innovative approach

Specifically, to the economic activities, incorporating entry space into a large size space, make the learning process in half the space more efficient. The complexity of computational processes, a feature of advanced technology organizations, for learning activities represents a linearly efficient computerized data separator, usable for large volumes of information-data to process, induced processes requiring significant financial allocations.

The concept of „kernels" („cores") is used for a proper description of the interior products in the characteristic space, having a functional relationship ψ and an area of space X positioned in a HILBERT space is defined as the „kernel" („core") in the form of the relationality $K(x, x') = \left(\psi(x), \psi(x')\right)$, we observe K an argument expressing the similarity between sequences and the incorporation mode ψ as a mapping of the set domain X, positioned in a space where the two similarities are inner products (Shalev-Shwartz, Shamir & Sridharan, 2010). Many algorithms specific to learning half-space processes are made based on the value of the kernel-expressing function, the structuring is done on pairs of points in the field, we identify a main advantage of this type of algorithmic approach, the implementation of linear separators in large dimensionalities spaces , without specifying the points located in that space or the expression ψ incorporated into that space.

We note that the totality versions of the S.V.M. are specific situations of the general form:

$$\boldsymbol{min_w}\left(\boldsymbol{f}\left(\left(\boldsymbol{w}, \boldsymbol{\psi}(\boldsymbol{x_1})\right), \ldots, \left(\boldsymbol{w}, \boldsymbol{\psi}(\boldsymbol{x_m})\right)\right) + \boldsymbol{R}\left(\|\boldsymbol{w}\|\right)\right) \quad \textbf{(7)}$$

in which $f{:}\, R^n{:} \to R$ is an arbitrary function and $R{:}\, R_+ \to R$ is an non-decreasing monotonous function.
In the operational economic applications, it is necessary to substantiate the decisions with the help of a sentences of fundamental truth, "Theorem of Representation".

The Representation Theorem: Assuming that ψ is a („mapping") a lui X in a HILBERT space, then there exists a vector $\alpha \in R^m$, such that $w = \sum_{i=1}^{m} \alpha_i \psi(x_i)$, which represents an optimal solution for the equation (7).

Specifying the previous presentation of SOFT - S.V.M., as in equation (5), the problem is reformulated as:

$$\boldsymbol{min_{\alpha \in R^m}}\left(\boldsymbol{\lambda \alpha^T G \alpha} + \frac{\mathbf{1}}{\boldsymbol{m}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{m}} \mathbf{max}\{\mathbf{0}, \mathbf{1} - \boldsymbol{y_i}(\boldsymbol{G\alpha})_{\boldsymbol{i}}\}\right) \quad \textbf{(8)}$$

where $(G\alpha)_i$ is the i^{st} element of the vector obtained by multiplying the GRAM matrix, G with the α vector, we notice that equation (8) is in the form of a square program, so an efficient solution exists.

The coefficients α are calculated for forecast, prognose for a new sequence, as follows:
$\left(w, \psi(x)\right) = \sum_{j=1}^{m} \alpha_j (\psi\left(x_j\right), \psi x)) = \sum_{j=1}^{m} \alpha_j K(x_j, x)$

In the case of the concrete operational situations of the advanced technology organizations, addressing the problems arising from the business environment with the help of the "kernel method" rather than inducing a directly process of optimization w in the characteristic space is supported by the fact that in certain concrete cases , the size of the characteristic space is extremely high on the time axis, the implementation of the "kernel method" is simple and extremely flexible.

The kernel perception as a way of expressing previous knowledge

A mapping of the ψ characteristics is perceived as an extrapolation of the linear class to a larger one, corresponding to the linear classifiers out of the characteristic spaces. We observe that the adequacy of any set of assumptions, hypotheses about the task of learning objectively depends entirely on the nature of this type of task. We develop a paradigm of thinking in which the incorporation of ψ is perceived as a way of expressing and using the knowledge available to the top organizational decision-maker regarding the challenge to be solved. There is a possibility that the learning task can be used to identify a sequence of characters, „signature", in a data file that made available to the top organizational decision maker, indicating a virus that alters the quality of the information or not.

Formally, we consider X a set of strings finite on a set of alphabet Σ, and let X_i t the set of all strings with a maximum length d, the set of assumptions, working hypotheses wanted to be learned is $H = \{h_v: v \in X_d\}$, where for a string $x \in X, h_v(x)$ is 1 if v is a substring of x si $h_v(x) = -1$, otherwise. We consider a mapping function ψ on the space R^s, where $s = |X_d|$, the coordinates of $\psi(x)$ on the space R ^ s, where s = | X_d |, the coordinates of ψ (x) correspond to a few strings v, they indicate if v is a substring of x, this means for any $\in X, \psi(x)$ is vector in $\{0, 1\}^{|X_d|}$.

Each member of the set H is achievable by composing some classifiers on $\psi(x)$, more than one half of the space, with the norm 1 and reaching the margin of 1.

For each $x \in X, \left||\psi(x)|\right| = O(\sqrt{d})$, the learning process is developed using S.V.M., a pattern sample of complexity showing that it is polynomial in d.

At the level of the applications specific to the advanced technology economic organizations, we find that the dimension of the characteristic space is exponential for d, the implementation by S.V.M. becoming a difficult process. We observe, however, that there is a computing feature of the inner product within the characteristic space, the kernel function, without an explicit cartographic presentation of the characteristic spatial sequences. We observe $K(x, x')$ as the number under the strings of x and x' easy to calculate for a corresponding time allocation for the polynomial d. It is thus demonstrated the mapping mode of the features, thus allowing spaces allocated for non-vectorial domains.

Kernel functions characteristics

The kernel matrix is a way of expressing the previously acquired knowledge.
We consider a similar function given by the form $K: X \times X \rightarrow R$, the problem that arises is the possibility of representing an inner product between $\psi(x)$ si ψ(x').

Observation: A symmetric function $K: X \times X \rightarrow R$ implements an inner product in certain HILBERT spaces if and only if it is positive semi-definite, called for the totality $x_1, \dots, x_m$, GRAM matrix, $G_{i,j} = K(x_i, x_j)$ is a positive semi-definite matrix.

Discussion: It is obvious that if at K level an inner product is implemented in a HILBERT space, it results that the GRAM matrix is positively semi-defined.
Extrapolating the discussion for the totality of possible operational situations, defining spatial functions over X as $R^X = \{f: X \rightarrow R\}$, for each $x \in X$, let $\psi(x)$ is the function $x \rightarrow K(.,x)$.
We define a vector space taking the totality of linear combinations of $K(.,x)$ elements, an inner product of this vector space admits a representation of the form;

$$(\sum_i \alpha_i K(.,x_i), \sum_j \beta_j K(.,x'_j)) = \sum_{i,j} \alpha_i \beta_j K(x_i, x'_j)$$

We affirm that this is a valid inner product as long as it is symmetric, property correlated with the symmetry of K, the product presented is linear and positively defined.
Result:

$$\big(\psi(x), \psi(x')\big) = \big(K(.,x), K(.,x)\big) = K(x, x')$$

Implementation of SOFT – S.V.M. with kernel, proposed model

Previously presented was the challenge of SOFT - S.V.M., the approach with the help of the "Kernel Method" is the subject of the scientific analysis developed further.
Equation (8) admits an algorithmic solution, but a simpler way to solve it is to address the problem of SOFT – S.V.M. optimization in spatial or characteristic,

$\boldsymbol{min_w(\frac{\lambda}{2}||w||^2 + \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y(w, \psi(x_i))\}}$,
(9)
using kernel ratings.

We note that vector $w^{(t)}$ introduced by S.G.D. procedure (Stochastic Gradients Descent) previously introduced is always in the linear span of $\{\psi(x_1), \dots, \psi(x_m)\}$, maintaining of $w^{(t)}$is equivalent to maintaining the correspondence of coefficients α.
We define $K(x,x') = \big(\psi(x), \psi(x')\big), \forall x, x'$, maintaining two vectors of R^m, corresponding to two vectors, $\theta^{(t)}$ and $w^{(t)}$ defined in the S.G.D. procedure $\beta^{(t)}$ is a vector such that

$$\boldsymbol{\theta^{(t)} = \sum_{j=1}^{m} \beta_j^{(t)} \psi(x_j)} \qquad \textbf{(10)}$$

si $\alpha^{(t)}$ is such that

$$\boldsymbol{w^{(t)} = \sum_{j=1}^{m} \alpha_j^{(t)} \psi(x_j)} \qquad \textbf{(11)}$$

The two vectors α, β are subject to an update process according to the proposed later procedure.

$$S.G.D. used\ to\ solve\ SOFT - S.V.M. with\ Kernel$$

$$Purpose: solution\ min_w(\frac{\lambda}{2}||w||^2 + \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y(w, \psi(x_i))\}$$

$$parameter : T$$

$$initilialization: \beta^{(1)} = 0$$

$$for\ t = 1, \dots, T$$

$$let\ \alpha^{(t)} = \frac{1}{\lambda t}\beta^{(t)}$$

$$let\ i\ uniformly\ randomly\ from\ [m]$$
$$for\ all\ j <> i, we\ choose\ \beta_j^{(t+1)} = \beta_j^{(t)}$$
$$if\ (y_i \sum_{j=1}^{m} \alpha_j^{(t)} K(x_j, x_i) < 1)$$
$$we\ choose\ \beta_i^{(t+1)} = \beta_i^{(t)} + y_i$$
$$otherwise$$
$$we\ choose\ \beta_i^{(t+1)} = \beta_i^{(t)}$$
$$(Output):\ w^{-} = \sum_{j=1}^{m} \alpha_j^{-} \psi(x_j),\ \alpha^{-} = \frac{1}{T} \sum_{t=1}^{T} \alpha^{(t)}$$

See several studies (Shalev-Shwartz, Shamir & Sridharan, 2010; Shalev-Shwartz & Ben-David, 2017).

It is demonstrable that the previous implementation is equivalent to the launching in execution of the S.G.D. for the characteristic space.
Considering w^{-} - the outputs of the S.G.D. described above, when applied on a spatial characteristic, let $w^{-} = \sum_{j=1}^{m} \alpha_j^{-} \psi(x_j)$ outputs after applying S.G.D. with kernels, then $w^{-} = w^{\wedge}$.
It is obvious that for $\forall t$, the result of running S.G.D. program, $\theta^{(t)}$ is held by equation (10) within the characteristic space.
By definition $\alpha^{(t)} = \frac{1}{\lambda t} \beta^{(t)}, w^{(t)} = \frac{1}{\lambda t} \theta^{(t)}$, the statement shows that $w^{(t)}$ is held by equation (11).
For $t = 1$ the statement is obvious, if $i \geq 1$, then:

$$y_i \left(w^{(t)}, \psi(x_i) \right) = y_i (\sum_j \alpha_j^{(t)} \psi(x_j), \psi(x_i)) = y_i (\sum_j \alpha_j^{(t)} K(x_j, x_i).$$

The condition of the two algorithms is equivalent to the updating of θ, it follows:

$$\theta^{(t+1)} = \theta^{(t)} + y_i \psi(x_i) = \sum_{j=1}^{m} \beta_j^{(t)} \psi(x_j) + y_i \psi(x_i) = \sum_{j=1}^{m} \beta_j^{(t+1)} \psi(x_j),$$

Which leads to the asserted conclusions.

Conclusions

The modern business environment contains economic organizations that operationalize and develop top technologies and even frontiers, the totality of which are faced with a strong need to increase computing power and storing used business data. From the perspective of the business models developed and implemented, it means that the vast majority of the economic organizations currently have very large size archives of data and information products, the human decident is most often overwhelmed by them. The challenge addressed to the top organizational decision-maker is to transform the huge volume of corporate data into an economically competitive advantage.

By translating the solving of some of the problems from the engineering sciences in the economic field, common applications are developed, from the anticipation of the products with a high degree of purchasing possibility by a client, the answers elaborated on the basis of questionnaires. Operational, engineering and economic reality imposed „Data Mining", a component of Artificial Intelligence, related field to „Machine Learning", Model Recognition, Mathematical Statistics, „Data Mining" performs a part of the suboptimal and random analyzes needed to develop the decision making process, regardless of its positioning within the structure of the studied advanced technology economic organization.

It is a unique possibility of strategic development of the new field, the arguments that offer this opportunity are the very large numbers of algorithmic approaches and algorithmic paradigms available. The fundamental task of „Data Mining" is that of inductive inference, prediction based on operational realities, approachable through a variety of different techniques. The degree of complexity of these paradigms requires human resourcing departments, involves high-skill entities, and their simplification often leads to "black box" approaches.

Vector Machine Support – S.V.M. is a method with a high degree of generality, efficiency, and powerful mathematics, generating results which at the outputs of the „output data" processes are evaluated from good to superlative, even optimal, compared to addressing issues through other methods. Understanding and transposing in operability of the concepts and paradigms of S.V.M., requires that there is a high standard of mathematical education at the level of top organizational management. The present study identifies and presents the relationship between S.V.M. and the scientific methods of analysis of the approached business issues, for the economic organizations using top I.T. and virtual technologies. Also described are the ways and algorithms needed to transpose in operability the S.V.M., the set of guidelines and the fundamental rules for the application of the S.V.M. concept, in all complexity and with all the advantages that result from it are specified. The theoretical and applicative considerations of the present scientific investigation, without claiming to be exhaustive, introduce the ideas, the methodology, as well as a succinct presentation of the specific software structures of S.V.M., intended for the applications. The general problem is that of induction, an important one that generates significant added value for organizational databases. The analysis and understanding of the existing correlations within these databases is a fundamental one with a high degree of complexity, flexible and relatively easy to implement operationally, the case of S.V.M. study. The S.V.M. approach is based and further developed on a simple and intuitive concept involving the separation of the two data classes in the first part using a linear function that shows the maximum distance of the used data.

The idea then becomes a powerful and efficient learning algorithm, when the problem of separability (marginal errors) and the implicit mapping of characteristic, more descriptive spaces by operationalization of the kernel functions are overcome. At the operational-strategic level, software packages are developed which allow for their implementation and launch in operational life, obtaining very good results with human effort and minimal organizational allocations. Taking into account the accelerated scientific progress, the complexity of the problems and challenges coming from the economic environment, as well as the competition of the economic organizations and the development of business models, well founded theoretically, we consider the field to be just in the beginning, major future developments becoming inevitable.

References

Anthony, M., & Barlet, P. (1999). *Neural Network Learning: Theoretical Foundations*, Cambridge University Press.

Barber, D. (2012). *Bayesian reasoning and machine learning,* Cambridge University Press.

Bishop, C.M. (2006). *Pattern recognition and machine learning*, Vol.1, New York, NY: Springer.

Bottou, L., & Cun, Y. L. (2004). Large scale online learning. In *Advances in neural information processing systems. Proceeding of the 2003 Conference* (pp. 217-224), Vol.16, MIT Press.

Cristianini, N., & Shaw-Taylor, J. (2000), *An introduction to support vector machines*, Cambridge University Press.

Cover, T., & Thomas, J. (1991). *Elements of information theory*, New York, NY: Wiley.

Floyd, S., & Warmuth, M. (1995). Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning,* 21(3), 269-304.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*, New York, NY: Springer.

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification. Retrieved from https://www.researchgate.net/profile/Chenghai_Yang/publication/272039161_Evaluating_unsupervised_and_supervised_image_classification_methods_for_mapping_cotton_root_rot/links/55f2c57408ae0960a3897985/Evaluating-unsupervised-and-supervised-image-classification-methods-for-mapping-cotton-root-rot.pdf.

Japkowicz, N., & Shah, M. (2011). *Evaluating Learning Algorithms:A Classification Perspective*, Cambridge University Press.

Jaynes, E.T. (2003). *Probability theory: the logic of science,* Cambridge University Press.

Jurafsky, D., & Martin, J.H. (2008). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, Prentice Hall.

Montgomery, D. (2012). *Design and Analysis of Experiments*, New York, NY: Wiley.

Rice, J.A. (2006). Mathematical Statistics and Data Analysis, Cengage Learning.

Shalev-Shwartz, S. (2007). *Online Leraning: Theory, Algorithms, and Applications*, PhD thesis, The Herbrew University.

Shalev-Shwartz, S., Shamir, O., & Sridharan, K. (2010). Learning kernel-based halfspaces with the zero-one loss, *COLT.* Retrieved from https://arxiv.org/pdf/1005.3681.pdf.

Shalev-Shwartz, S., & Ben-David (2017). *Understanding Machine Learning, from Theory to Algorithms*, Cambridge University Press.

Siddiqi, N. (2005). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring,* New York, NY: Wiley.

Stewart, J. (2012). *Calculus*, (7e ed.), Cengage Learning.

Tukey, J.W. (1977). *Exploratory Data Analysis*, Addison-Wesley.

Widdows, D. (2004). *Geometry and Meaning,* Stanford, CA: Center for the Study of Language and Information.