Support Vector Machines: A Tool for Pattern Recognition and Classification[1]

A.P. Tchangani

Dpt. GEII, IUT de Tarbes, Université Toulouse III - Paul Sabatier,

1 rue Lautréamont, BP 1624, 65016 Tarbes Cedex,

FRANCE

E-Mail: ayeley.tchangani@iut-tarbes.

Keywords: Pattern Recognition, Classification, Support Vector Machines, Supervised Learning, Mathematical (Quadratic) Programming.

Ayeley Philippe Tchangani received Ingénieur Degree from Ecole Centrale de Lille, Lille, France in 1995; an M.Sc. (DEA) and a Ph.D. (Doctorat) from Université des Sciences et Technologies de Lille, Lille, France in 1995, 1999 respectively in Control and Automation. From November 1999 to December 2000 he was a Postdoctoral Fellow at French -South Africa Technical Institute in Electronics, Technikon Pretoria, South Africa. Since February 2001 he has been with Université Toulouse III (IUT de Tarbes), Tarbes, France where he is currently an *Assistant Professor* (Maître de Conférences); he teaches Control Systems and Automation. He is also an *Assistant Research Professor* at Laboratoire Génie de Production (LGP), Ecole Nationale d'Ingénieurs de Tarbes (ENIT). Dr. Tchangani's research interests are in modeling, analyzing and deriving algorithms for control, decision making and evaluation of systems in uncertain and conflicting environment. He is a Member of IEEE and SIAM.

Abstract: The power of computation and large memory of computers nowadays offer a great opportunity for information processing and storage. But information is not knowledge and one needs methods that permit to go from information to knowledge. Extracting automatically knowledge from storage data becomes then one of great challenge for the Information Technology (IT) industry. Pattern Recognition (PR) is the study of how machines can observe the environment, learn to distinguish pattern of interest from their background and make sound and reasonable decisions about the category of the pattern. The automatic recognition, classification, description, grouping of pattern is an important problem in engineering and sciences such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, remote sensing, manufacturing, etc. Computer programs that help many professional in their daily work such as doctors diagnosing disease, policemen identifying suspects, engineers supervising manufacturing plants and energy production systems, etc. depend in some way on pattern recognition. One important field and goal of pattern recognition is classification: supervised or unsupervised also known as clustering. In this paper we present a mathematical tool named support vector machines (SVM) that permit to derive efficient algorithms of learning and classification.

1. Notations

The following notations will be used through this paper:

- $\Re$ denotes the set of real numbers.
- $\Re^d$ denotes the real space of dimension *d*.
- For a vector $x \in \Re^d$, x' denotes its transpose.
- For two vectors $x, y \in \Re^d$, $x'y$ is their inner product, that is $x'y = \sum_{i=1}^{d} x_i y_i$.
- $A \in \Re^{n \times m}$ denotes a *n* rows and *m* columns matrix with real entries.

2. Introduction

Pattern recognition is a domain of information processing that got a great interest since decades. This interest is mainly due to the increasing power of computers in terms of memory (storage of great amount of information) and rapidity (real time processing). A central problem one is concerned with in pattern recognition is assigning a given object to one of several possible categories. This process is known as discrimination and classification. The problem of (supervised) pattern recognition is then: one is given a

1 The preliminary version of this paper was written when the author was giving Pattern Recognition Module of Post Graduate Diploma in Electrical Engineering (PGDEE) at French - South African Technical Institute in Electronics (F'SATIE) in Pretoria. He was also supported during that time by South African National Research Foundation (NRF) as a Post Doctoral Fellow. The support of these institutions is greatly acknowledged.

certain number of categories C_i, $i = 1, 2, ..., n$ and the purpose is to classify each pattern x_j in one of them ($x_j \in C_j$) according to some rules. The *discrimination* refers to the process of deriving classification rules (learning) from sample classified objects and the *classification* refers to applying the rules to new objects of unknown class (recognition). Stated like this, it is immediately apparent that the range of applications of pattern recognition is very broad. In the following section some potential applications as well as their domain will be outlined.

2.1. Potential Area of Applications of Pattern Recognition

Pattern recognition techniques are widely used in various domains such as medicine (cancer diagnosis, homosexual and heterosexual detection based on the analysis of androsterone and etiocholanolone, etc.), security (human identification based on voice print, fingerprint, handwriting, etc.), agriculture (identification of crops using satellite image), emergency (in the case of natural disaster satellite image can be used to estimate the degree of infrastructures damage), military (identifying an enemy target or movement), manufacturing (assembly of product, machine failure detection), etc. Here is a non exhaustive list of potential applications.

2.1.1 Medicine (see [2], [7], [8])

- Chemotherapy diagnosis in breast cancer: the problem here is deciding if a breast cancer patient needs or not a chemotherapy. The discrimination is made on the basis of existence or not of metastasized lymph nodes removed from the patient's armpit during surgery. This studied is carried at University of Wisconsin (see also [7], [11], [12]).
- Classification of person in term of Normal, Hepatitis B, HIV Early, HIV Late, Leukemia, Malnutrition by using some measurements of blood [8].
- Classification of person in terms of heterosexual or homosexual by using measurements of their urinary androsterone and etiocholanolone [2] (see Example 1).

2.1.2. Human Identification by Biometry ([5], [16])

One or more of some characteristics like signature, fingerprint, voice print pattern for each individual in a particular organization can be stored in a machine and then the actual characteristic is compared to the stored pattern using some criteria in order to identify the present individual.

2.1.3. Agriculture ([4])

In some countries (countries of European Union for example) the production of some crops is regulated and the farmers receive some subvention according to the area used for this particular crop. The use of satellite image to identify each crop and estimate the area occupied by this crop is some time useful.

2.1.4. Emergency ([4])

The use of satellite image to identify the damage of infrastructure (roads, bridge, highway, etc.) during a natural disaster such as earthquake or inundation can help to organize efficiently assistance of population. This technique can also be used to estimate the progression of a forest fire for instance.

2.1.5. Military ([16])

Target identification or enemy movement detection are important parameters for military. The satellite images are often used by army to identify a target to be bombed, the movement of an enemy, etc. Another technique of target or enemy identification is the use of acoustic signal (sonar) for submarine for instance.

2.1.6. Manufacturing ([2], [17], [18])

Most of the time in the manufacturing industry, the assembly of products is made automatically by a machine which uses characteristics such color, form, weight, etc. to discriminate between products. Another use of pattern recognition technique in manufacturing is machine failure detection based on the analysis of vibration wave form, noise made by the machine, etc. See also [13] for other applications.

2.1.7. Finance ([14])

In the domain of finance, investors use to classify the risk of investment in a particular company according to some of its financial characteristics by studying the historical data (financial ratios) of the company and data coming from similar companies (see Example 2).

One can imagine many other applications that can use pattern recognition techniques. In the following paragraph we will briefly present the principles of pattern recognition.

2.2. Principles of Pattern Recognition

The classification of a pattern in term of categories is made by using some mathematical tools to derive classification algorithms. The main components of a pattern recognition system are shown on Figure 1; the role of each component is briefly explained below.

- **Sensors**: This component senses some information on the object to be classified. These information are then sent to feature extraction system. The sensor is to be understood in a large sense including the registration of financial ratios in a stock exchange market for instance.
- **Feature extraction system**: The task of this component is to extract from the information given by the sensor some relevant features or attributes that will be used to efficiently classify the object.
- **Classification system**: This is the decision unit that will decide on the basis of some criteria to which class the object under consideration belong to.

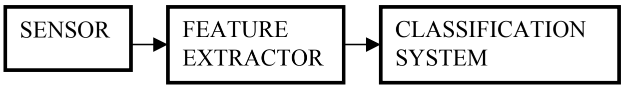


Figure 1: Main components of a Pattern Recognition System

The problem of pattern recognition or classification can be divided into two sub-problems: learning and testing (or classifying). For the learning process, a set T of pattern named training set is available and the purpose is to derive discrimination rules using some mathematical tools based on this training set. At this stage there are two possibilities:

- categories are known and one want to establish a correspondence between each pattern in the training set and the category set in order to minimize a certain cost (mis-classification error for instance); this is known as *supervised learning*;
- the underlying relationship and categories must be discovered by the learning methods; this is *unsupervised learning* or *clustering*.

Testing (or classification) consists in classifying a new pattern based on rules derived during learning process. This paper considers essentially the supervised learning case. The next section will give an overview of mathematical tools used for learning process.

3. Mathematical Tools for Pattern Recognition

Algorithms of learning and classification are derived by using some mathematical tools. In the following paragraphs we will present some of these tools.

3.1. Template Matching

The template matching is the determination of similarity between two entities (points, curves, shapes). The use of this technique in pattern recognition is performed as follow: a template or a prototype of the pattern to be recognized is available (stored). The pattern to be recognized is matched against the stored template taking into account all allowable pose (translation and rotation) and scales change. This technique is less used in practice because it is time consuming.

3. 2. Statistical Methods

Statistical methods are probably the eldest methods used in pattern recognition. The pattern is represented by a vector of features or measurements and the purpose is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint region in the space of features. The surfaces separating disjoint compact regions are called decision boundaries. Most of the time decision boundaries are hyper planes [2], [14]. Another technique that falls in the framework of statistical learning is the Bayesian technique and Bayesian networks where decision is made based on posterior (after observing features) probability that the given pattern belongs to a particular category; many algorithms based on this technique are available in the literature (see [9] for instance and references therein).

3.3. Neural Networks

Neural networks are massively parallel computing system consisting of extremely large number of simple processors (neurons) with many interconnections. They attempt to simulate the way the human brain works. They are shown to be universal approximators of nonlinear functions and so they are capable to learn very complex nonlinear input-output relationship (see for example [3]).

Another promising tool, that we will consider in the remaining of this paper is support vector machines. This method is born from the work by Vapnik [19] (see [15] for a presentation and references therein). This tool has some strong advantages as non necessity of normalizing data like in neural networks case or non perturbation by data permutation [15] and the algorithms are based on mathematical programming. Its success in practical applications leads to a growing litterature, see for instance [1], [10], [17].

4. Support Vector Machines: Presentation

This technique is a new technique and is an object of intense researches at this moment. It has shown some useful applications as in the case of breast cancer treatment [11], [12]. It can learn linear and nonlinear relationships. Though it can be used in the case of many categories, we will consider only the case of two categories represented by sets *A* and *B* that are regions of $\Re^d$ and supervised learning case. The problem then consists in classifying *N* pattern represented by their features vectors $x_i \in \Re^d$ between *A* and *B*. The sets *A* and *B* can be linearly separable or non linearly separable. *A* and *B* are said to be linearly separable if there exists an hyper plane *H* defined by

$$H = \{x \in \Re^d : \omega' x + b = 0\}$$

where $\omega \in \Re^d$ is a given vector and *b* is a scalar such that all points of *A* lay on one side of *H* and all points of *B* on the other side. Mathematically it means that

$$x_i \in A \;\; if \;\; \omega' x_i + b > 0 \;\; and \;\; x_i \in B \;\; if \;\; \omega' x_i + b < 0.$$

A and *B* are said to be nonlinearly separable if the previous property does not hold. These two cases will be considered in the following paragraphs.

4.1. Linearly Separable Case

In the case of two linearly separable sets, the support vector machines learning process is stated as: find $\omega \in \Re^d$ and $b \in \Re$ such that:

$$\begin{aligned} \omega' x_i + b &\geq \;\; 1 \;\; for \;\; x_i \in A, \\ \omega' x_i + b &\leq -1 \;\; for \;\; x_i \in B. \end{aligned} \tag{1}$$

The equalities2 $\omega' x_i + b = 1 \; and \; \omega' x_i + b = -1$ define two hyper planes H_1 and H_2 respectively and there is no elements laying between these hyper planes. The objective then is to find hyper planes that maximize the length of the layer between them. A geometrical consideration proves that the distance between H_1 and H_2 is $\frac{2}{\|\omega\|}$. Maximizing this distance is the same as minimizing $\frac{1}{2}\|\omega\|^2$. The whole process is reduced then to the quadratic programming problem:

$$\min_{\omega, b} \left\{ \frac{1}{2}\|\omega\|^2 \right\},$$
$$s.t. \quad \begin{cases} \omega' x_i + b \geq 1, \; for \; x_i \in A, \\ \omega' x_i + b \leq -1, \; for \; x_i \in B. \end{cases} \tag{2}$$

Though this problem can be solved directly by mathematical programming solvers such as OSL, CPLEX, CONOPT, MINOS and related modelling software GAMS, AMPL or Matlab, it is shown that more efficient algorithms can be derived based on the particularities of the problem. It is shown by using Lagrange multipliers that the important points are those which fall on the hyper planes H_1 and H_2 known as the support vectors (see Figure 2). And so the algorithms of learning (see [11], [12], [15]) are designed to find these support vectors. The learning process consists then in finding the equations of hyper planes H_1 and H_2 determined by the support vectors.

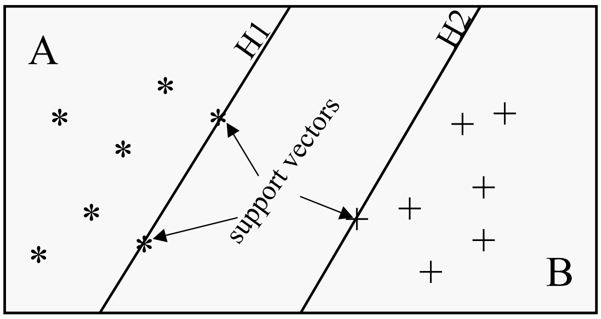


Figure 2: Example of two Linearly Separable Sets A (points *) and B (points +)

The classification is made as follow: a new pattern x^* is classified as

$$x^* \in A \; if \; sgn(\omega' x^* + b) = 1,$$
$$x^* \in B \; if \; sgn(\omega' x^* + b) = -1.$$

where *sgn* is the sign function defined by

$$sgn(a) = \begin{cases} 1 \; if \; a > 0, \\ -1 \; if \; a < 0. \end{cases}$$

4.2. Nonlinearly Separable case

In practice it is possible, due to the necessary measurement error and other uncertainties, that some elements of A lay in B and vice versa so that these sets are no longer linearly separable. In this case it is necessary to modify constraint (1) as

2 Notice that equalities $\omega' x_i + b = 1 \; and \; \omega' x_i + b = -1$ are not restrictive as by scaling one can put any equation of the form $\omega' x_i + b = \delta \; (\delta \neq 0)$ into this framework.

$$\begin{aligned} &\omega' x_i + b \geq \ \ 1 - \zeta_i \ \ for \ x_i \in A, \\ &\omega' x_i + b \leq -1 + \zeta_i \ \ for \ x_i \in B, \\ &\zeta_i \geq 0, \forall \, i \end{aligned} \tag{3}$$

where $\zeta_i, i = 1, 2, ..., N$ (N = number of training pattern) are some constant scalars. The quantity $\sum_{i=1}^{N} \zeta_i$ is an upper bound of the training error [15] so it must be taken into account in the cost function, that is the function to be minimized is taken as $\frac{1}{2}\|\omega\|^2 + C\left(\sum_{i=1}^{N} \zeta_i\right)^k$ where C is a constant to be chosen by the user and k is an integer. The optimization problem is then

$$\begin{aligned} &\min_{\omega, b, \zeta} \left\{ \frac{1}{2}\|\omega\|^2 + C\left(\sum_{i=1}^{N} \zeta_i\right)^k \right\}, \\ &s.t. \quad \begin{cases} \omega' x_i + b \geq \ \ 1 - \zeta_i \ \ for \ x_i \in A, \\ \omega' x_i + b \leq -1 + \zeta_i \ \ for \ x_i \in B, \\ \zeta_i \geq 0 \quad \forall \, i. \end{cases} \end{aligned} \tag{4}$$

with $\zeta = \begin{bmatrix} \zeta_1 & \zeta_2 & . & . & \zeta_N \end{bmatrix}'$. When $k = 1$ or 2 we recover a quadratic programming problem.

Remark 1: Problem (4) contains problem (2) so it is more general.

4.3. Nonlinear Support Vector Machines

The separating surfaces considered in the previous sections are linear surfaces, that is hyper planes. In practice, the sets A and B can be separated by nonlinear surfaces such as balls, ellipsoids or any imaginable surface. The problem is then how to extend the previous method in order to find such surfaces. The solution [15] is mapping first the actual data into a high order space (possible with infinite dimension) S using the map Φ defined by $\Phi : \Re^d \to S$, and then apply previous method. In order for an algorithm designed for selecting the support vectors to be applied, the map Φ must verify some properties: there must exist a kernel K such that $K(x_i, x_j) = \Phi(x_i)' \Phi(x_j)$. In fact it is not necessary to know explicitly the map Φ, the important issue is to dispose of kernel K. Some examples of such kernels are [15]:

- $K(x, y) = (x' y + 1)^p$ where p is an integer,
- $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$ with σ a given scalar,
- $K(x, y) = \tanh(\kappa x' y - \delta)$ where κ and δ are given scalars.

The following paragraph will present briefly some successful real world applications of support vector machines techniques.

4.4. Successful Practical Applications of SVM Techniques

The support vectors machines techniques presented so far have been applied with success for solving some real world classification problems. In [7] (see also [11], [12] and [1]) the problem of deciding whether a breast cancer patient needs or not a chemotherapy based on the existence or not of the metastasized lymph nodes removed from the patient's armpit during surgery has been carried up with success. In the domain of industrial diagnosis, failure detection issue has been formulated and solved using support vector data analysis, a technique based on support vector machines paradigm in [17], [18] (see also [17] for other applications by the same authors). The literature on the applications of support vector machines technique for real world problem solving is growing rapidly and one can find many applications from the specialized websites such as [1], [10] or [13] and references therein. Some applications that are considered as benchmarks for support vectors machines in biological cybernetics, medicine, text categorization, etc. can be found and downloaded from [10].

5. Applications

Let us now consider some applications to show the efficiency of the SVM based methods in the classification process. According to the small size of problems considered, the optimizations problems (2) or (4) are solved directly using Matlab Optimization Toolbox with its function *quadprog* (quadratic Programming).

5.1. Linear Case

Two examples corresponding to real world problems are considered; one in medical domain (Example 1) and the other in financial domain (Example 2).

5.1.1. Example 1

The following tables (Table I and Table II, see [2]) represent the measurements of urinary *androsterone* and *etiocholanolone* in healthy heterosexual and homosexual males in *mg/24* hours. The purpose is to design a discrimination surface between heterosexual and homosexual according to the ratio of androsterone and etiocholanolone.

Table I: Data for Heterosexual Group

Antrosterone	Etiocholanolone
3.9	1.8
4.0	2.3
3.8	2.3
3.9	2.5
2.9	1.3
3.2	1.7
4.6	3.4
4.3	3.1
3.1	1.8
2.7	1.5
2.3	1.4

Tables II: Data for Homosexual Group

Antrosterone	Etiocholanolone
2.5	2.1
1.6	1.1
3.9	3.9
3.4	3.6
2.3	2.5
1.6	1.7
2.5	2.9
3.4	4.0
1.6	1.9
4.3	5.3
2.0	2.7
1.8	3.6
2.2	4.1
3.1	5.2
1.3	4.O

Let us apply the SVM approach to this problem. The pattern are two dimensional vectors $x = [x_1 \quad x_2]'$ where x_1 represents the quantity of androsterone and x_2 the quantity of etiocholanolone. A visualization of data of Tables I & II show that these groups are linearly separable. The result of the optimization problem (2) using Matlab with Optimization Toolbox is

$$\omega = [4.6512 \quad -4.1860]' \quad and \quad b = -3.8372;$$

that is the separating line (central line on Figure 3 is given by the equation

$$4.6512x_1 - 4.1860x_2 - 3.8372 = 0.$$

One can see that this two sets are linearly separated and decision process can be made as:

- heterosexual if $\mathrm{sgn}(4.6512x_1 - 4.1860x_2 - 3.8372) = 1$,
- homosexual if $sgn(4.6512x_1 - 4.1860x_2 - 3.8372) = -1$

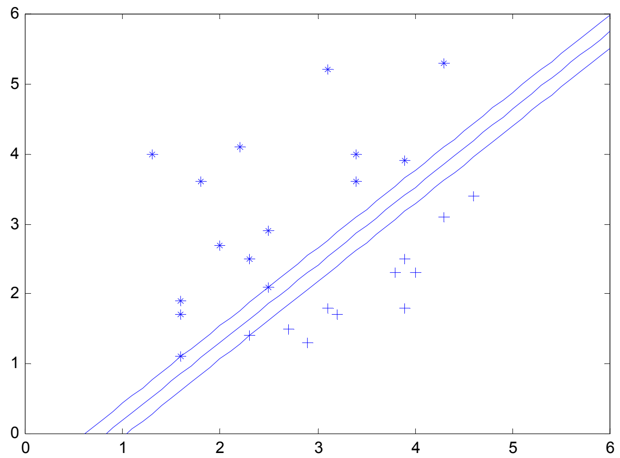


Figure 3: Linear Separation of Heterosexual (+) and Homosexual (*); the Horizontal Axis Corresponds to x_1 and the Vertical axis to x_2.

5.1.2. Example 2

The purpose here consists in discriminating between two groups of firms F_i (most admired (G_1) and least admired (G_2)) by using some financial data [14]. There are 5 ratios: MKBOOK (market to book value, x_1), ROTC (return on total capital, x_2), ROE (return on equity, x_3), REASS (return on assets, x_4), EBITASS (earnings before interest and taxes to total assets, x_5). The data are given on the Table III.

Table III: Data for Example 2

	Gr.	x_1	x_2	x_3	x_4	x_5
F_1	G_1	2.304	0.182	0.191	0.377	0.158
F_2	G_1	2.703	0.206	0.205	0.469	0.210
F_3	G_1	2.385	0.188	0.182	0.581	0.207
F_4	G_1	5.981	0.236	0.258	0.491	0.280
F_5	G_1	2.762	0.193	0.178	0.587	0.197
F_6	G_1	2.984	0.173	0.178	0.546	0.227
F_7	G_1	2.070	0.196	0.178	0.443	0.148
F_8	G_1	2.762	0.212	0.219	0.472	0.254
F_9	G_1	1.345	0.147	0.148	0.297	0.079
F_{10}	G_1	1.716	0.128	0.118	0.597	0.149
F_{11}	G_1	3.000	0.150	0.157	0.530	0.200
F_{12}	G_1	3.006	0.191	0.194	0.575	0.187
F_{13}	G_2	0.975	-0.031	-0.280	0.105	-0.012
F_{14}	G_2	0.945	0.053	0.019	0.306	0.036
F_{15}	G_2	0.270	0.036	0.012	0.269	0.038
F_{16}	G_2	0.739	-0.074	-0.150	0.204	-0.063
F_{17}	G_2	0.833	-0.119	-0.358	0.155	-0.054
F_{18}	G_2	0.716	-0.005	-0.305	0.027	0.000
F_{19}	G_2	0.574	0.039	-0.042	0.268	0.005
F_{20}	G_2	0.800	0.122	0.080	0.339	0.091
F_{21}	G_2	2.028	-0.072	-0.836	-0.185	-0.036
F_{22}	G_2	1.225	0.064	-0.430	-0.057	0.045
F_{23}	G_2	1.502	-0.024	-0.545	-0.050	-0.026
F_{24}	G_2	0.714	0.026	-0.110	0.021	0.016

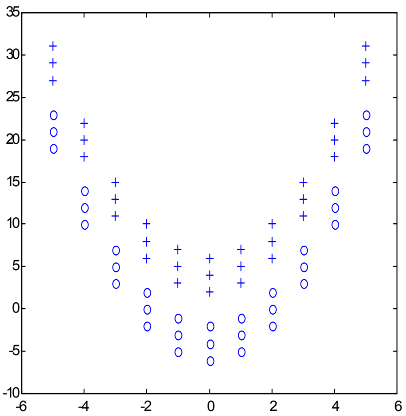
Here, it is not possible to visualize data as they belong to a space of dimension *5*; each pattern *x* is determined by $x = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5]' \in \Re^5$; so we use directly optimization problem (4). For a large value of C ($C \geq 15$) and *k = 1*, we obtain that these two groups, based on the financial ratios considered, are linearly separable, that is $\zeta = 0$ and the separating plane is defined by

$$H = \{x \in \Re^5 : 3.4253x_1 + 1.5744x_2 + 3.5463x_3 + 0.8690x_4 + 0.7502x_5 - 4.6806 = 0\}$$

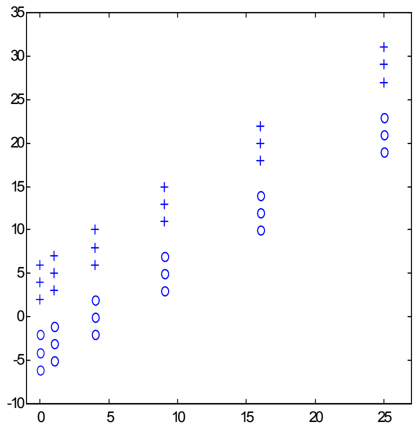
and the classification rule is given as previously with *sign* function. The same solution is obtained with *k = 2* and a very large constant *C* (*C* must be around 10^6).

5.2. Nonlinear Case

Consider data depicted on Figure 4 (a) where the set *A* consists in points (+) and the set *B* in points (o).



(a)



(b)

Figure 4: Data for Nonlinear Case (a) and Transformed Data by Φ; the Horizontal Axis Corresponds to x_1 and the Vertical axis to x_2.

These data are artificially generated just to show the possibility of SVM based approach in nonlinear case. It is clear that the two sets A and B are not linearly separable but can be separated by a parabola. Let us apply the following nonlinear transformation:

$$\Phi : \Re^2 \rightarrow \Re^2, \Phi(x) = [\Phi_1(x) \quad \Phi_2(x)]' = [x_1^2 \quad x_2]' \tag{5}$$

which corresponds to the kernel

$$K(x, y) = (x_1 y_1)^2 + x_2 y_2. \tag{6}$$

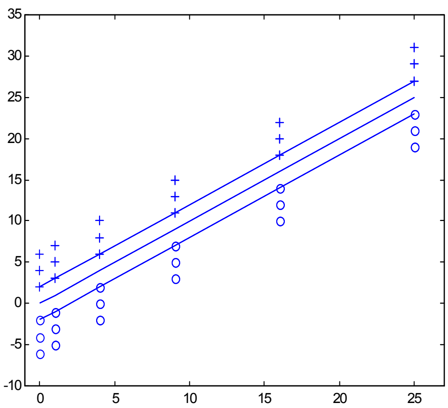
The application of this transformation on the previous data leads to data depicted on Figure 4 (b) which shows two linearly separable sets. The application of linear case SVM method gives the solution $[-0.5 \quad 0.5]'$ *and* $b = 0$. The classification of a new pattern $x \in \Re^2$ is carried as follow:

$$x \in A \; if \; \operatorname{sgn}(-\Phi_1(x) + \Phi_2(x)) = \; 1 \quad and \quad x \in B \; if \; \operatorname{sgn}(-\Phi_1(x) + \Phi_2(x)) = -1,$$

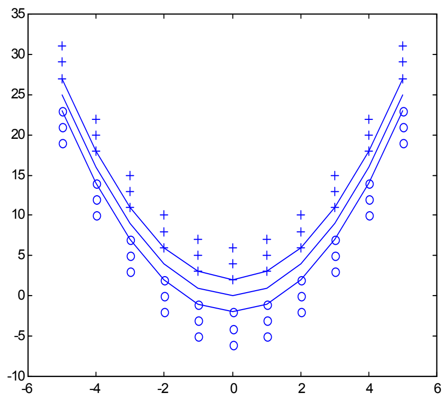
or equivalently

$$x \in A \; if \; \operatorname{sgn}(-x_1^2 + x_2) = \; 1 \quad and \quad x \in B \; if \; \operatorname{sgn}(-x_1^2 + x_2) = -1.$$

This shows that the nonlinear separating surface is the parabola of equation $x_2 = x_1^2$. The Figures 5 (a) & (b) show the representation of separating process on the transformed data and on the original data respectively.



(a)



(b)

Figure 5: Separation by Linear Surface (a) and Separation by Nonlinear Surface (parabola) (b); the Horizontal Axis Corresponds to x_1 and the Vertical Axis to x_2.

Remark 2: Notice that for examples discussed here, only the nonlinear case needs a kernel selection. For linear case, data of example 1 could be visualized (two dimensions data, Figure 3 without lines) and one could see that they are linearly separable; for example 2 it is not possible to visualize data and so nonlinearly separable formulation was used directly and gives satisfaction. For nonlinear case, the visualization of data (Figure 4 (a)) shows a parabola type representation and this guided the choice of nonlinear transformation (5) that leads to kernel (6).

6. Conclusion

The problem of pattern recognition (classification) has been considered through support vector machines. This mathematical tool is showing promising applications. Only basic aspects that are directly applicable for solving real world problems of this tool are considered here, more achievements can be found in the growing literature related to data mining such as [1], [11], [12], [15] and references in [10]. The small examples considered show that this tool may be useful for many applications for pattern recognition and classification in real world. Furthermore, for small applications such as that considered in this paper, existing optimization packages can directly be used to solve them.

REFERENCES

1. Data Mining Institute, http://www.cs.wisc.edu/dmi/

2. Publications on Kernel Methods, http://www.kernel-machines.org/

3. Quantitative Imaging Group, http://www.ph.tn.tudelft.nl/~davidt/papers.html

4. HAND D.J., Discrimination and Classification, **Wiley Series in Probability and Mathematical Statistics**, John Wiley & Sons, 1981.

5. HAYKIN S., **Neural Networks: A Comprehensive Presentation**, Prentice Hall, 1999.

6. HOFFBECK J. P. and LANDGREBE D. A., **Classification of Remote Sensing Images having High Spectral Resolution**, Remote Sensing of Environment, Vol. 57, No. 3, pp 119-126, September 1996.

7. JAIN A. K. and PANKANTI S., **Fingerprint Classification and Recognition, the Image and Video Processing Handbook**, A. Bovik (ed), Academic Press, April 2000.

8. LEE Y.-J., **Smooth Support Vector Machines, Preliminary Thesis Proposal**, Data Mining Institute, University of Wisconsin, 2000, http://www.cs.wisc.edu/dmi/.

9. LEE Y.-J., MANGASARIAN O.L. and WOLBERG W.H., **Survival-Time Classification of Breast Cancer Patients,** Data Mining Institute Technical Report 01-05, May 2001, http://www.cs.wisc.edu/dmi/

10. LOONEY C. G., **Pattern Recognition Using Neural Networks**, Oxford University Press, 1997.

11. JENSEN F. V., **Lecture Notes on Bayesian Networks and Influence Diagrams,** Department of Computer Science, Aalborg University, 1999.

12. MANGASARIAN O.L. and MUSICANT D.R., **Robust linear and Support Vector Regression**, Data Mining Institute Report, http://www.cs.wisc.edu/dmi/

13. MANGASARIAN O.L., **Mathematical Programming in Data Mining**, Data Mining Technical Report, Data Mining Institute, University of Wisconsin, 1997.

14. SHARMA S., **Applied Multivariate Techniques**, John Wiley & Sons, 1996.

15. SMOLA A.J. and SHÖLKOPF, **A Tutorial on Support Vector Regression**, NeuroCOLT2 Technical Report Series, NC2-TR-1998-030, 1998.

16. STONIC V. and BRADLEY K., **Labs for Signals and Systems using Matlab**, PWS Publishing Company, 1996.

17. TAX, D.M.J., YPMA, A. and DUIN, R.P.W., **Pump Failure Detection using Support Vector Data Description** , Proceedings of Intelligent Data Analysis, Amsterdam, 1999.

18. YPMA, A., TAX, D.M.J. and DUIN, R.P.W. **Robust machine fault detection with Independent Component Analysis and Support Vector Data Description**, Proceedings of the 1999 IEEE Workshop on Neural Networks for Signal Processing, 1999, Madison.

19. VAPNIK V., **The nature of statistical learning theory**, Springer-Verlag, New York, 1995.