

Predicción de hipertensión arterial usando máquinas de vectores de soporte

César Seijas⁽¹⁾, Antonino Caralli⁽¹⁾, Sergio Villasana⁽¹⁾, Laura Sáenz⁽²⁾, Francisco Arteaga⁽³⁾

⁽¹⁾ Unidad de Investigación en Bioelectrónica, Departamento de Electrónica, Escuela de Eléctrica

⁽²⁾ Centro de Investigaciones de Mecánica, CIMEC

⁽³⁾ Centro de Investigación y Tecnología en Automatización Electrónica y Control, CITAEC

Universidad de Carabobo, Facultad de Ingeniería, Valencia, Venezuela

E-mail: cseijas@uc.edu.ve, acaralli@ieee.org, svillaza@uc.edu.ve

lsaenz@uc.edu.ve, farteaga@uc.edu.ve

Resumen

En este trabajo se usó las Máquinas de Vectores de Soporte (*SVM*, de sus siglas en inglés) como predictor o clasificador (*SVC*) del padecimiento de Hipertensión Arterial (*HTA*) en humanos. El *SVC* desarrollado usó factores fisiológicos, que según los expertos pueden producir *HTA*, para decidir el padecimiento o no de la enfermedad. Los conjuntos de entrenamiento y validación del *SVC* construido constan de 600 y 200 datos respectivamente, tomados aleatoriamente de una base de datos de la *Asociación Norteamericana de HTA*, año 2000, con 7200 evaluaciones de *HTA* y 28 factores fisiológicos antes mencionados. La *SVC* construida usó como “kernel” funciones de base radial, lográndose un error medio cuadrático menor de 3%. Se verificó el excelente desempeño como clasificador de las *SVM*, tanto con datos de distribución conocida, como de datos experimentales, correspondientes a la evaluación de variables fisiológicas humanas obtenidas en consultas cardiológicas.

Palabras clave: Máquinas de vectores de soporte, clasificación, hipertensión arterial.

Prediction of arterial hypertension using support vector machines

Abstract

This work presents the Support Vector Machines (*SVM*) as predictor or classifier (*SVC*) of the Human Arterial Hypertension (*HTA*) illness. The developed *SVC* used physiological factors, which according to experts can produce *HTA*, to decide the suffering or not of the illness. Training and validation sets of the built *SVC* are respectively made up of 600 and 200 samples, which were taken randomly from the North American Association data base of *HTA*, year 2000, with 7200 *HTA* evaluations and 28 physiological factors as before mentioned. Designed *SVC* used as kernel the radial basis function, being achieved a mean square error less than 3%. The excellent performance of the *SVM* as classifier was verified, with data of know distribution as well as with data of experimental data, corresponding to the evaluation of human physiological variables obtained from cardiological consulting.

Keywords: Support vector machines, classification, arterial hypertension.

I. INTRODUCCIÓN

Las Máquinas de Vectores de Soporte (*SVM*) son algoritmos emergentes del área de la ciencia computacional conocida como *Inteligencia Artificial* que han sido usados exitosamente en problemas de clasificación y regresión. El fundamento matemático es debido a Vapnik [1] en 1964 que fue desarrollado y generalizado en los 90. A diferencia de las *Redes Neuronales Artificiales (RNA)* que invocan, durante la fase de entrenamiento, el principio de *Minimización del Riesgo Empírico (ERM)*, de sus siglas en inglés), las *SVM* se basan en el principio de *Minimización del Riesgo*

Estructural (SRM), siglas en inglés). En general, *SRM* ha mostrado mejor desempeño que *ERM* [1]. La *SVM* minimiza un límite superior al riesgo esperado a diferencia de *ERM* que minimiza el error sobre los datos de entrenamiento. Entre algunas aplicaciones de las *SVM*, se pueden citar: detección de caras en imágenes de grupos de personas [2], clasificación [3],[4] y regresión [5], predicción de series de tiempo financieras [6] y biológicas [7], y sistemas de control [8], entre otras aplicaciones. En este trabajo desarrolla un proceso de predicción del posible padecimiento de Hipertensión Arterial (*HTA*) en personas, a través de un mecanismo inteligente, basado en la Máquina de Vectores de Soporte (*SVM*, de sus siglas en

inglés), lo que corresponde a un problema de clasificación de pacientes proclives a padecer de HTA o no.

II. FUNDAMENTOS TEÓRICOS

A. Clasificación usando SVM

En el caso de aplicar las SVM para clasificación, la aplicación puede restringirse a un problema de dos clases sin pérdida de generalidad. El problema es separar dos clases por una función inducida de ejemplos disponibles. El objetivo es clasificar datos o ejemplos fuera del conjunto de entrenamiento, esto es generalizar correctamente.

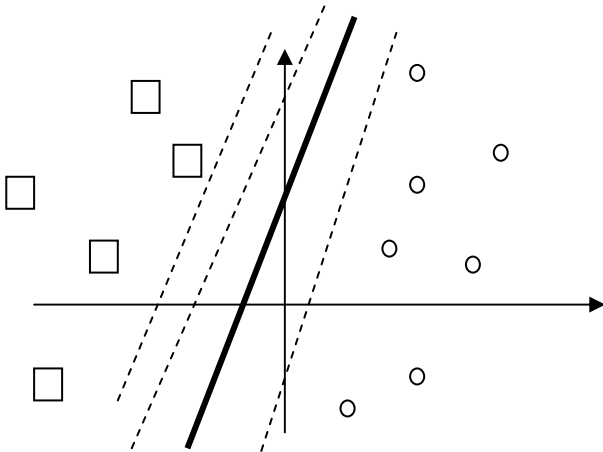


Figura 1. Clasificación de objetos con hiperplanos.

En la Figura 1 se ilustra un problema de clasificación de objetos donde se busca definir un hiperplano separador que separe las clases u objetos: cuadrados de los ovalados. Obsérvese que existen infinitos hiperplanos separadores (señalados con líneas punteadas) pero solo uno maximiza la distancia entre el mismo y el objeto mas cercano de cada clase (línea sólida), a este clasificador lineal se le denomina *Hiperplano Separador Óptimo (HSO)*. La definición formal del HSO es la siguiente: considere el problema de separar los conjuntos de entrenamiento que pertenecen a las clases x e y , como se indica a continuación:

$$D = \{(x^1, y^1), \dots, (x^l, y^l)\}, x \in R^n, y \in \{-1, 1\} \quad (1)$$

con un hiperplano,

$$\langle w, x \rangle + b = 0, \quad w \in R^n, b \in R \quad (2)$$

El conjunto de vectores (datos) se dice que son óptimamente separados, si el hiperplano los separa sin error y la distancia entre el vector mas próximo al

hiperplano es máxima. En ese caso el hiperplano definido por (2) es el HSO. Sin pérdida de generalidad, usualmente se considera un hiperplano canónico, donde los parámetros w, b se restringen según,

$$\min_i |\langle w, x^i \rangle + b| = 1 \quad (3)$$

es decir, la norma del vector de pesos w debe ser igual al inverso de la distancia del punto mas cercano en el conjunto de datos al HSO. Un HSO en forma canónica debe satisfacer las siguientes restricciones,

$$y^i [\langle w, x^i \rangle + b] \geq 1, \quad i = 1, \dots, l. \quad (4)$$

donde la distancia $d(w, b; x)$ de un punto x al HSO es,

$$d(w, b; x) = \frac{|\langle w, x^i \rangle + b|}{\|w\|} \quad (5)$$

El HSO se consigue maximizando el margen ρ , donde el margen se considera como la distancia entre el HSO y los vectores (datos) mas próximos a dicho hiperplano correspondientes a cada clase. La maximización de ρ está sujeta a las restricciones definidas por la ecuación (4); el margen ρ , se obtiene como,

$$\begin{aligned} \rho(w, b) &= \min_{x^i: y^i = -1} d(w, b; x^i) + \min_{x^i: y^i = +1} d(w, b; x^i) \\ &= \min_{x^i: y^i = -1} \frac{|\langle w, x^i \rangle + b|}{\|w\|} + \min_{x^i: y^i = +1} \frac{|\langle w, x^i \rangle + b|}{\|w\|} \\ &= \frac{1}{\|w\|} \left(\min_{x^i: y^i = -1} |\langle w, x^i \rangle + b| + \min_{x^i: y^i = +1} |\langle w, x^i \rangle + b| \right) \\ &= \frac{2}{\|w\|} \end{aligned} \quad (6)$$

en consecuencia el HSO minimiza,

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (7)$$

independiente de b , para el siguiente límite

$$\|w\| < A \quad (8)$$

entonces de las ecuaciones (7) y (8)

$$d(w, b; x) \geq \frac{1}{A} \tag{9}$$

es decir, los *HSO* no pueden estar mas cerca de $1/A$ de los puntos de datos. La solución al problema de optimización planteado por la ecuación (7), sujeto a las restricciones establecidas por (8) corresponde al punto de silla del siguiente funcional *Lagrangiano*:

$$\Phi(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y^i [\langle w, x^i \rangle + b] - 1) \tag{10}$$

En la ecuación (10), los α_i son los *multiplicadores de Lagrange*. La solución del problema anterior tiene un dual dado por,

$$\max_{\alpha} W(\alpha) = \max_{\alpha} W \left(\min_{w, b} \Phi(w, b, \alpha) \right) \tag{11}$$

el mínimo con respecto a w y b del *Lagrangiano*, Φ , se deriva de:

$$\frac{\partial \Phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \tag{12}$$

$$\frac{\partial \Phi}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i$$

luego de las ecuaciones (10), (11) y (12) se tiene:

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \tag{13}$$

y por tanto, la solución es,

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \tag{14}$$

sujeta a las restricciones,

$$\alpha_i \geq 0, \quad i=1, \dots, l, y$$

$$\sum_{j=1}^l \alpha_j y_j = 0$$

el *HSO* es,

$$w^* = \sum_{i=1}^l \alpha_i y_i x_i \tag{15}$$

$$b^* = -\frac{1}{2} \langle w^*, x_r + x_s \rangle$$

donde x_r y x_s son los *vectores de soporte* de cada una de las clases que satisfacen,

$$\alpha_r \alpha_s > 0, \quad y_r = -1, \quad y_s = +1 \tag{16}$$

definido el *HSO*, se puede diseñar un clasificador, usando una función fuertemente selectiva como lo es la función *signo* (*sgn*), luego un clasificador o *SVC* “duro” viene dado por,

$$f(x) = \text{sgn}(\langle w^*, x \rangle + b) \tag{17}$$

de las condiciones de *Kuhn-Tucker* [1],

$$\alpha_i \cdot (y^i [\langle w, x^i \rangle + b] - 1) = 0, \quad i=1, \dots, l \tag{18}$$

solo los puntos x_i que satisfagan,

$$y^i [\langle w, x^i \rangle + b] = 1 \tag{19}$$

tendrán *multiplicadores de Lagrange* no nulos; esos x_i se denominan *Vectores de Soporte (SV)*. Si los datos son linealmente separables, todos los *SV* residen en el margen y por tanto la cantidad de *SV* es pequeña. Consecuentemente el hiperplano es determinado por un pequeño subconjunto del conjunto de entrenamiento; es decir, descartando todos los otros vectores o datos del conjunto de entrenamiento y re-entrenando se obtendrá el mismo *HSO* y *SV*. Esta es una de las características que hacen atractivas las *SVM*, ya que puede resumirse la información contenida en el conjunto de datos a los *SVs* producidos.

En el caso general, los datos del conjunto de entrenamiento no son linealmente separables. Existen dos estrategias para generalizar el problema, las cuales dependen del conocimiento a priori del problema en análisis y del estimado de ruido en los datos. En el caso donde se considera que un hiperplano puede separar correctamente la mayoría de los datos puede introducirse una función de costo adicional, que penalice los errores de clasificación. En ese sentido se puede definir un *HSO* generalizado introduciendo variables no negativas, $\xi_i \geq 0$, y una función de castigo,

$$F_{\sigma}(\xi) = \sum_i \xi_i^{\sigma} \quad \sigma > 0 \tag{20}$$

donde los ξ_i representan una medida de los errores de clasificación. En el caso en análisis, el problema de optimización ahora debe minimizar tanto el error de clasificación como el acotamiento de la dimensión *VC* del

SVC. Las restricciones impuestas por la ecuación (4), para el caso no separable, deben ser modificadas como,

$$y^i \left[\langle w, x^i \rangle + b \right] \geq 1 - \xi_i, i = 1, \dots, l. \quad (21)$$

donde $\xi_i \geq 0$. El *HSO* generalizado es determinado por el vector w , que minimiza el funcional,

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (22)$$

obsérvese que en la ecuación anterior, se introduce el parámetro C o capacidad de la *SVM*, el cual se fija de modo que refleje, en alguna medida, el ruido estimado en los datos.

La solución al problema de optimización planteado en (22) corresponde ahora al punto de silla del *Lagrangiano*,

$$\Phi(w, b, \alpha, \xi, \beta) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l (y^i [w^T x^i] - 1 + \xi_i) - \sum_{j=1}^l \beta_j \xi_j \quad (23)$$

siendo α y β los *multiplicadores de Lagrange*. La solución es de la forma,

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^l \alpha_k \quad (24)$$

con las restricciones,

$$\begin{aligned} 0 \leq \alpha_i \leq C \quad i = 1, \dots, l \\ \sum_{j=1}^l \alpha_j y_j = 0 \end{aligned} \quad (25)$$

es decir, la solución al problema de optimización es idéntica al caso separable excepto por una modificación al límite superior de los *multiplicadores de Lagrange*.

Usualmente, no es posible separar con una frontera lineal, en la dimensión del espacio de los datos; en ese caso, la *SVM* puede proyectar el vector de entrada x , en un hiper-espacio de rasgos z . Con una proyección no lineal adecuada, será posible separar los vectores en ese hiper-espacio. El problema de optimización de (22) es ahora, de la forma,

$$\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{k=1}^l \alpha_k \quad (26)$$

donde $K(x, x')$ es la función “*kernel*” que ejecuta la proyección no lineal en el espacio de rasgos, con las

mismas restricciones señaladas por (25), $K(x, x')$ es de la forma.

$$K(x, x') = \langle \phi(x), \phi(x') \rangle \quad (26)$$

un *SVC* duro que implementa el *HSO* en el espacio de rasgos es:

$$f(x) = \text{sgn} \left(\sum_{i \in SVs} \alpha_i K(x_i, x) + b \right) \quad (27)$$

donde

$$\langle w^*, x \rangle = \sum_{i=1}^l \alpha_i y_i K(x_i, x) \quad (28)$$

$$b^* = -\frac{1}{2} \sum_{i=1}^l \alpha_i y_i [K(x_i, x_r) + K(x_i, x_r)]$$

El problema principal de aplicar o proyectar vectores de entrada a un espacio hiperdimensional se conoce como “la maldición de la dimensionalidad”, que consiste en un crecimiento potencial del tiempo de cálculo computacional. El objetivo de usar funciones “*kernel*” es permitir que las operaciones se ejecuten en el espacio de entrada, por tanto el producto escalar no necesita ser evaluado en el espacio de rasgos. Las funciones que califican como funciones “*kernel*” deben satisfacer ciertas condiciones (condiciones de *Mercer* [1])

$$K(x, x') = \sum_m^{\infty} a_m \Phi_m(x) \Phi_m(x'), \quad a_m > 0 \quad (29)$$

$$\iint K(x, x') g(x) g(x') dx dx' > 0, \quad g \in L_2 \quad (30)$$

A continuación se mencionan funciones que califican como funciones “*kernel*”: polinomiales, funciones de base radial gaussianas, de base radialexponencial, perceptrón multicapa, series de Fourier, splines y suma y productos (tensores) de *kernels*.

III. PARTE EXPERIMENTAL

A continuación se describe el diseño y la evaluación del *SVC*. El *SVC* diseñado se aplicó en el diagnóstico precoz *HTA* en humanos, usando como datos de entrenamiento, datos fisiológicos usualmente evaluados en una consulta médica de rutina, de donde se obtuvo el conjunto de vectores de entrenamiento. El conjunto de entrenamiento consta de 600 datos o vectores, mientras que como conjunto de validación, se usaron 200 vectores,

los datos de entrenamiento y validación se obtuvieron de la base de datos de la *Asociación Norteamericana de HTA*, año 2000, institución adscrita a la *Organización Mundial de la Salud*. Cada vector consta de 28 componentes cada uno, de los cuales, las primeras 27 componentes corresponden a los factores que los expertos consideran, produce predisposición a padecer de *HTA*, siendo sub-clasificados en factores primarios o secundarios según su influencia en el riesgo de sufrir el padecimiento. Los factores primarios son: sobrepeso, padecimiento de enfermedades cardiovasculares, antecedentes familiares de *HTA*, diabetes, padecimiento de enfermedades pulmonares y padecimiento de enfermedades renales.

Entre los factores secundarios se consideran: antecedentes familiares cardiovasculares o pulmonares, hábitos de tabaquismo, edad, niveles anormales de ácido úrico, azúcar, colesterol, triglicérido, stress, entre otros.

Uno de los problemas a resolver en el diseño del *SVC* es el de la identificación de la función “kernel” apropiada, así como de los parámetros de sintonía de dicho “kernel”; con este objetivo se evaluaron varias topologías y se compararon, en base a velocidad de convergencia y exactitud en el diagnóstico, concluyéndose que la función “kernel” con mejor desempeño fue la del tipo de función con base radial. Como plataforma de desarrollo del algoritmo *SVC* se usó el software matemático *MATLAB* y una herramienta de aplicación académica para diseños de *SVM* desarrollada por S. Gunn [9], el cual fue analizado y adaptado al desarrollo de este proyecto por sus autores. El programa ambientado en el software matemático *Matlab*, es manejable por el usuario, a través de una interfaz gráfica que se denomina *Uiclass*.

Como primer paso en el procesamiento de los datos de entrada, estos deben ser normalizados, lo cual lo ejecuta la rutina *svdatanorm(x,ker)*, cuyos parámetros son x el vector de entrenamiento de entrada a normalizar y la función *kernel* de ensayo (para establecer el dominio de normalización). A continuación se realiza la clasificación mediante la rutina *SVC(x_m,y,ker,C)* que tiene como parámetros de entrada: el vector de entrada normalizado x_n , el vector objetivo y , la función *kernel* y la capacidad C que regula el error de clasificación (22). La rutina *SVC* produce como salida un matriz de tres componentes: el número de *SV*, el vector de multiplicadores de Lagrange (α_i) y el término de polarización (b). Para evaluar el *SVC*, se usa la rutina *SVCoutput(x,y,x_p,ker,a₁,b₁,f)*, que entre otros parámetros de entrada, es alimentada con el vector

de prueba x_p , los parámetros del *SVC* previamente entrenado (vector de multiplicadores de Lagrange, α_i y el término de polarización, b). Para medir la calidad del pronóstico y la visualización del proceso se construyeron las rutinas: *SVError* y *SVPlot*

IV. RESULTADOS DE LA SIMULACIÓN

Como conjunto de entrenamiento se experimentó con diferentes tamaños de conjuntos de datos tomados aleatoriamente de la base de datos citada en párrafos anteriores, pero tratando de que estos vectores correspondiesen, con distribución uniforme, a casos tanto de padecimiento de *HTA*, como de personas sanas. El mínimo error de entrenamiento correspondió a matrices de 600 filas y 27 columnas extraída de la base de datos antes citada, como se muestra en la figura 2.

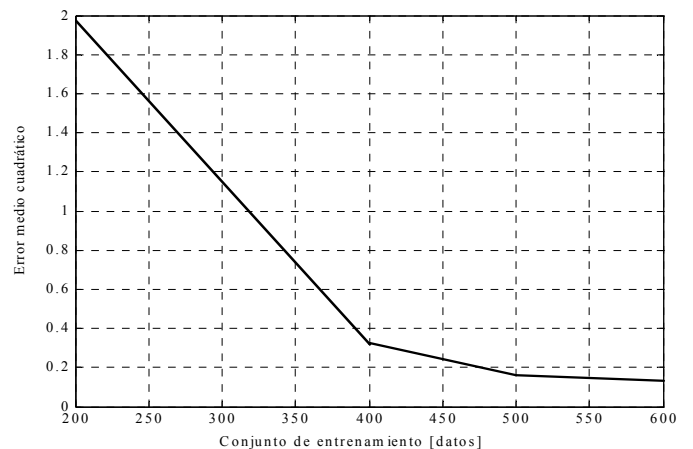


Figura 2. Error medio cuadrático vs. Tamaño de matriz de entrenamiento.

Se ensayó con diferentes funciones *kernel* y el parámetro de capacidad del *SVC*, C , observando que el menor error de validación, es decir con datos fuera del conjunto de entrenamiento se logró con funciones de base radial (*RBF*), optando en consecuencia, a restringir los experimentos de evaluación de desempeño a este *kernel*, como se presenta en la Tabla 1.

Tabla 1. Selección del “kernel”.

Kernel	Error de Validación
Lineal	6.4598
Polinomiales	3.5707
RBF	0.1333
Spline	0.5362

Se ensayaron diferentes valores del coeficiente de dispersión σ de la *RBF*, lográndose el menor error porcentual (menos de 3%), bajo una combinación de $\sigma = 0.1$ y $C = 1000$, como se muestra en la Tabla 2.

Tabla 2. Parámetros de la SVC con “kernel” RBF a $C = 1000$.

σ	T_e (seg)	NVS	Error de Validación
0.01	87.2	600	1.355
0.1	88.0	600	1.455
1.0	86.0	600	1.555
3.0	87.7	600	3.002

Se propicia con este modelo, la simulación de una alteración fisiológica humana, que padecen, 4 de cada 10 personas [10], sin proceder a métodos invasivos. Se verificó el excelente desempeño como clasificador de la *Máquina de Vectores de Soporte*, tanto con datos de distribución conocida, como de datos experimentales, correspondientes a la evaluación de variables fisiológicas humanas en consultas cardiológicas. El mejor comportamiento como función “*kernel*” lo exhibió la función del tipo de base radial. Los resultados presentados en este artículo expresan un futuro promisorio a las *SVM* en el área de la ingeniería aplicada a la medicina

VII. REFERENCIAS

- [1] Vapnik V.N., "The Nature of Statistical Learning Theory", Springer-Verlag, New York, ISBN 0-387-94559-8, 1995.
- [2] Richman M. & T. Parks, "Detection of Noses and Faces in Consumer Images", Center for applied Mathematics, Cornell Univ., U.S.A., 1998.
- [3] Back. Andrew, "Classification using Support Vector Machines", Riken Brain Science Institute, Wako-shi, Saitama, Japan, 1998.
- [4] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V, "Gene Selection for Cancer Classification using Support Vector Machines", Barnhill Bioinformatics, Savannah, Goergia, U.S.A., 1999.
- [5] Osuna E., Mukherjee, Girosi F., "Nonlinear Prediction of Chaotic Time Series using SVM", M.I.T., Cambridge, U.S.A., 1997.
- [6] Tay F. & L. Cao, "Application of Support Vector Machines in Financial Time Series Forecasting", Department of Mechanical Engineering, National University of Singapore, Singapore, 2001.
- [7] Frontzek T., T. Navin & R. Eckmiller, "Predicting the Nonlinear Dynamics of Biological Neurons using Support Vector Machines with Different Kernels", Department of Computer Science VI, University of Bonn, F. R. Germany, 2001.
- [8] De Kruif B. & T. de Vries, "On using a SVM in learning Feed-Forward Control", Institute of Mechatronics, Univ. of Twente, The Netherlands, 2001.
- [9] Gunn Steve, "Support Vector Machines for Classification and Regression", Faculty of Engineering and Applied Science, University of Southampton, U.S.A., May 1998.
- [10] Lindholm LH, Hansson L, Schersten B, Eldom T. - Morbidity and mortality in the Swedish Trial in Old Patients with hypertension (STOP) - Hypertension 1992; 251-255.