수정된 kernel-adatron알고리즘에 의한 Support Vector Machines의 학습

조용현 박창환[0]
대구효성가톨릭대학교 공과대학 전자정보공학부
{yhcho, g9521007}@cuth.cataegu.ac.kr

Training of Support Vector Machines Using the Modified Kernel-adatron Algorithm

Young-Hyun Cho and Chang-Hwan Park[0]
School of Electronics and Information Engineering,
Catholic University of Taegu-Hyosung

요 약

본 논문에서는 모멘트 항을 추가한 수정된 kernel-adatron알고리즘을 제안하고 이를 support vector machines의 학습기법으로 이용하였다. 이는 기울기상승법에서 일어나는 최적해로의 수렴에 따른 발진을 억제하여 그 수렴속도를 좀더 개선시키는 모멘트의 장점과 kernel-adatron알고리즘의 구현용이성을 그대로 살리기 위함이다. 제안된 학습기법의 SVM을 실제 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, Campbell등의 kernel-adatron알고리즘을 이용한 SVM의 결과와 비교할 때 학습시간과 시험 데이터의 분류률에서 더욱 우수한 성능이 있음을 확인할 수 있었다.

1. 서론

학습과 대규모 병렬처리에 기인한 신경망을 이용한 접근은 데이터의 overfitting에 따른 모델을 생성할 수 있어 일반화측면에서 어려움이 뒤따르게 된다. 이는 학습을 위한 최적화 알고리즘과 가장 좋은 모델을 선택하는데 이용된 통계적 척도의 결과이다. 이러한 신경망이 가지는 제약들을 해결하기 위해서 support vector machine(SVM)이 제안되었으며, 문자, 얼굴, 그리고 물체 인식 등의 실제분야에 성공적으로 적용되었다[1-5].

SVM이 overfitting을 효과적으로 막아주는 것은 VC(Vapnik-Chervonenkis) 이론으로 설명될 수 있으며[1], 그것의 학습은 볼록함수를 최대화함으로써 수행되는 것으로 이는 polynomial time내에 발견될 수 있는 유일해가 존재한다는 것을 의미한다. 또한 망의 가중치는 선형 부등과 항등 조건을 가진 2차 프로그래밍(quadratic programming : QP) 문제를 풂으로서 찾을 수 있다.

최근 수많은 실제 분야에서 SVM이 성공적으로 적용되었으나 아직까지 machine learning에서 신경망 등과는 달리 표준도구로 인정받지 못하고 있다. 이는 QP문제를 푸는 데는 복잡한 계산이 요구되며 시스템을 구현하는데도 어려움이 있기 때문이다. 이때 요구되는 메모리 량은 데이터 수의 자승에 해당하며 학습 또한 대단히 느리게 이루어진다. 특히, SVM의 학습알고리즘은 복잡하고 미묘하며 때로는 구현이 어렵다. 이러한 학습에서의 제약점을 해결하기 위한 많은 연구가 진행되고 있다[3-5]. 그러나 대부분의 기법들은 QP 알고리즘이 가지는 메모리 문제나 수치적인 2차 프로그래밍의 수행으로 인한 시간의 제약을 가지고 있다. 특히, Campbell 등은 커널(kernel) 방법과 퍼셉트론 규칙을 병합한 반복기법의 kernel-adatron알고리즘을 제안하였다[3]. kernel-adatron알고리즘은 adatron의 용이한 구현성과 커널에 의한 비선형 특징공간에서의 동작을 조합한 하이브리드 학습기법이다. 여기서는 특징공간에서의 여유를 최대화하기 위하여 기울기상승법을 이용하고 등식의 조건을 만족시키는 라그랑지안 계수을 얻기 위해서 할선법(secant method)을 이용하였다. 그러나 기울기상승법은 해의 변화를 라그랑지안의 미분에 따라 변화시킴으로서 최적해로 수렴할 때 발진이 발생할 수도 있다. 따라서 우수한 구현성은 그대로 살리면서도 최적해로의 수렴에서 발생하는 발진을 막아 좀 더 빠른 속도로 SVM을 학습시킬 수 있는 효과적인 기법의 연구가 요구된다.

본 연구에서는 기울기상승법에서 해의 변화에 과거의 속성을 반영하는 모멘트(momentum)항을 추가한 수정된 kernel-adatron알고리즘을 제안하고 이를 SVM의 학습기법으로 이용하였다. radial basis function의 커널을 이용한 제안된 학습기법의 SVM을 실제 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제[6]에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, kernel-adatron알고리즘을 각각 SVM에 적용한 결과와 비교 고찰하였다.

2. Support Vector Machines

SVM은 기존의 신경망 등에서 이용된 경험적 위험을 최소화하는 원리보다는 구조적 위험을 최소화하는 근사적 구현이다. 구조적 위험을 최소화하는 방법은 학습오차 비율의 합으로 범위가 결정되는 시험오차항과 학습머신의 VC-차원에 의존하는 항에 기반을 두고 있다. 여기서는 SVM을 분류문제를 대상으로 소개한다.

2개의 분리된 클래스에 속하는 학습벡터의 집합을 선형적으로 분리가능한 문제를 생각해 보자. 이는 가중치 벡터 $\mathbf{\omega}$ 와 바이어스 θ 로 구성되는 $(\mathbf{\omega}\cdot\mathbf{x})+\theta=0$의 초월면을 가지도록 표본 $\{(x_i, d_i)\}_{i=1}^{N}$을 학습시키는 것이다. 여기서 $\mathbf{x}_i$는 i번째 예에 대한 입력패턴이고, d_i 는 ±1의 값을 가지는 원하는 응답이다. 초월면$(\mathbf{\omega}\cdot\mathbf{x})+\theta=0$은 표준형태(canonical form)로 나타낸 다음의 조건을 만족한다. 즉,

$$(\mathbf{\omega}\cdot\mathbf{x}_i)+\theta \ge 1, \quad \text{if} \quad d_i=1$$
$$(\mathbf{\omega}\cdot\mathbf{x}_i)+\theta < -1, \quad \text{if} \quad d_i=-1 \tag{1}$$

이다. 위의 표준형태를 단일항의 식으로 나타내면

$$d_i(\mathbf{\omega}\cdot\mathbf{x}_i)+\theta \ge 1, \quad i=1,2,\cdots,N \tag{2}$$

이다. 식 (2)에서 등호의 조건을 만족하는 입력패턴 $(\mathbf{x}_i, d_i)$을 support vector라 하며, 따라서 분류를 위한 학습은 제약조건 식 (2)을 만족하는 최적의 초월면을 찾는 것이다. 여기서 최적은 최대여유를 가지는 것이며, 최대여유 초월면은 최적으로 2개의 클래스를 분리할 수 있는 초월면이다. 결국 2개의 클래스와 초월면사이의 거리에 따른

여유ρ (ω, θ)는 $2/\|w\|$가 되며, 입력패턴을 최적으로 분류하는 초월면은 다음의 비용함수 $\theta(\omega)$를 최소화한다. 비용함수 $\theta(\omega)$는 $\theta(\omega)=\frac{1}{2}\|w\|^2$ 이며, ω의 볼록함수으로 제약조건 식 (2)은 ω에 선형임을 알 수 있다.

SVM을 정리하면, 학습패턴이 주어질 때 제약조건 식 (2)을 만족하는 가중치 벡터 ω와 바이어스 θ를 찾는 최적화 문제를 기본문제(primal problem)로 생각할 수 있으며, 이때 ω는 비용함수도 최소화한다. 이 최적화 문제를 해결하기 위해 라그랑지안 계수법을 이용하면 다음과 같은 라그랑지안 함수를

$$J(\omega, \theta, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i [d_i(w^T x_i + \theta) - 1] \quad (3)$$

얻을 수 있다. 식에서 α_i는 라그랑지안 계수들이다. 식 (3)의 최적화 문제에 대한 해는 ω와 θ에 대해서는 최소화되며 $\alpha_i \geq 0$에 대해서는 최대화되어야 한다. 따라서 ω와 θ에 대한 $J(\omega, \theta, \alpha)$의 최소는 그 각각에 대한 미분으로 얻어질 수 있다. 즉,

$$\frac{\partial J}{\partial \theta} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} \alpha_i d_i = 0$$

$$\frac{\partial J}{\partial w} = 0 \quad \rightarrow \quad w = \sum_{i=1}^{N} \alpha_i x_i d_i \quad (4)$$

이다. 한편 α_i를 구하기 위해 기본문제에 대한 라그랑지안 함수 $J(\omega, \theta, \alpha)$를 이원문제(dual problem)의 목적함수 $Q(\alpha)$로 표현하면

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j (x_i^T x_j) \quad (5)$$

과 같고, 조건은 $\sum_{i=1}^{N} \alpha_i d_i = 0$와 $\alpha_i \geq 0$ (i=1,2, … ,N)이다. 식 (5)의 목적함수는 일반적인 QP문제의 형태로 학습패턴의 항으로만 구성되며, 입력패턴의 내적(dot product)의 형태인 $(x_i^T x_j)$로 표현된다. 따라서 분류문제를 식 (5)의 이원문제로 생각하면, 이는 학습패턴 $\{(x_i, d_i)\}_{i=1}^{N}$이 주어질 때, 제약조건 $\sum_{i=1}^{N} \alpha_i d_i = 0$와 $\alpha_i \geq 0$ (i=1,2,… ,N)을 만족하는 목적함수 식 (5)을 최대화하는 라그랑지안 계수 α_i을 찾는 것이다.

이상에서 본 분류식은 선형의 결정면만을 설명한 것으로 비선형 분류도 가능하게 하기 위해서는 좀 더 일반적인 결정면을 가지도록 하여야 한다. 이를 위한 방법으로 SVM에서는 입력벡터 $\mathbf{x}$를 고차원의 특징공간 $\mathbf{z}$로의 사상을 이용한다. 그러나 이를 위한 목적함수 $Q(\alpha)$를 최대화하는 것은 고차원의 특징공간에서 내적 $(\mathbf{z}(\mathbf{x}) \cdot \mathbf{z}(\mathbf{x}_i))$의 계산을 요구한다. 어떤 조건하에서, 내적 계산은 아주 비효율적이나 커널함수 K를 사용함으로써 효율적인 계산이 가능하다. 여기서 커널함수와 특징공간과의 관계는 다음과 같다. 즉,

$$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{z}(\mathbf{x})^T \cdot \mathbf{z}(\mathbf{x}_i) \quad (6)$$

이며, 이때 사상을 위한 커널함수로는 polynomial, radial basis, 그리고 s-자형 함수 등이 이용된다[2]. 따라서 비선형 분류도 가능하게 하는 최적의 초월면을 구성하기 위한 목적함수 $Q(\alpha)$는

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (7)$$

로 표현되며, 이는 선형 분류의 목적함수에서 입력벡터의 내적을 커널함수로 대치한 것에 불과하다. 여기서의 조건도 선형적으로 분리 불가능한 경우와 동일하다. 또한 특징공간에서 최적의 분리를 위한 초월면의 분류식은 다음과 같이 나타낼 수 있다. 즉,

$$f(\mathbf{x}) = sign[\sum_{i=SVs} \alpha_{io} d_i K(x, x_i) + \theta_o] \quad (8)$$

이다. 여기서 SVs는 support vector를 나타내며, 최적의 가중치 벡터 ω_o와 최적의 바이어스 θ_o는

$$w_o = \sum_{i=SVs} \alpha_{io} d_i z(x_i) \quad (9)$$

$$\theta_o = -\frac{1}{2} \sum_{i=SVs} w_o d_i [K(x_i, x_r) + K(x_i, x_s)]$$

로 계산될 수 있다.

지금까지 살펴본 SVM을 이용한 패턴의 분류는 가장 가까운 패턴들로부터 거리가 최대인 초월면을 찾는 것으로 요약될 수 있다. 이를 위한 과정은 분류문제를 QP문제로 변환하여 해결하는 것이다. 그러나 QP과정에서는 계산적인 여러 가지 제약점이 있으며, 문제의 규모가 커질수록 그 제약점은 더욱 더 심화된다. 따라서 본 연구에서는 빠르고 구현성이 뛰어난 새로운 학습알고리즘을 제안한다.

3. 모멘트에 의한 수정된 kernel-adatron 알고리즘

Campbell등이 제안한 kernel-adatron알고리즘은 퍼셉트론과 유사한 알고리즘인 adatron과 비선형 특징공간에서의 동작을 위한 kernel을 조합한 하이브리드 학습기법이다. kernel-adatron알고리즘에서는 최대화를 위한 α_i의 변화량 $\delta\alpha_i$의 계산에 기울기상승법을 그대로 이용하고 있어 최적해로 수렴할 때 학습률 η에 따라서는 발진이 발생할 수도 있다. 따라서 발진을 막아 줄 수만 있다면 좀 더 빠르게 최적해로 수렴시킬 수 있을 것이다. 이를 위해 본 연구에서는 $\delta\alpha_i$를 계산할 때 이전 변화의 속성을 가지는 모멘트를 추가하였다. 이는 현재의 계산방향이 이전의 계산방향을 따르도록 하는 것이다. 이렇게 하면 kernel-adatron알고리즘이 가지는 우수한 구현성은 그대로 살리면서도 최적해로의 수렴에서 발생하는 발진을 억제시켜 좀 더 빠른 속도로 SVM을 학습시킬 수 있다. 제안된 모멘트를 이용한 $\delta\alpha_i$의 계산식은

$$\delta\alpha_i(t+1) = \eta\left(1 - \sum_{j=1}^{N} \alpha_j(t) d_j K(x_i, x_j) - \lambda(t) d_i\right) + m(\alpha_i(t) - \alpha_i(t-1)) \quad (10)$$

와 같으며, 여기서 m은 모멘트이다. 따라서 모멘트 항을 이용한 새로운 kernel-adatron 알고리즘의 과정을 정리하면 다음과 같다. 즉,

단계 1 : 라그랑지안 계수 $\alpha_i(0)$와 γ, 파라미터 c, 학습률 η, 모멘트 m, 그리고 알고리즘의 최대 반복수 t_{max}를 설정

단계 2 : i,j = 1에서 N까지 커널요소 $K(x_i, x_j)$를 계산

단계 3 : 알고리즘의 반복 수 t = 1에서 t_{max} 까지 다음 단계 4에서 8까지를 수행

단계 4 : t = 0이면, 라그랑지안 계수 λ(0) = γ로 t = 1이면, 라그랑지안 계수 λ(1) =-γ 로 그렇지 않으면 $\lambda(t) = \lambda(t-1) - \omega(t-1) \times \left(\frac{\lambda(t-1) - \lambda(t-2)}{\omega(t-1) - \omega(t-2)}\right)$을 계산

단계 5 : i = 1에서 N까지 다음 단계 6과 7을 수행

단계 6 : $M_i = \sum_{j=1}^{N} \alpha_j(t) d_j K(x_i, x_j)$를 계산

단계 7 : $\delta\alpha_i(t) = \eta(1 - M_i d_i - \lambda(t) d_i) + m(\alpha_i(t) - \alpha_i(t-1))$를 계산
$(\alpha_i(t) + \delta\alpha_i(t)) \leq 0$이면, $\alpha_i(t) = 0$
$0 < (\alpha_i(t) + \delta\alpha_i(t)) < c$이면, $\alpha_i(t) = \alpha_i(t) + \delta\alpha_i(t)$
$(\alpha_i(t) + \delta\alpha_i(t)) \geq c$이면, $\alpha_i(t) = c$로 한다.

단계 8 : $\omega(t) = \sum_{j=1}^{N} \alpha_j(t) d_j$를 계산

단계 9 : 반복수 t가 t_{max}를 넘거나 여유χ를 계산하여 근사적으로 1이면 학습을 종료하며, 그렇지 않으면 단계 3으로 간다.

$$\chi = \frac{1}{2}\left(\min_{(i|d_i=1,\ \alpha_i \leq c)} (M_i) - \min_{(i|d_i=-1,\ \alpha_i \leq c)} (M_i) \right)$$

제안된 알고리즘으로 얻어진 라그랑지안 계수 α_i와 λ의 값을 이용하면 시험패턴들을 분류할 수 있다. 이때 특징공간에서 최적으로 패턴들을 분리시킬 수 있는 초월면의 분류식은

$$f(x) = sign[\sum_{i \in SVs} \alpha_i d_i K(x, x_i) + \lambda] \qquad (11)$$

와 같다.

4. 시뮬레이션 결과 및 분석

제안된 kernel-adatron알고리즘을 이용한 SVM의 성능을 평가하기 위해서 실험에서 커널함수는 가우스(Gaussian)의 radial basis 함수 $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$을 이용하였다. 라그랑지안 계수 α_i의 초기값은 랜덤시드(random seed)를 이용하여 0과 1 사이의 임의의 값으로 설정하였으며, 알고리즘의 종료는 반복수가 20,000번(t_{max} = 20,000) 이상이거나 여유χ 가 1이상일 때로 하였다. 제안된 학습기법의 SVM을 실제 암환자를 2부류(초기와 악성)로 분류하는 문제를 대상으로 팬티엄II-MMX200(㎒) 컴퓨터를 이용하여 시뮬레이션한 후 그 타당성을 확인하였으며, Campbell등의 kernel-adatron알고리즘을 이용한 SVM에 의한 결과와도 비교 고찰하였다.

실험에 이용한 암환자 분류문제의 데이터는 미국 위스콘신 대학병원에서 제공되는 "Wisconsin breast cancer databases"를 대상으로 하였다[6]. 실험에 이용된 데이터들에서 환자의 식별번호인 첫 번째 속성은 제거하였고, 모든 조건속성들은 정수 10으로 나누었으며, 분류속성은 2(초기)이면 -1로 4(악성)이면 1로 표현하였다. 다음의 그림 1은 5명의 환자를 대상으로 본 실험에서 사용되는 데이터를 나타내는 예이다.

0.6, 0.2, 0.1, 0.1, 0.1, 0.1, 0.7, 0.1, 0.1, 1
0.5, 0.4, 0.4, 0.9, 0.2, 1, 0.5, 0.6, 0.1, 1
0.2, 0.5, 0.3, 0.3, 0.6, 0.7, 0.7, 0.5, 0.1, 1
0.6, 0.6, 0.6, 0.9, 0.6, ?, 0.7, 0.8, 0.1, -1 (실험에서 제외)
1, 0.4, 0.3, 0.1, 0.3, 0.3, 0.6, 0.5, 0.2, 1

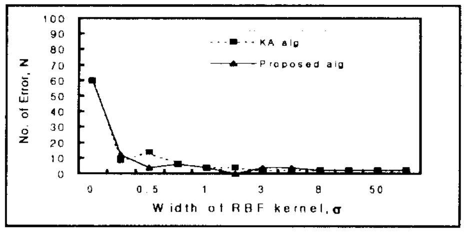
(그림 1) 5명의 환자에 대한 속성데이터 예

그림에서 ?는 그 속성 데이터가 알려져 있지 않아 실험에서는 학습 데이터로 이용되지 않은 경우이다.

그림 2는 라그랑지 계수 γ =1.0, 파라미터 c = 0.1, 학습률 η =3.0, 그리고 모멘트 m=0.1로 하여 radial basis 함수의 폭 σ 의 변화에 따른 시험데이터에 대한 일반화 성능을 보인 것이다. 그림에서 보면 제안된 알고리즘의 경우는 Compbell의 kernel-adatron알고리즘보다도 더욱 우수한 일반화 성능을 가지며, 그 성능변화가 상대적으로 적어 커널함수의 내부 파라미터인 σ 의 변화에 덜 의존함을 알 수 있다. 그림의 결과로부터 σ 의 값은 1.0에서 50사이의 값이 적당한 값이다.

표 1은 radial basis 함수의 폭 σ 의 변화에 따른 SVM의 학습알고리즘의 수행회수 N과 일반화 성능인 분류오차 E, 그리고 SV수 n을 각각 나타낸 것이다. 표에서도 제안된 알고리즘이 상대적으로 빠르면서도 우수한 일반화 성능이 있음을 확인할 수 있다. 특히, kernel-adatron에서 시험데이터를 완전하게 분류할 수 있는 경우가 하나도 나타나지 않는 것은 학습시에 미리 결정해야 하는 파라미터의 설정이 상대적으로 어려운 것으로 추측된다. 한편 함수의 폭 σ 의 변화에 따른 분류오차 E가 0.1미만인 경우에 대해서만 살펴보면, 제안된 알고리즘이 가장 넓은 변화폭을 가짐을 알 수 있으며, 학습속도 측면에서도 빠른 결과를 보이고 있다. 이는 제안된 알고리즘은 σ 값의 변화에 가장 영향을 적게 받기 때문이다. 동일한 일반화 성능에서 제안된 방법은 kernel-adatron알고리즘에 비해서는 약 1.5 배 이상 빠름도 알 수 있다. SV의 개수도 제안된 알고리즘이 상대적으로 적음을 알 수 있다. 따라서 제안된 알고리즘은 kernel-adatron알고리즘에서 발생되는 학습속도와 일반화 성능 두 가지 측면을 함께 개선시킬 수 있다.

여기서 학습오차와 일반화 능력 사이의 상관관계를 제어하는 파라미터인 α 의 상한을 결정하는 파라미터 c와 학습률 η 의 영향은 SVM의 전체 성능에는 크게 영향을 미치지 못함을 알 수 있다. 결국 제안된 방법은 Compbell의 kernel-adatron알고리즘에서 발생되는 발진현상을 해결할 수 있음을 확인할 수 있다.



(그림 2) radial basis 함수의 폭 σ 의 변화에 따른 SVM의 일반화 오차 E

σ	KA alg.			Proposed alg.		
	N	E	n(pn,nn)	N	E	n(pn,nn)
0	20000	0.6	40(20,20)	2	0.6	40(20,20)
0.2	4	0.08	20(2,18)	4090	0.12	44(28,16)
0.5	3	0.14	26(18,8)	18	0.04	18(8,10)
0.8	3	0.06	14(8,6)	2	0.06	14(8,6)
1.0	3	0.04	16(8,8))	2	0.04	16(8,8)
2.0	2	0.02	14(6,8)	2	0	22(10,12)
3.0	2	0.02	26(12,14)	2	0.04	14(6,8)
5.0	2	0.02	40(20,20)	2	0.04	18(8,10)
8.0	2	0.02	40(20,20)	2	0.02	26(12,14)
10	2	0.02	40(20,20)	2	0.02	40(20,20)
50	2	0.02	40(20,20)	2	0.02	40(20,20)
100	2	0.02	40(20,20)	2	0.02	40(20,20)

<표 1> 커널 함수의 폭 σ 의 변화에 따른 학습알고리즘의 수행회수와 일반화 성능

5. 결 론

본 연구에서는 기울기상승법에서 해의 변화에 과거의 속성을 반영하는 모멘트 항을 추가한 kernel-adatron알고리즘을 제안, 이를 SVM의 학습기법으로 이용하였다. 이렇게 하면 기울기상승법에서 발생하는 최적해로의 수렴에 따른 발진을 억제하여 그 수렴속도를 좀 더 개선시키는 모멘트의 장점과 kernel-adatron알고리즘의 구현 용이성을 그대로 살릴 수 있다.

제안된 학습알고리즘의 SVM을 실제 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, Campbell 등의 kernel-adatron알고리즘을 이용한 SVM과 비교할 때 학습시간과 시험데이터의 분류오차로 표현된 일반화 성능에서 더욱 우수한 결과를 보였다. 향후 제안된 알고리즘의 SVM을 좀 큰 규모의 문제와 회귀문제 등과 같은 좀 더 다양한 분야에의 연구가 계속 진행되어야 할 것이다.

참고문헌

[1] V.Vapnik, The Nature of Statistical Learning Theory, Springer Vergag, 1995
[2] M.O. Stitson, J.A.E. Weston, A. Gammerman, V. Vovk, and V.Vapnik," Theory of Support Vector Machines," Technical report CSD-TR-96-17, Royal Holloway, Univ. of London, May 1998
[3] C. Campbell and N. Cristianini, "Simple Learning Algorithms for Training Support Vector Machines, "http://lara.enm.bris.ac.uk/ cig/gzipped/KA-ieee.ps.gz
[4] E.E. Osuna, R. Freund, and F. Girosi," Training Support Vector Machines : An Application to Face Detection," Proc. Computer Vision and Pattern Recognition '97, Pueto Rico, June 1997
[5] E.E. Osuna, R. Freund, and F. Girosi," An Improved Training Algorithm for Support Vector Machines," Proc. of IEEE NNSP'97, New York, Sep. 1997
[6] W. H. Wolberg, "Wisconsin Breast Cancer Database," Univ. of Wisconsin Hospital,July 1992
(ftp://mlftp.diee.unica.it/pub/mac...sin/breast-cancer-wisconsin.data)