퍼지 및 지능시스템학회 논문지 2003, Vol. 13, No. 3, pp. 316-321

정해진 기저함수가 포함되는 Nu-SVR 학습방법

Nu-SVR Learning with Predetermined Basis Functions Included

김영일, 조원희, 박주영

Young-Il Kim, Won-Hee Cho, Jooyoung Park

고려대학교 제어계측공학과

Dept. of Control and Instrumentation Engineering, Korea University

요 약

최근들어, 서포트 벡터 학습은 패턴 분류, 함수 근사 및 비정상 상태 탐지 등의 분야에서 상당한 관심을 끌고 있다. 여러 가지 서포트 벡터 학습 방법들 중 누-버전(nu-versions)으로 불리는 방법들은 서포트 벡터의 개수를 제어해야할 필요가 있는 경우에는 특히 유용한 것으로 알려져 있다. 본 논문에서는, ν-SVR로 불리는 누-버전 서포트 벡터 학습 방법과 미리 정해진 기저함수를 모두 활용하는 함수 근사 문제를 고려한다. ε-SVR, ν-SVR 및 세미-파라메트릭 함수 근사 방법론 등을 복습한 후에, 본 논문은 정해진 기저함수를 이용할 수 있는 방향으로 기존의 ν-SVR 방법을 확장하는 방안을 제시한다. 그리고, 제안된 방법의 적용가능성이 예제를 통하여 보여진다.

Abstract

Recently, support vector learning attracts great interests in the areas of pattern classification, function approximation, and abnormality detection. It is well-known that among the various support vector learning methods, the so-called nu-versions are particularly useful in cases that we need to control the total number of support vectors. In this paper, we consider the problem of function approximation utilizing both predetermined basis functions and a nu-version support vector learning called ν-SVR. After reviewing ε-SVR, ν-SVR, and a semi-parametric approach, this paper presents an extension of the conventional ν-SVR method toward the direction that can utilize predetermined basis functions. Moreover, the applicability of the presented method is illustrated via an example.

Key Words : 서포트 벡터 학습, 함수 근사, 기저 함수

1. 서 론

최근 들어, 서포트 벡터 학습 방법은 관련 이론이 정립되고 각종 응용 사례가 보고되면서, 지능시스템 분야에서 매우 주요한 도구 중 하나로 자리를 잡아가고 있다[1-2]. 서포트 벡터 학습 방법은 최대 마진(maximal margin)을 이용하거나[3], RKHS(reproducing kernel Hilbert space) 위에서의 정칙화 기반(regularization- based) 함수 근사[4], 혹은 베이지안 확률이론[5] 등의 방법을 사용하여 유도할 수 있는데, 특히 일반화 능력(generalization capability)을 효과적으로 확보할 수 있는 관계로 패턴 분류[6], 함수 근사[7], 비정상 상태 탐지[8] 등의 여러 가지 문제에 다양하게 적용되어 왔다. 본 논문에서는, 이와 같이 다양한 분야에서 우수성을 인정받고 있는 서포트 벡터 학습 방법을, 정해진 기저함수가 포함되는 경우의 함수 근사 문제에 적용하는 방안을 확립해 보고자 한다.

접수일자 : 2003년 4월 28일
완료일자 : 2003년 6월 4일
본 연구는 고려대학교 특별연구비에 의하여 수행되었음

서포트벡터 학습 방법은 한 개의 은닉층을 갖는 MLP (multi-layer perceptron), 혹은 RBFN(radial basis function networks) 등의 신경망을 대상으로 하여 다음과 같은 장점을 갖는 해를 제공할 수 있는 학습방법이다: 우선, 신경망을 위한 고전적인 학습 방법들과는 달리, 은닉 노드의 개수를 자동으로 결정할 수 있다. 그리고, 기울기 강하 기법(gradient descent method) 등의 학습방법론이 가지고 있었던 지역적 최적해(local optimum)로 수렴하는 문제가 없어서 반드시 성능지수에 관한 전역적 최적해(global optimum)를 찾을 수 있다. 따라서, 연결강도(weights)의 초기 값에 따라 학습결과가 달라지는 문제점이 없다. 마지막으로, 유도과정이 통계적 학습이론(statistical learning theory)으로 설명될 수 있기 때문에 일반화 능력(generalization capability)이 우수한 결과를 학습 결과로 얻을 수 있다. 본 논문에서는 서포트 벡터 학습이 갖는 이러한 일반적인 장점에 추가해서, 주어진 문제에 대한 정보를 추가로 활용하는 방안에 대해서 고찰한다. 즉, 주어진 데이터에 대해서 일정한 성능지수를 최적화할 수 있는 근사함수(approximating function)를 찾는 문제에서, 문제의 특성상 정해진 기저 함수들(basis functions)이 근사함수에 포함되는 것이 바람직한 경우를 생각해보자. 서포트 벡터 학습 분야에서는 이같이 정해진 기저 함수를 포

함할 수 있는 서포트 벡터 학습 방법을 세미-파라메트릭 방법(semi-parametric approach)이라고 부르는데, 함수 근사를 위한 전형적인 서포트 벡터 학습 방법인 ε-SVR 기법에 세미-파라메트릭 방법을 적용한 연구로는 [10] 등이 있다. 본 논문에서는, 서포트 벡터가 전체 학습 데이터 중 차지하는 비율을 조절할 수 있는 ν-SVR 기법에 세미-파라메트릭 방법을 접목하는 문제를 고려하고, 간단한 예제를 대상으로 제안된 방법론의 응용 가능성을 살펴보았다.

본 논문의 구성은 다음과 같다. 2장에서는 서포트 벡터 학습을 이용한 함수 근사 문제에 대해 설명하고, 3장에서는 본 논문의 핵심 주제인 정해진 기저 함수가 포함된 ν-SVR 기법에 관련된 내용을 기술한다. 그리고, 4장에서는 제안된 방법의 적용 가능성을 보이기 위해 예제를 대상으로 시뮬레이션을 수행하고 그 결과를 분석한다. 마지막으로, 5장에서는 결론과 향후 과제 등을 제시한다.

2. 서포트 벡터 학습을 이용한 함수 근사

2. 1. ε-SVR

서포트 벡터 학습을 이용한 함수근사 문제는 보통 SVR(support vector regression)이라고 불린다[7]. 다음에서는 입실론(epsilon) SVR 방법이라고 불리는 ε-SVR 기법의 핵심을 간단히 소개하기로 한다. 주어진 학습데이터 $\{(x_i, y_i)\}_{i=1}^{m}$에 대해 입실론-무반응 오차(ε-insensitive error)[7]

$|y_i - f(x_i)|_\varepsilon \triangleq \max\{|y_i - f(x)| - \varepsilon, 0\}$의 총합을 작게 하는 매끄러운 근사함수인 $f(x) = \langle w, \phi(x)\rangle + b$ $f(x) = \langle w, \phi(x)\rangle + b$를 찾는 것이 ε-SVR의 목적이다. 여기에서, $\phi(x)$는 입력 벡터 x를 고차원 특징공간 F로 매핑한 결과임에 유의하자. 위에서 언급한 ε-SVR의 목적은, 다음과 같은 최적화 문제를 풀어서 달성할 수 있다[2]:

$$
\begin{aligned}
\min \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \qquad (1)\\
s.t. \quad & y_i - (\langle w, \phi(x_i)\rangle + b) \le \varepsilon + \xi_i\\
& (\langle w. \phi(x_i)\rangle + b) - y_i \le \varepsilon + \xi_i^*\\
& \xi_i,\ \xi_i^* \ge 0,\ \forall i
\end{aligned}
$$

여기에서, C는 상수로서 0보다 큰 값을 가지며, 함수 f의 완만함(flatness)과 학습 데이터에 대한 오차 ξ_i, ξ_i^* 사이의 상대적 중요도를 결정한다.

이러한 최적화 문제에 대응하는 라그랑제 함수(Lagrange function)는 다음과 같다:

$$
\begin{aligned}
L = & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \qquad (1)\\
& - \sum_{i=1}^{m}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, \phi(x_i)\rangle + b)\\
& - \sum_{i=1}^{m}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, \phi(x_i)\rangle - b)\\
& - \sum_{i=1}^{m}(\eta_i\xi_i + \eta_i^*\xi_i^*)\\
& (\text{단},\ \alpha_i^{(*)}, \eta_i^{(*)} \ge 0)
\end{aligned}
$$

여기에서 w, b, ξ_i, ξ_i^*는 주 변수(primal variables)이고, $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$는 라그랑제 승수(Lagrange multipliers)로 도입된 쌍대 변수(dual variable)이다. 문제 (2)의 최적 해는 주 변수와 쌍대 변수로 이루어지는 전체 변수 공간 내에서 안장점(saddle point)이 되므로[1], 다음 식을 이용하면 주 변수들을 소거할 수 있다.

$$
\begin{aligned}
& \frac{\partial L}{\partial w} = 0 \Leftrightarrow w - \sum_{i=1}^{m}\alpha_i\phi(x_i) + \sum_{i=1}^{m}\alpha_i^*\phi(x_i) = 0 \qquad (3)\\
& \therefore\ w = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i)\\
& \frac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0\\
& \frac{\partial L}{\partial \xi_i^{(*)}} = 0 \Leftrightarrow \alpha_i^{(*)} + \eta_i^{(*)} = C\\
& \therefore\ \alpha_i^{(*)} \in [0, C],\ \forall i
\end{aligned}
$$

(여기에서, (*) 표기는 *가 있는 경우와 없는 경우가 모두 해당됨을 의미함)

식 (3)의 결과를 라그랑제 함수에 대입하여 주 변수들을제거하면 다음과 같은 쌍대 문제(dual problem)를 얻게 된다.

$$
\begin{aligned}
\max\ \tilde{D} = & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle\phi(x_i), \phi(x_j)\rangle \qquad (4)\\
& + \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{m}(\alpha_i + \alpha_i^*)\varepsilon\\
s.t. \quad & \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0\\
& \alpha_i^{(*)} \in [0, C],\ \forall i
\end{aligned}
$$

여기에서, 관련 분야에서 널리 사용되고 있는 가우시안 RBF 커널(radial basis function kernel)을 이용한 특징 공간으로의 매핑함수 $\phi: R^n \to F$를 사용하면 다음과 같은 형태의 커널 트릭(kernel trick)이 성립된다[2]:

$$\langle\phi(x), \phi(y)\rangle = k(x, y) = \exp(-\|x - y\|^2/2\sigma^2)$$

그러므로 식 (4)의 쌍대 문제는 다음과 동치가 된다.

$$
\begin{aligned}
\max D = & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \qquad (5)\\
& + \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{m}(\alpha_i + \alpha_i^*)\varepsilon\\
s.t. \quad & \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0\\
& \alpha_i^{(*)} \in [0, C],\ \forall i
\end{aligned}
$$

이제, QP(quadratic programming) 형태로 표현된 쌍대 문제 (5)를 풀면 최적의 α_i와 α_i^*가 구해지고, 최적의 b값은 Kuhn-Tucker 조건[1]으로부터 얻어진다. 이에 따라, ε-SVR 기법을 통하여 얻은 근사 함수 f는, $w = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i)$가 성립함을 이용하면, 다음과 같이 표현할 수 있게 된다.

$$f(x) = \langle w, \phi(x)\rangle + b = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + b \qquad (6)$$

2.2. ν-SVR 학습방법

ε-SVR 학습방법에서 나타나는 문제점은 ε의 값을 학습하기 전에 정해주어야 한다는 것이다. 그러나, 대부분의 함수 근사 문제에서는 주어진 학습 데이터에 합당한 무반응 오차 범위 ε값을 미리 알 수 없다. 이러한 문제를 해결하기 위해

서 ν-SVR 학습방법[9]에서는 ε의 값을 최적화 문제의 변수들 중 하나로 인식하고, 다른 변수들과 함께 가능한 한 작은 ε의 값을 갖도록 최적화 문제에 포함시켜 표현함으로써, 함수 근사를 위한 최적화 문제를 해결함과 동시에 최적의 ε 값을 구하는 방법을 취한다. 또한, ν-SVR 학습방법은 ν값의 변화에 따라 서포트 벡터의 개수를 조절할 수 있는 장점이 있다. ν-SVR 기법을 위한 최적화 문제는 다음과 같이 구성된다:

$$\begin{aligned} \min \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^* + \nu\varepsilon) \\ s.t. \quad & y_i - (\langle w, \phi(x_i)\rangle + b) \le \varepsilon + \xi_i \\ & (\langle w, \phi(x_i)\rangle + b) - y_i \le \varepsilon + \xi_i^* \\ & \xi_i^{(*)},\ \varepsilon \ge 0,\ \forall i \quad (\text{단, } \nu \in (0,1]) \end{aligned} \tag{7}$$

이 문제의 목적 함수와 제약 조건으로부터, 문제 (7)에 대응하는 라그랑제 함수는 다음과 같음을 알 수 있다:

$$\begin{aligned} L = & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^* + \nu\varepsilon) - \gamma\varepsilon \\ & - \sum_{i=1}^{m}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, \phi(x_i)\rangle + b) \\ & - \sum_{i=1}^{m}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, \phi(x_i)\rangle - b) \\ & - \sum_{i=1}^{m}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ & (\text{단, } \alpha_i^{(*)}, \eta_i^{(*)}, \gamma \ge 0) \end{aligned} \tag{8}$$

그리고, 최적 해에 대한 안장점 조건을 이용하여 주 변수를 소거하면 다음과 같은 쌍대 문제를 얻게 된다.

$$\begin{aligned} \max D = & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ & + \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)y_i \\ s.t. \quad & \sum_{i=1}^{m}(\alpha_i + \alpha_i^*) \le Cm\nu,\quad \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i^{(*)} \in [0, C],\ \forall i \end{aligned} \tag{9}$$

쌍대 문제 (9)를 풀면, 최적의 α_i와 α_i^*가 구해지고, 이에 따르는 최적의 b와 ε값은 Kuhn-Tucker 조건을 이용하여 얻을 수 있다. 그리고, ν-SVR 학습방법을 통하여 얻은 근사 함수 f는, ε-SVR 기법의 경우와 마찬가지로 다음과 같이 표현할 수 있게 된다.

$$f(x) = \langle w, \phi(x)\rangle + b = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{10}$$

3. 정해진 기저함수가 포함되는 ν-SVR 학습방법

정해진 기저함수가 포함된 서포트 벡터 학습 방법이라함은, RBFN이나 MLP같은 신경망 기반 근사기에 정해진 독립 기저함수(independent basis function)인 $\{\phi_1(\cdot), \ldots, \phi_n(\cdot)\}$가 추가된 구조를 사용하여 주어진 학습 데이터를 근사하는 문제를 서포트 벡터 학습으로 찾는 푸는 것을 의미한다. 이 때 사용되는 근사 함수 f의 수식은 다음과 같이 표현될 수 있다[10]:

$$f(x) = \langle w, \phi(x)\rangle + \sum_{i=1}^{m}\beta_i\phi_i(x) \tag{11}$$

ε-SVR 기법과 유사한 방법을 사용하여 주어진 학습 데이터를 식 (11) 형태의 함수로 근사하는 방법은 [10]에서 연구된 바 있다. 앞에서 설명된 바와 같이, 본 논문의 주된 목적은 식 (11) 형태를 이용한 함수 근사 문제를 위한 ν-SVR 버전을 제시하는 것이다. 이러한 목적을 이루기 위해 다음에서는 우선 [10]에서 제시된 입실론 버전, 즉 세미-파라메트릭 ε-SVR 기법을 간단하게 살펴보기로 한다. 참고문헌 [10]에서는 식 (11)을 이용한 함수 근사를 위하여 다음의 최적화 문제를 활용하는 것을 제안하였다:

$$\begin{aligned} \min. \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=m}^{l}(\xi_i + \xi_i^*) \\ s.t. \quad & \langle w, \phi(x_i)\rangle + \sum_{j=1}^{n}\beta_j\phi_j(x_i) - y_i \le \varepsilon + \xi_i^* \\ & y_i - \langle w, \phi(x_i)\rangle - \sum_{j=1}^{n}\beta_j\phi_j(x_i) \le \varepsilon + \xi_i \\ & \xi_i, \xi_i^* \ge 0,\ \forall i \end{aligned} \tag{12}$$

위의 목적함수와 제약조건들로부터, 문제 (12)의 라그랑제 함수는 다음과 같이 나타낼 수 있다:

$$\begin{aligned} L = & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \\ & + \sum_{i=1}^{m}\alpha_i\Big\{y_i - \langle w, \phi(x_i)\rangle - \sum_{j=1}^{n}\beta_j\phi_j(x_i) - \varepsilon - \xi_i\Big\} \\ & + \sum_{i=1}^{m}\alpha_i^*\Big\{\langle w, \phi(x_i)\rangle + \sum_{j=1}^{n}\beta_j\phi_j(x_i) - y_i - \varepsilon - \xi_i^*\Big\} \\ & + \sum_{i=1}^{m}\eta_i(-\xi_i) + \sum_{i=1}^{m}\eta_i^*(-\xi_i^*) = 0 \\ & (\text{단, } \alpha_i^{(*)}, \eta_i^{(*)} \ge 0) \end{aligned} \tag{13}$$

그리고, 라그랑제 함수 (13)에서 주변수에 대한 편미분을 0으로 놓는 방법으로 주 변수를 소거하면 다음과 같은 쌍대 문제를 얻게 된다.

$$\begin{aligned} \max \quad & -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ & -\varepsilon\sum_{i=1}^{m}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*) \\ s.t. \quad & \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi_j = 0,\ \forall j \\ & \alpha_i, \alpha_i^* \in [0, C],\ \forall i \end{aligned} \tag{14}$$

이제, QP 형태인 쌍대 문제 (14)를 풀어주면 최적의 α_i와 α_i^*가 구할 수 있게 되고, 여기에 Kuhn-Tucker 조건을 이용하여 β_j값들을 마저 구해주면, 우리가 구하고자 하는 근사 함수가 다음 형태로 구해진다:

$$f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + \sum_{j=1}^{n}\beta_j\phi_j(x) \tag{15}$$

본 논문에서 추구하는 누-버전, 즉 세미-파라메트릭 ν-SVR 학습 방법은 기본적으로 2.2절의 ν-SVR 방법을 참고하여 위의 절차를 밟아 가면 유도할 수 있다. 식 (11) 형태의 근사함수를 이용하여 주어진 데이터를 ν-SVR 기법과 같은 원리로 학습하는 문제를 다루기 위해서는 다음과 같은 최적화 문제를 활용할 수 있다:

$$\min.\ \ \frac{1}{2}\|w\|^2 + C\sum_{i=m}^{l}(\nu\varepsilon + \xi_i + \xi_i^*) \tag{16}$$

$$s.t.\ \ \langle w, \phi(x_i)\rangle + \sum_{j=1}^{n}\beta_j\phi_j(x_i) - y_i \le \varepsilon + \xi_i^*$$

$$y_i - \langle w, \phi(x_i)\rangle - \sum_{j=1}^{n}\beta_j\phi_j(x_i) \le \varepsilon + \xi_i$$

$$\xi_i^{(*)}\ ,\ \varepsilon \ge 0\ ,\ \forall i\ (\text{단, } \nu \in (0,1])$$

이 문제를 같은 답을 갖는 유한 차원의 쌍대 문제로 전환시키기 위해서는, 우선 목적함수와 제약조건들을 이용하여 다음과 같은 라그랑제 함수를 구성하여야 한다.

$$\begin{aligned} L =\ & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\nu\varepsilon + \xi_i + \xi_i^*) + \gamma(-\varepsilon) \\ &+ \sum_{i=1}^{m}\alpha_i\Big\{y_i - \langle w, \phi(x_i)\rangle - \sum_{j=1}^{n}\beta_j\phi_j(x_i) - \varepsilon - \xi_i\Big\} \\ &+ \sum_{i=1}^{m}\alpha_i^*\Big\{\langle w, \phi(x_i)\rangle + \sum_{j=1}^{n}\beta_j\phi_j(x_i) - y_i - \varepsilon - \xi_i^*\Big\} \\ &+ \sum_{i=1}^{m}\eta_i(-\xi_i) + \sum_{i=1}^{m}\eta_i^*(-\xi_i^*)\ =\ 0 \end{aligned} \tag{17}$$

$$(\text{단, } \alpha_i^{(*)}, \eta_i^{(*)}, \gamma \ge 0)$$

문제 (16)의 최적 해는 주 변수와 쌍대 변수가 모두 포함된 전체 변수 공간 내에서 안장점이 되므로, 다음에 설명된 바와 같이 주 변수에 대한 편미분을 0으로 놓는 과정을 이용하여 주 변수를 소거할 수 있다.

$$\frac{\partial L}{\partial w} = 0 \Leftrightarrow w - \sum_{i=1}^{m}\alpha_i\phi(x_i) + \sum_{i=1}^{m}\alpha_i^*\phi(x_i) = 0 \tag{18}$$

$$\therefore\ \ w = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i)$$

$$\frac{\partial L}{\partial \varepsilon}\ =\ 0\ \Leftrightarrow\ Cm\nu - \gamma - \sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0$$

$$\therefore\ \ \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\ \le\ Cm\nu$$

$$\frac{\partial L}{\partial \beta}\ =\ 0\ \Leftrightarrow\ \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi_j(x_i) = 0$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = 0\ \Leftrightarrow\ C - \alpha_i^{(*)} - \eta_i^{(*)} = 0$$

$$\therefore\ \ \alpha_i^{(*)} \in [0, C]\ ,\ \forall i$$

이제, (18)의 결과를 식 (17)에 대입하여 주 변수들을 소거해주면, 다음과 같은 쌍대 문제를 얻게 된다.

$$\begin{aligned} \max\ & -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ & + \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*) \end{aligned} \tag{19}$$

$$\begin{aligned} s.t.\ \ & \sum_{i=1}^{m}(\alpha_i + \alpha_i^*) \le Cm\nu \\ & \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi_j = 0\ ,\ \forall j \\ & \alpha_i, \alpha_i^* \in [0, C], \forall i \end{aligned}$$

마지막으로, QP 형태인 문제 (19)를 풀어서 최적의 α_i와 α_i^*가 구해주고, Kuhn-Tucker 조건을 이용하여 β_j와 ε를 구하는 과정을 밟아주면, 우리가 구하고자 하는 근사 함수 f가 다음과 같은 형태로 얻어진다:

$$f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x_i, x) + \sum_{j=1}^{n}\beta_j\phi_j(x) \tag{20}$$

4. 모의 실험

본 논문에서 제안한 방법의 적용 가능성을 평가하기 위하여, 다음 식을 이용하여 생성된 학습 데이터를 근사하는 문제[10]를 고려해 보았다.

$$f(x) = \sin x + sinc(2\pi(x-5)) \tag{21}$$

학습 데이터로는, 식 (21)을 이용하여 출력된 데이터에 잡음(noise)이 추가로 포함되는 $\{(x_i, y_i) \mid y_i = f(x_i) + \xi_i\}$를 선택하였다. 이때, 가우시안 분포를 따르는 에러 항 ξ_i의 평균과 표준편차는 각각 0과 0.2로 설정하고, 가우시안 RBF 커널의 폭(width) σ 값으로는 0.25를 사용하였다. 그리고, 최적화 문제 (16)에 사용된 상수 C값은 0.5로 설정되었다. 본 논문에서 제안한 정해진 기저함수가 포함되는 ν-SVR 학습방법과 일반적인 ν-SVR 학습방법을 비교하기 위하여, ν값을 0.1에서부터 0.9까지 0.02의 간격으로 나누어 각각 학습시켜 보았다. 이 때, 본 논문에서 제안한 방법이 사용하는 기저함수의 집합으로는 $\{\sin x, \cos x, 1\}$이 사용되었다. 따라서 이 문제의 경우에는, 주어진 학습 데이터 이외에 각 주파수 1 [rad/sec]의 sin 혹은 cos 함수, 그리고 상수 함수가 주어진 데이터를 묘사하는데 중요한 역할을 할 수 있다는 사전 정보를 추가로 가지는 셈이 된다.

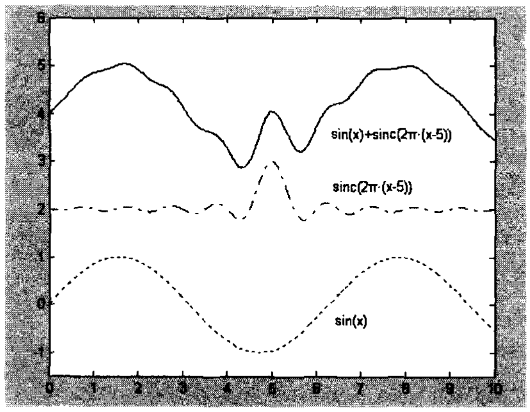


그림 1. 예제에서 고려된 함수
Fig. 1. Functions considered in the example

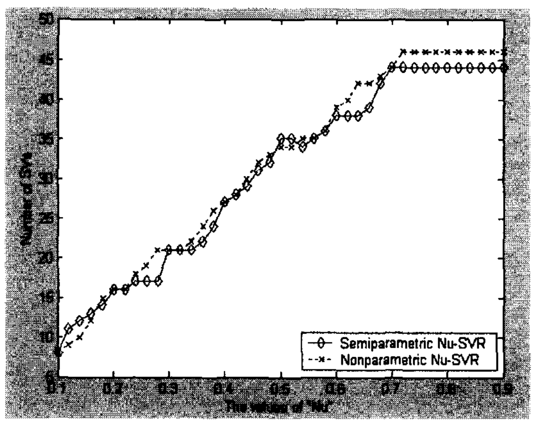


그림 2. ν 값의 변화에 따른 서포트 벡터의 개수
Fig. 2. Number of support vectors as a function of the nu-value

그림 1에서는 식 (21)에 표현된 함수를 각 구성 요소와 함께 그림으로 나타내었다. 그림 2에서는 ν값의 변화에 따른 서포트 벡터의 개수를 그림으로 나타내었다.

우리는 그림 2에서 ν값이 증가함에 따라 서포트 벡터의 개수도 증가함을 확인할 수 있다. 이들 그림에서 "semiparametric Nu-SVR"로 표시된 부분은 본 논문의 방법론, 즉, 정해진 기저함수가 포함된 ν-SVR 기법을 적용해서 구한 결과를 나타내고, "nonparametric Nu-SVR"로 표시된

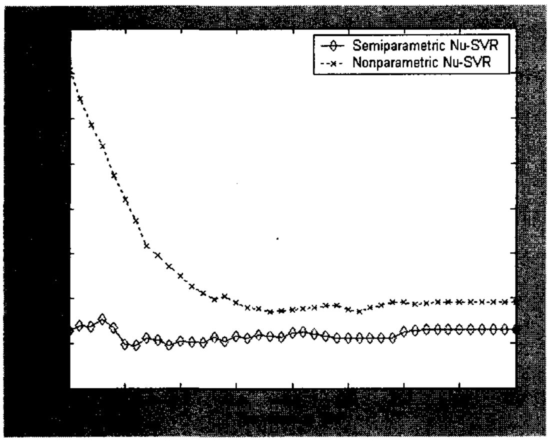


그림 3. ν 값의 변화에 따른 L_1 오차
Fig. 3. L1 error as a function of the nu-value

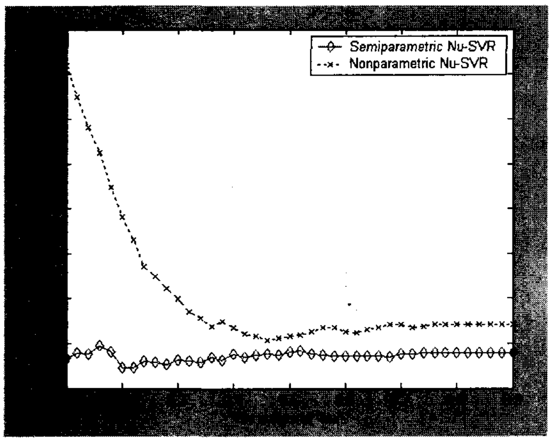


그림 4. ν 값의 변화에 따른 L_2 오차
Fig. 4. L2 error as a function of the nu-value

부분은 2.2절에서 소개된 일반적인 ν-SVR 기법에 따랐을 때 얻게되는 결과를 보인 것이다. 그림 3과 4에서는 ν값의 변화에 따른 L_1, L_2 오차 값을 비교하여 나타내었다. 여기에서, L_1, L_2 오차 값은 각각 다음을 의미한다:

$$L_1 \text{ 오차} = \frac{(|error_1| + \cdots + |error_n|)}{n}$$

$$L_2 \text{ 오차} = \sqrt{\frac{(error_1)^2 + \cdots + (error_n)^2}{n}}$$

이들 그림으로부터 우리는 ν값이 증가함에 따라 두 가지 학습방법의 학습결과로 얻어낸 오차 값들 사이의 차이가 줄어듬을 관찰할 수 있다. 그리고, ν값이 작은 경우 즉, 서포트 벡터의 개수가 작은 경우에는 정해진 기저함수가 포함되는

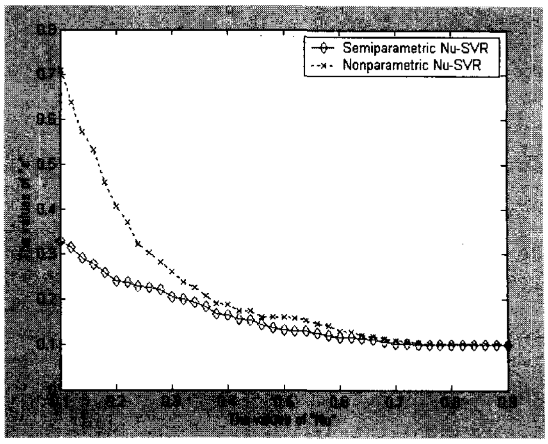


그림 5. 학습의 결과로 얻어진 ε 값들
Fig. 5. Values of ε resulting from training

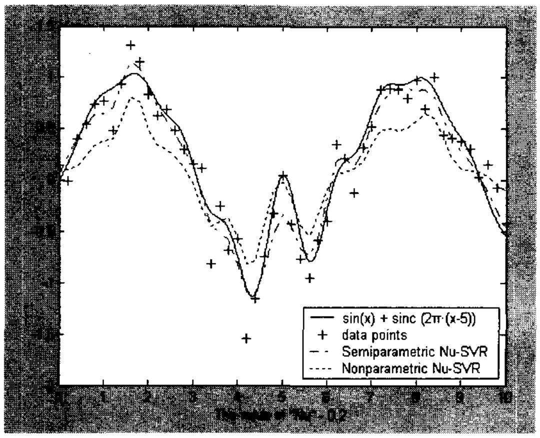


그림 6. ν= 0.2 일 때의 학습결과
Fig. 6. Training results when ν=0.2

ν-SVR 학습방법의 학습결과가 일반적인 ν-SVR 학습방법의 결과에 비해 오차 값이 작게 나타남으로서, 더 좋은 학습결과를 보임을 알 수 있다. 그림 5는 ν값에 변화에 따라 학습의 결과로 얻어진 ε값의 변화를 나타내었다. 앞에서의 결과와 마찬가지로, ν값이 작은 경우에는 정해진 기저함수가 포함되는 ν-SVR 학습방법의 결과가 일반적인 ν-SVR 학습방법에 비해 학습의 결과로 얻어진 최적화 된 ε값이 비교적 작게 나타남으로서, 더 좋은 결과를 보이고 있다. 그림 6은 그림 3과 4에서 학습의 결과로 얻어진 오차 값이 가장 작은 ν값 0.2에 대하여, 일반적인 ν-SVR 학습방법과 정해진 기저 함수과 포함된 ν-SVR 학습방법의 결과를 나타낸 그림이다. 정해진 기저 함수가 포함된 ν-SVR 학습방법을 이용한 학습결과 얻어진 함수 f의 그림이 일반적인 ν-SVR 학습방법의 결과로 얻어진 함수 f의 그림보다, 식 (21)에 나타낸 함수 $f(x)$를 더 잘 따라가는 것을 알 수 있다. 이상의 관찰로부터, 본 논문의 기저함수가 포함된 ν-SVR 학습방법은 일반적인 ν-SVR 방법과 마찬가지로 ν값의 변화에 따라 서포트 벡터의 개수가 변화하는 특성을 가지면서도, 일반적인 ν-SVR 학습에 비해 오차와 ε의 값을 상당히 줄일 수 있음을 알 수 있다.

5. 결 론

본 논문에서는 서포트 벡터 학습을 이용한 함수 근사 기법을 전반적으로 소개한 후에, 기존의 방법론들을 참고하여 정해진 기저함수가 포함된 ν-SVR 기법을 제시하고 필요한 중간 과정을 유도하여 보았다. 관련 연구에서 고려된 간단한 함수 근사 문제를 대상으로 비교·검토해 본 결과, 주어진 함수 근사 문제가 우리에게 기저함수에 대한 추가적인 힌트를 줄 경우에는 본 논문에서 제시한 방법론이 기존의 일반적인 ν-SVR 방법보다 더 좋은 결과를 제공함을 관찰할 수 있었다. 이와 관련하여 이후에 추가적으로 연구되어져야 할 과제로는, 다양한 예제에 대한 광범위한 실험을 통하여 제시한 방법론의 장단점을 파악하고, 여타 방법론과의 체계적인 성능 비교를 수행해보는 문제 등을 꼽을 수 있다.

참 고 문 헌

[1] N. Cristianini an J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods, Cambridge University* Press, 2000.

[2] B. Schölkopf and A. J. Smola, *Learning with kernels*, MIT Press, 2002.

[3] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181-201, 2001.

[4] F. Girosi, "An equivalence between sparse approximation and support vector machines," *Neural Computation*, vol. 10, no. 6, pp. 1455-1480, 1998.

[5] J. T. Kwok, "The evidence framework applied to support vector machines," *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1162-1173, 2000.

[6] C. Burges, "A tutorial on support vector machines for pattern recognition," *Knowledge Discovery and Data Mining*, vol. 2, no. 2, pp. 121-167, 1998.

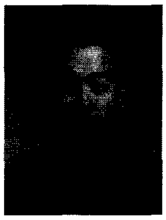
[7] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Neuro Colt Technical Report NC-TR-1998-030*, Royal Holloway College, University of London, UK, 1998.

[8] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, and A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, pp. 1443-1471, 2001.

[9] B. Schölkopf, A. Smola, R. Williamson, and P. L. Bartlett. "New support vector algorithms," *Neural Computation*, vol. 12, no. 5, pp. 1207-1245, 2000.

[10] A. Smola, T. Frieß, and B. Schölkopf, "Semiparametric Support Vector and Linear Programming Machines," *Nuero COLT Technical Report Series NC2-TR-1998-024*, August, 1998.

저 자 소 개



김 영 일 (Young-Il Kim)
2002년 : 고려대학교 제어계측공학과 졸업(학사)
2002년~현재 : 고려대학교 대학원 제어계측공학과 석사 과정
관심분야 : SVM, 신경망, 강화학습

Phone : (02) 922-0863
E-mail : sstuff@korea.ac.kr

조 원 희 (Won-Hee Cho)
2002년 : 고려대학교 제어계측공학과 졸업(학사)
2002년~현재 : 고려대학교 대학원 제어계측공학과 석사 과정

관심분야 : SVM 응용, 강화학습

Phone : (02) 922-0863
E-mail : loadneo@hanmail.net

박 주 영 (Jooyoung Park)
1983년 : 서울대학교 전기공학과 졸업(학사)
1985년 : 한국과학기술원 졸업(석사)
1985년3월-1988년7월 : 한국전력 월성 원자력발전소 근무
1992년 : University of Texas at Austin, 전기 및 컴퓨터공학과 졸업(박사)
1992년8월~1993년2월 : 한국전력 전력경제연구실 근무
1993년3월~현재 : 고려대학교 제어계측공학과 교수

관심분야 : 신경망이론, 퍼지제어, 비선형시스템
E-mail : parkj@korea.ac.kr