

# 분류 및 회귀문제에서의 분류 성능과 정확도를 동시에 향상시키기 위한 새로운 바이어스 스케줄링 방법

(A New Bias Scheduling Method for Improving Both  
Classification Performance and Precision on the Classification  
and Regression Problems)

김은미<sup>†</sup>   박성미<sup>†</sup>   김광희<sup>\*\*</sup>   이배호<sup>\*\*\*</sup>  
(Eun-Mi Kim)   (Seong-Mi Park)   (Kwang-Hee Kim)   (Bae-Ho Lee)

**요약** 분류 및 회귀문제에서의 일반적인 해법은, 현실 세계에서 얻은 정보를 행렬로 사상하거나, 이 진정보로 변형하는 등 주어진 데이터의 가공과 이를 이용한 학습에서 찾을 수 있다. 본 논문은 현실세계에 존재하는 순수한 데이터를 근원공간이라 칭하며, 근원 데이터가 커널에 의해 사상된 행렬을 이원공간이라 한다. 근원공간 혹은 이원공간에서의 분류문제는 그 역이 존재하는 문제 즉, 완전해가 존재하는 문제와, 그 역이 존재하지 않거나, 역의 원소 값들이 무한히 커지는 불량조건 혹은 특이조건인 두 가지 형태로 존재한다. 특히, 실제 문제에 있어서 완전 해를 가진 문제이기 보다는 후자에 가까운 형태로 나타나게 된다. 결론적으로 근원데이터나 이원데이터를 이용한 문제를 해결하기 위해서는 많은 경우에 완전 해를 갖는 문제로 변형시키는 정규화과정이 필요하다. 본 논문에서는 이러한 정규화 인수를 찾는 문제를 기존의 GCV, L-Curve, 그리고 이원공간에서의 데이터를 RBF 신경회로망에 적용시킨 커널 학습법에 대한 각각의 성능을 비교실험을 통해 고찰한다. GCV와 L-Curve는 정규화 인수를 찾는 대표적인 방법으로 두 방법 모두 성능면에서 동등하며 문제의 조건에 따라 다소 차이를 보인다. 그러나 이러한 두 방법은 문제해를 구하기 위해서는 정규화 인수를 구한후 문제를 재정의하는 이원적인 문제해결이라는 취약점을 갖는다. 반면, RBF 신경회로망을 이용한 방법은 정규화 인수와 해를 동시에 학습하는 단일화된 방법이 된다. 이때 커널을 이용한 학습법의 성능을 향상하기 위해, 전체학습과 성능의 제한적 비례관계라는 설정아래, 각각의 학습에 따라 능동적으로 변화하는 동적모멘텀의 도입을 제안한다. 동적모멘텀은 바이어스 학습을 포함한 방법과 포함하지 않은 방법에 각각 적용분석하였다. 끝으로 제안된 동적모멘텀이 분류문제의 표준인 Iris 데이터, Singular 시스템의 대표적 모델인 가우시안 데이터, 그리고 마지막으로 1차원 이미지 복구문제인 Shaw 데이터를 이용한 각각의 실험에서 분류문제와 회귀문제 양쪽 모두에 있어 기존의 GCV, L-Curve와 동등하거나 우수한 성능이 있음을 보인다.

**키워드** : RBF 네트워크, 신경회로망, 회귀분석, 분류, 커널, 정규화, 동적모멘트

**Abstract** The general solution for classification and regression problems can be found by matching and modifying matrices with the information in real world and then these matrices are learning in neural networks. This paper treats primary space as a real world, and dual space that primary space matches matrices using kernel. In practical study, there are two kinds of problems; complete system which can get an answer using inverse matrix and ill-posed system or singular system which cannot get an answer directly from inverse of the given matrix. Further more the problems are often given by the latter condition; therefore, it is necessary to find regularization parameter to change ill-posed or singular problems into complete system. This paper compares each

· 이 논문은 2004년도 전남대학교 특별연구사업비 지원에 의하여 연구되었음

† 학생회원 : 전남대학교 컴퓨터정보통신공학과  
em1342@hotmail.com

seongmi@chonnam.ac.kr

\*\* 정회원 : 국방과학연구소 연구원

kshift@hanafos.com

\*\*\* 정회원 : 전남대학교 컴퓨터정보통신공학과 교수  
bhlee@chonnam.ac.kr

논문접수 : 2005년 3월 16일

심사완료 : 2005년 8월 29일

performance under both classification and regression problems among GCV, L-Curve, which are well known for getting regularization parameter, and kernel methods. Both GCV and L-Curve have excellent performance to get regularization parameters, and the performances are similar although they show little bit different results from the different condition of problems. However, these methods are two-step solution because both have to calculate the regularization parameters to solve given problems, and then those problems can be applied to other solving methods. Compared with GCV and L-Curve, kernel methods are one-step solution which is simultaneously learning a regularization parameter within the learning process of pattern weights. This paper also suggests dynamic momentum which is learning under the limited proportional condition between learning epoch and the performance of given problems to increase performance and precision for regularization. Finally, this paper shows the results that suggested solution can get better or equivalent results compared with GCV and L-Curve through the experiments using Iris data which are used to consider standard data in classification, Gaussian data which are typical data for singular system, and Shaw data which is an one-dimension image restoration problems.

**Key words** : RBF, Regression, Classification, Kernel, Regularization, Dynamic Momentum. Neural Network

## 1. 서론

본 논문에서는 실세계에서 얻은 정보나 데이터를 단순 행렬로 변환한 공간을 근원공간(primary space)이라 하고 이 근원공간을 다차원 공간으로 확장 변환한 행렬 공간을 이원공간(dual space)이라 한다. 이러한 근원데이터와 이원데이터는 실세계에서 발생하는 문제를 컴퓨터를 이용, 주어진 문제를 해결할 수 있도록 가능케 하는 중요한 연결고리 역할을 한다. 즉, 목적치를 아는 근원공간 혹은 이원공간에서의 데이터가 존재한다면 이는 단순한 QP문제로 재정의 될 수 있으며 때에 따라서는 근원데이터 행렬이나 이원데이터 행렬의 역행렬을 취함으로써 간단하게 구하고자하는 해를 얻을 수 있는 완전해를 갖는 문제로 재 정의될 수 있다. 하지만 일반적인 경우, 근원행렬이나 이원행렬은 불량조건문제(ill-posed problem) 또는 특이 구조(singular system)의 형태로 나타나며 이 경우 역행렬을 취하게 되면 그 값이 너무 커지거나 무한하게 된다. 이러한 불량조건문제를 해결하기 위해 1963년 Tikhonov에 의해 정규화 이론(Regularization Theory)[1]이 제안되었으며 이로 하여금 불량조건문제를 구조화된 문제(well-posed problem)로 재정의 할 수 있게 되었다. 이러한 Tikhonov에 의한 정규화 이론을 바탕으로 이에 대한 많은 연구가 이상적인 정규화 인수(regularization parameter)를 찾기 위해 이루어졌으며 이러한 정규화인수를 찾는 대표적인 방법으로 GCV와 L-Curve가 등장하게 되었다. 그러나 이러한 방법들은 그 알고리즘과 성능의 우수함에도 불구하고 정규화 인수를 구하는 방법과 주어진 문제의 해를 구하는 방법이 이원화되었다는 결정적인 문제를 가지고, 실질적인 응용의 한계를 갖고 있다[2]. 이러한 이원화된 문제를 해결, 그 응용범위를 넓히기 위하여, 패턴 가중

치와 바이어스 항을 동시에 학습하는 새로운 정규화 인수를 찾는 방법인 바이어스 학습법이 커널과 신경망을 이용하여 제안되었다. 하지만 제안된 바이어스 학습법은 RBF 신경회로망에 기반으로 한 학습법으로 기존의 신경회로망에서 가지는 고전적인 문제, 즉, 지역 값과 과잉학습에 대한 문제를 잠정적으로 내포하며 이는 실제 학습에 있어 빈번하게 발생한다[3].

본 논문에서는 이러한 바이어스 학습법이 가지는 문제점을 해결하기 위한 방법으로 동적모멘텀을 제안한다. 동적모멘텀은 학습 성능과 학습회수를 제한적 비례관계로 정의하며 학습 정도에 따라 모멘텀의 값이 능동적으로 변화하는 방법으로 순차 학습법에서의 수렴시 발진을 억제하기 위해 과거의 속성 값을 반영하는 모멘텀의 응용적 스케줄링 방법이 된다. 본 논문에서는 이러한 동적모멘텀의 성능을 평가하기 위해 자기 다른 의미의 데이터 셋을 이용, 3단계에 의해 GCV, L-Curve와 함께 비교실험 하였다. 첫 번째 단계에서는 일반적인 성능 측정을 위하여, 분류문제의 표준데이터로 널리 사용되는 IRIS데이터를 이용한 분류성능과 각각의 방법으로 얻은 정규화 인수(Regularization parameter)를 이용한 오류율 측정을 위한 실험이다. 두 번째 단계에서는 보다 큰 문제에 대한 학습데이터와 실험데이터 각각에 대한 분류 성능 및 정규화 인수와 오류율에 관한 분석을 위한 실험이다. 이때 사용된 가우시안 데이터는 불량조건문제의 대표적인 데이터라 할 수 있다. 마지막 3단계에서의 실험은 실제 이미지 복구 문제를 이용한 각각의 알고리즘의 우수성에 대한 평가로 1차원 이미지 복구 문제인 Shaw데이터를 이용한 성능 평가이다. 끝으로 본 논문에서는 동적모멘텀이 적용된 커널 학습법 KLMSBLDM이 동적모멘텀을 사용하지 않은 학습법에 비해 보다 작

은 오류율을 가지며 분류·회계문제 성능을 동시에 향상하기 위해서는 동적모멘텀의 적용이 반드시 필요함을 제안한다.

## 2. 본 론

### 2.1 정규화(Regularization)

정규화는 불랑조건문제를 해결하기 위해 아래와 같은 비용함수를 갖는다[2,3].

$$\text{Primary Space} : \mathcal{J}(\mathbf{a}, \lambda) = \frac{1}{2} \epsilon^2 + \lambda \frac{1}{2} \mathbf{a}' \mathbf{a}, \quad (1)$$

$$\epsilon = \mathbf{d} - \mathbf{Y}' \mathbf{a}$$

$$\text{Dual Space} : \mathcal{J}(\boldsymbol{\alpha}, \lambda) = \frac{1}{2} \epsilon^2 + \lambda \frac{1}{2} \boldsymbol{\alpha}' \mathbf{K} \boldsymbol{\alpha}, \quad (2)$$

$$\epsilon = \mathbf{d} - \mathbf{K} \boldsymbol{\alpha}$$

아래와 같이 정의된 식에서

$$\mathbf{K} \boldsymbol{\alpha} = \mathbf{d} \quad (3)$$

위 식이 비 특이 구조(non-singular system)일 경우 완전해법이 존재하는 다음과 같은 식을 얻게 된다.

$$\boldsymbol{\alpha} = \mathbf{K}^{-1} \mathbf{d} \quad (4)$$

식 (2)를 라그랑지안 승수법으로 해를 구하면 아래와 같다[3].

$$L(\boldsymbol{\alpha}) = -\frac{1}{2} \lambda \boldsymbol{\alpha}' \boldsymbol{\alpha} - \lambda \frac{1}{2} \boldsymbol{\alpha}' \mathbf{K} \boldsymbol{\alpha} + \lambda \boldsymbol{\alpha}' \mathbf{d} \quad (5)$$

$$\frac{\partial L}{\partial \boldsymbol{\alpha}} = -\lambda(\lambda \boldsymbol{\alpha} - \mathbf{K} \boldsymbol{\alpha} + \mathbf{d}) = 0 \quad (6)$$

$$(\mathbf{K} + \lambda \mathbf{I}) \boldsymbol{\alpha} = \mathbf{d} \quad (7)$$

이때 정규화 인수  $\lambda$ 는  $(\mathbf{K} + \lambda \mathbf{I})$ 이 충분히 역이 가능한 스칼라(scalar) 값을 가진다. 따라서 위 식은 아래와 같이  $\boldsymbol{\alpha}$ 의 해를 구할 수 있다.

$$\boldsymbol{\alpha}_{reg} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{d} \quad (8)$$

이때 식 (4)는 아래 식이 된다.

$$\mathbf{d} = \mathbf{g}_k(\mathbf{y}) = \mathbf{K} \boldsymbol{\alpha}_{reg} \quad (9)$$

### 2.2 GCV(Generalized Cross-Validation)

$F_\lambda(\mathbf{x}_k)$ 는 임의의 정규화 인수(Regularization Parameter)  $\lambda$ 에 의해 회귀 원함수  $f(\mathbf{x})$ 를 정규화한 것으로 아래와 같이 정의된다[3].

$$F_\lambda(\mathbf{x}_k) = \sum_{i=1}^N a_{ki}(\lambda) d_i \quad (10)$$

$$\mathbf{F}_\lambda = \mathbf{A}(\lambda) \mathbf{d} \quad (11)$$

이때 대칭행렬(Symmetric Matrix)  $\mathbf{A}(\lambda)$ 를 영향행렬(Influence Matrix)이라 한다.

위 식을 정규화에 적용하면 아래와 같다.

$$\mathbf{F}_\lambda = \mathbf{K} \boldsymbol{\alpha}_{reg} \quad (12)$$

$$= \mathbf{K}(\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{d}$$

이고

$$\mathbf{A}(\lambda) = \mathbf{K}(\mathbf{K} + \lambda \mathbf{I})^{-1} \quad (13)$$

최적화된 정규화 인수  $\lambda$ 는 아래와 같이 주어진 회귀 원 함수  $f(\mathbf{x})$ 에 대한 평균 제곱 오류(average square error)를 최소화한다.

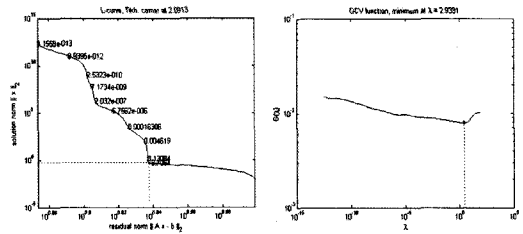
$$R(\lambda) = \frac{1}{N} \sum_{i=1}^N [f(\mathbf{x}_i) - F_\lambda(\mathbf{x}_i)]^2 \quad (14)$$

하지만 일반적으로 회귀 원 함수  $f(\mathbf{x})$ 는 알 수 없다. 대신 소량의 백색 잡음(white noise)이 첨가된 함수 값  $\mathbf{d}$ 를 사용하는데 정규화 인수에 대하여 평균제곱 오류는 단조증가 함수임으로 가중치를 부여하여 GCV함수를 유도하였다[3]. 최적화된 정규화 인수  $\lambda$ 는 주어진 입력 패턴에 대해 아래의 GCV 목적 함수를 최소화한다[3].

$$V(\lambda) = \frac{\frac{1}{N} \|(\mathbf{I} - \mathbf{A}(\lambda)) \mathbf{d}\|^2}{\left[ \frac{1}{N} \text{tr}[\mathbf{I} - \mathbf{A}(\lambda)] \right]^2} \quad (15)$$

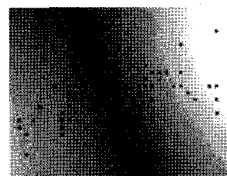
### 2.3 L-Curve

근원공간(primary space)에서  $\log \|\mathbf{a}_{reg}\|_2$ 과  $\log \|\mathbf{Y}' \mathbf{a}_{reg} - \mathbf{d}\|_2$ 가 이루는 그래프 또는 이원공간(dual space)에서  $\log \|\boldsymbol{\alpha}'_{reg} \mathbf{K} \boldsymbol{\alpha}_{reg}\|_2^{1/2}$ 과  $\log \|\mathbf{K} \boldsymbol{\alpha}_{reg} - \mathbf{d}\|_2$ 가 이루는 그래프는 아래 그림 1의 L자 모양의 곡선이 된다. 이때 최적화된 정규화 인수는 수직 좌표의 섭동(perturbation)오차와 수평 좌표의 잔여오차간(residual error)의 균형을 가장 잘 이룰 수 있는 점에서 정의된다. 이 점은 위의 그림에서 수직과 수평 그래프가 이루는 코너에 존재한다. L-Curve에서는 최적화된 정규화

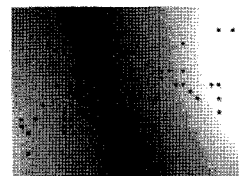


가. L\_curve

나. GCV



다. KLMSE



라. KLMSEDM

그림 1 Iris data(가, 나, 다, 라)

인수에 해당하는 점을 근사적으로 찾기 위해 코너에서 최대 곡률을 가진 점을 계산한다. 결국 L-Curve방법에서는 코너에서 최대 곡률을 가진 점의 정규화 인수가 최적의 정규화 인수가 된다[2].

**2.4 바이어스 학습(Bias Learning)**

바이어스 항을 포함하는 RBF 네트워크의 구조는 아래와 같은 과소결정 선형 구조(Underdetermined Linear System)이다[3].

$$\mathbf{K}\alpha + \alpha_0 = \mathbf{d} \tag{16}$$

정규화 RBF 네트워크의 구조는 아래와 같은 과소결정 비선형 구조(Underdetermined Nonlinear System)이다[3].

$$(\mathbf{K} + \lambda \mathbf{I})\alpha = \mathbf{d} \tag{17}$$

릿지 회귀분석(Ridge Regression)은 입력 패턴  $\mathbf{x}_i$ 와 출력  $d_i$ 에 대해 아래와 같이 정의된다[3].

$$d_i = f(\mathbf{x}_i) + \epsilon_i \tag{18}$$

$$E[\epsilon_i] = 0 \text{ for all } i$$

이때 식 (16)과 식 (17)의 바이어스 항  $\alpha_0$ 와 정규화 인수  $\lambda$ 의 관계를 정립하면 아래와 같다.

$$\mathbf{d} = \mathbf{K}\alpha + \lambda\alpha \tag{19}$$

$$= \mathbf{K}\alpha + \xi$$

$$= \mathbf{K}\alpha + \bar{\xi} + (\xi - \bar{\xi})$$

$$= \mathbf{K}\alpha + \alpha_0 + \epsilon$$

$$\xi \equiv \lambda\alpha, \tag{20}$$

$$\bar{\xi} \equiv \frac{\sum \xi_i}{N} = \frac{\lambda \sum \alpha_i}{N},$$

$$\alpha_0 \equiv \bar{\xi},$$

$$\epsilon \equiv (\xi - \bar{\xi})$$

바이어스 항  $\alpha_0$ 와 정규화 인수  $\lambda$ , 가중치  $\alpha$ 의 관계를 정립하면 다음과 같다.

$$\alpha_0 = \bar{\xi} \equiv \frac{\sum \xi_i}{N} = \frac{\lambda \sum \alpha_i}{N} \tag{21}$$

또는

$$\lambda = \frac{\alpha_0}{\sum \alpha_i} N \tag{22}$$

기존의 정규화 네트워크에서 문제의 해는 다음과 같다[3].

$$F_\lambda(\mathbf{y}) = \alpha_{\text{reg}}^\dagger \mathbf{K}(:, \mathbf{y}) \tag{23}$$

수정된 정규화 네트워크의 해는 다음과 같다.

$$F_\lambda(\mathbf{y}) = \alpha_{\text{reg}}^\dagger \mathbf{K}(:, \mathbf{y}) + \alpha_0 \tag{24}$$

여기에서  $\alpha_{\text{reg}} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{d}$ 이 된다. 실제 바이어스 학습(Bias Learning)에서는  $\alpha$ 와  $\alpha_0$ 가 확장 커널을 이용하여 동시에 학습하게 된다.

패턴 행렬은 바이어스  $\alpha_0$ 를 학습하기 위해 아래와 같이 확장하였다.

$$\mathbf{H}^t \omega = \mathbf{d} \tag{25}$$

$$\mathbf{H} \equiv [\mathbf{K} \text{ ones}(N, 1)]^t, \quad \omega \equiv \begin{bmatrix} \alpha \\ \alpha_0 \end{bmatrix}$$

이상과 같은 방법으로 바이어스  $\alpha_0$ 는 패턴 가중치  $\alpha$ 가 학습하는 동안 함께 학습함으로써 기존의 사용자 지정 방식의 정규화 인수는 패턴가중치와 함께 학습하게 된다. 이때 확장 커널 행렬  $\mathbf{H}^t \omega = \mathbf{d}$ 는 입력 공간  $\mathbf{Y}^t \alpha = \mathbf{d}$ 와 대응한다.

**2.5 동적모멘텀(Dynamic Momentum)**

동적모멘트를 이용한 학습 방법은 학습 초기 모멘트의 크기를 크게 하여 수렴 속도를 높였으며 점차 학습 수행 횟수가 많아져 수렴에 가까워질수록 모멘트의 크기는 정밀하게 조정된다. 따라서 조정된 모멘트의 크기가 초기 최대 모멘트 크기보다 작은 범위 내에서 조정되게 된다[4].

$$\mathbf{M} = \frac{m(k+1)}{\tau} \tag{26}$$

$$\text{if } \mathbf{M} > m$$

$$\mathbf{M} = 0;$$

$$\tau = r^2$$

여기서  $\tau$ 는 초기 모멘트의 값의 크기를 최소화하기 위해 사용한 제어 변수로 예상된 최초 한계치에서의 학습 횟수가 되며  $m$ 은 기존의 상수 모멘트로 동적 모멘트를 이용한 학습방법에서는 적용될 모멘트의 상한 값이 된다. 결국 동적모멘트  $\mathbf{M}$ 의 값이 모멘트의 상한 값인  $m$ 보다 클 경우 동적모멘트를 다시 0으로 하고  $\tau$ 를  $r^2$ 으로 다음번의 예상된 한계치를 최대로 둬으로써 예상된 한계치 내에서 다시 모멘트를 증가하면서 학습하게 한다.

즉,  $\mathbf{M}$ 이 단조 증가되는 것을 제한함으로써 기존의 상수 모멘트인  $m$ 을 초기 선언된 모멘트 내에서 가변적 변화가 가능하도록 하였다. 이때 처음 정의된 한계치에서의 모멘트의 증가는  $\frac{m(k+1)}{\Delta r}$ 의 증가치를 가지며 학습하게 된다. 또한 기존의 모멘트 값의 변경에만 의존하던 모멘트 항의 정의에 예상된 한계치에 대한 학습 횟수를 추가함으로써 보다 적극적인 모멘트 스케줄링을 정의 할 수 있다[4-7].

### 3. 실험

#### 3.1 실험환경

실험을 위한 시스템의 구현은 Matlab을 이용하였으며 본 실험에서는 사용한 데이터는 분류문제의 표준 데이터로 널리 알려진 IRIS 데이터, 보다 큰 문제에 관한 성능 분석을 위한 특이행렬의 전형인 가우시안 데이터, 그리고 마지막으로 1차원 이미지복구문제인 SHAW 데이터를 이용한 3단계 실험으로 구성된다. IRIS데이터는 3개의 클래스(Setosa, Versicolour, Virginica)로 구성되며, Setosa와 다른 두 클래스는 완벽한 선형분류가 가능하며, Versicolour과 Virginica는 완벽한 선형·비선형 분류가 불가능한 분류문제가 된다. 각 클래스의 개수는 50개로 전체 150개의 데이터를 가지며 각 데이터는 다시 sepal length, sepal width, petal length, petal width로 나누어진다. 이 실험에서는 선형분류가 불가능한 Versicolour과 Virginica의 Sepal 데이터를 실험 데이터로 사용하였다. 2단계 실험에서 사용된 가우시안 데이터는 Matlab에서 지원하는 가우시안 분포 생성함수 normrnd()를 사용하여 학습 데이터 500개와 테스트 데이터 1000개를 생성 사용하였다. 이때 사용된 평균 벡터와 분산은 첫번째 클래스에서 [0 0]과 [1 1]을 두 번째 클래스에서 [2 0]과 [2 2]를 사용하였으며, 실험결과에 베이지한 결정정제를 도식화함으로써 베이지안 분류기의 분류성능과 비교·분석할 수 있도록 하였다. 이 실험을 통하여 GCV와 L-Curve, 그리고 커널 학습법이 실험데이터 뿐 만 아니라 일반적인 문제에게까지 확대될 수 있음을 증명하고, 실제 특이치 행렬인 가우시안 데이터에서 정규화가 반드시 필요함을 실험을 통하여 알아본다. 그리고 마지막 SHAW는 1차원 이미지 복구 문제로, 그 문제의 형태가 이전의 분류문제와 달리 전형적인 회계문제로 본 논문에서 제안된 동적모멘텀의 학습법이 실제 문제에서도 동일한 성능 향상이 있음을 보여준다. 이하, 바이어스를 적용한 커널 학습법을  $KLMS_{BL}$ , 동적모멘트를 적용한 것을  $KLMS_{BLDM}$ 라 한다.

#### 3.2 실험결과

IRIS 데이터를 이용한 1단계 실험에서는 각각의 알고리즘의 이용하여 정규화 인수를 구하고 결정된 정규화 인수를 식 (8), (21) 및 (22)를 이용하여 바이어스와 오류율을 산출하였다. 이러한 방식으로 계산되어진 데이터를 순수한 학습에 따라 구한 데이터와 구분하기 위해 학습법의 뒤에 Inverse 표기하였다. 이때 사용된 RMS 오류 계산 방법은 다음과 같다.

RMS Error =

$$\sqrt{\frac{\|d - K\alpha\|^2}{N}} \quad (27)$$

RMS Error with Bias =

$$\sqrt{\frac{\|d - (K\alpha + \alpha_0)\|^2}{N}} \quad (28)$$

정규화 인수인  $\lambda$ 의 크기는  $KLMS_{BL} > KLMS_{BLDM} > GCV > L\text{-Curve}$  순으로 크게 나왔으며, 각각의 알고리즘이 산출해내는 정규화 인수는 모두 다른 것으로 나타났다. 바이어스 항의 크기는  $KLMS_{BL} > KLMS_{BL} \text{Inverse} > GCV > KLMS_{BLDM} > L\text{-Curve} > KLMS_{BLDM} \text{Inverse}$  순으로 나타났으며 바이어스의 크기와 정규화 인수와의 직접적인 연관은 없어 보이나 대체적으로 정규화 인수가 클수록 바이어스 항도 커지는 것을 볼 수 있다. 성능 면에서는 근소한 차이가 있으나 모든 알고리즘이 74%이상의 분류성능을 가지는 것으로 나타났으며 이는 Iris의 최적의 분류성능인 70~80%에 근접함으로써 성능 면에서는 동일하다고 볼 수 있다.  $KLMS_{BL}$ 을 제외한 나머지 알고리즘들은 바이어스 항을 포함한 것이 바이어스 항을 포함하지 않은 것에 비해 전체적으로 정확도가 높아진 것을 알 수 있으며, 특히  $KLMS_{BL}$ 은 원래 학습의 결과와 정규화 인수로만 산출해낸 결과가 다른 알고리즘과 대비 심한 괴리를 보이며, 이는 비록 분류성능에서 다른 알고리즘에 비해 75%라는 우수한 성능을 가졌으나 트레이닝 데이터에 대한 과잉 학습의 결과로 볼 수 있다. 이것은 동적모멘텀을 적용한  $KLMS_{BLDM}$ 과는 다른 관점으로 분류되어 적야한다.  $KLMS_{BLDM}$ 은 같은 75%의 분류성능을 가지나 순수 학습과 Inverse를 이용한 방식에서 RMS 오류 및 바이어스 항을 포함한 RMS 오류의 실험결과 오류율에 대한 변화가 GCV 및 L-Curve와 유사한 성능을 보이며 두 방식에 오류율의 차이가 거의 없는 것으로 볼 때 과잉 학습의 결과가 아니라는 것을 알 수 있다. 즉, 앞서 밝힌 바와 같이  $KLMS_{BL}$ 은 신경망의 고전문제인 지역값과 과잉학습의 문제를 극복하지 못하는 것을 알 수 있으며 제안된 동적모멘텀이 이러한 문제를 효과적으로 해결할 수 있음을 실험을 통하여 알 수 있다.

가우시안 데이터를 이용한 2단계 실험은 보다 큰 규모의 불량조건에 관한 문제로 정의되며 실험결과  $KLMS_{BL}$ 과  $KLMS_{BL} \text{Inverse}$ 를 제외한 다른 알고리즘들은 비슷한 수준의 RMS오류를 가진 것으로 나타났다.

특히, 2단계 실험에서도 IRIS를 이용한 실험에서와 마찬가지로  $KLMS_{BL}$ 의 순수학습에 의한 RMS오류 및 성능의 평가와 정규화 인수를 이용하여 계산한 오류 및 성능이 다른 방법들과는 달리 많은 차이를 보이며, 이는  $KLMS_{BL}$ 이 여전히 학습 데이터에 과잉학습 하는 것을 볼 수 있다. 분류성능에서 학습데이터와 실험데이터의 분류성능이 82.6% 와 83%로 안정된 모습을 보이기 하나 이는 가우시안 데이터의 문제 형태에 따른 실험의

표 1 Results of the regularization parameter using Iris data

	$\lambda$	bias	performance(%)	RMS Error	RMS Error with Bias
1	2.091283e+000	2.077852e-003	74	8.706841e-001	8.706816e-001
2	2.933131e+000	2.491655e-003	74	8.745416e-001	8.745381e-001
3	9.111957e+000	2.679960e-002	75	8.658558e-001	8.666892e-001
4	9.111957e+000	3.951047e-003	74	9.029183e-001	9.029097e-001
5	6.405503e-001	2.224852e-003	75	8.665070e-001	8.666279e-001
6	6.405503e-001	9.800551e-004	74	8.655781e-001	8.655776e-001

●L-Curve ● GCV ● KLMSBL  
 ●KLMSBL Inverse ●KLMSBLDM ●KLMSBLDM Inverse

표 2 Results of the regularization parameter using Gauss data

	$\lambda$	bias	PF		
			Tr	Tr	Tr
1	1.169441e+001	-5.311977e-002	82	7.444592e-001	7.425617e-001
			Ts	Ts	Ts
			82.3	7.352792e-001	7.332775e-001
			Tr	Tr	Tr
2	9.714378e+000	-4.756501e-002	82	7.401309e-001	7.386009e-001
			Ts	Ts	Ts
			82.4	7.312036e-001	7.285924e-001
			Tr	Tr	Tr
3	3.932931e+001	-2.562286e-001	82.6	7.981821e-001	7.287200e-001
			Ts	Ts	Ts
			83	7.960949e-001	7.169601e-001
			Tr	Tr	Tr
4	3.932931e+001	-4.457293e-002	81	7.951101e-001	7.900969e-001
			Ts	Ts	Ts
			81.8	7.861331e-001	7.817757e-001
			Tr	Tr	Tr
5	2.294522e+001	-1.772471e-001	82.6	7.681478e-001	7.297773e-001
			Ts	Ts	Ts
			83.2	7.638209e-001	7.186710e-001
			Tr	Tr	Tr
6	2.294522e+001	-7.437588e-002	81.6	7.676260e-001	7.640144e-001
			Ts	Ts	Ts
			81.9	7.580675e-001	7.547165e-001
			Tr	Tr	Tr

●L-Curve ● GCV ● KLMSBL  
 ●KLMSBL Inverse ●KLMSBLDM ●KLMSBLDM Inverse  
 PF : performance(%), Tr : Training data, Ts : Testing data  
 RMSE : Root Mean Square Error  
 RMSEB : Root Mean Square Error with Bias

결과로 보이며 비록 비슷한 수준의 분류성능이 학습 및 실험데이터에서 나왔다고 하나, 정규화 인수를 이용한 문제해결에서 RMS 오류를 순수 학습과 비교할 경우 다른 알고리즘에 비해 큰 값을 갖게 된다. 결과적으로 이는 KLMS<sub>BL</sub>이 학습데이터에 관해 과잉 학습되었다는 것을 간접적으로 알 수 있게 한다. 결론적으로 2단계 실험에서는 KLMS<sub>BL</sub>이 여전히 학습데이터에 관해 과잉 학습되고 있음을 증명하였으며 제안된 모멘텀을 이용한 학습법이 규모가 큰 불건전조건하에서도 안정적으로 분

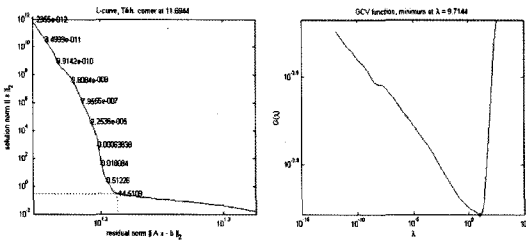
류 성능 향상과 회귀 오류율의 감소에 도움을 주고 있음을 알 수 있다.

마지막 실험으로 1차원 이미지 복구 문제인 Shaw 데이터틀 이용한 실험에서는 L-Curve와 역행렬을 이용한 데이터를 제외한 다른 알고리즘들이 유사한 결과를 보이고 있으며 이전의 실험결과에서 GCV와 L-Curve 그리고 동적모멘텀을 이용한 방법들이 각 문제에 관한 오류율의 성능 순위가 각기 다르게 나타나는 점으로 L-Curve, GCV 그리고 모멘텀을 이용한 방법은 동일한

표 3 Results of the regularization parameter using Shaw data

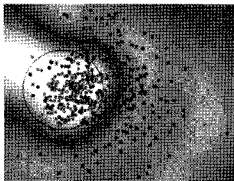
	$\lambda$	Bias	RMSE	RMSEB
①	N/A	N/A	4.877630e+002	N/A
②	6.436940e-008	2.105983e-007	2.439225e-006	2.430389e-006
③	3.183243e-014	-1.764525e-013	1.816696e-002	1.816696e-002
④	5.592667e-002	2.531608e-002	4.195736e-002	4.735803e-002
⑤	5.592667e-002	2.651775e-002	4.085471e-002	3.107920e-002
⑥	4.853420e-002	2.194394e-002	3.287784e-002	3.656827e-002
⑦	4.853420e-002	2.333542e-002	3.612810e-002	2.758075e-002

① L-Curve ② GCV ③ KLMSBL ④ KLMSBL Inverse  
 ⑤ KLMSBLDM ⑥ KLMSBLDM Inverse ⑦ KLMSBLDM Inverse  
 RMSE : Root Mean Square Error  
 RMSEB : Root Mean Square Error with Bias

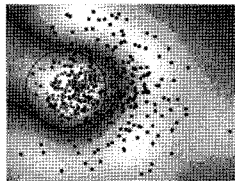


가. L-Curve

나. GCV



다. KLMSBL



라. KLMSBLDM

그림 2 Gaussian data(가, 나, 다, 라)

성능을 가지는 것으로 볼 수 있으며, 전체 실험을 통해서 모멘텀을 적용한 학습법이 분류 및 회귀문제에 있어 성능 향상을 가져온 것으로 나타났다.

결론적으로 커널을 이용한 학습법에 있어서 모멘텀의 적용은 필수적이며 모멘텀을 적용한 학습법은 문제의 형태에 따라 기존의 정규화 방법들보다 우수하거나 동등한 성능을 가진 것을 알 수 있다.

#### 4. 결론

기존의 L-Curve[2], GCV[3] 및 커널 학습법의 하나인 KLMS를 이용한 분류 및 회귀 성능 분석 실험을 통해서 KLMS는 신경망의 고전적인 학습문제의 하나인 지역값과 과잉 학습에 관한 잠정적 문제점을 갖고 있으며, 이는 Iris 데이터를 이용한 1 단계 실험과 가우시안 데이터를 이용한 2 단계 실험에서 순수 학습과 정규화 인수의 Inverse를 통한 결과를 비교해봄으로써 KLMS가 트레이닝 데이터에 과잉학습 되었음을 보여주었다.

또한 KLMSBL이 모멘텀을 적용한 후 분류 성능이 높아 졌을 뿐만 아니라 정규화 인수에 대한 오류율 또한 GCV 및 L-Curve에 비해 현저히 낮아지는 것을 실험을 통하여 알아보았다. 불량 조건하에서 보다 복잡한 문제인 2단계 실험은 동적모멘텀의 적용이 트레이닝 및 테스트 데이터에 대해 보다 일반화되고, 오류율이 낮은 정규화 인수를 구하는데 있어 성능향상에 도움이 됨을 실험을 통하여 알아보았다. 마지막 단계의 실험에서는 실제 이미지 복구 문제에 대해 동적모멘텀을 적용해 보는 실험으로 동적모멘텀을 적용한 KLMSBLDM이 KLMSBL에 비해 낮은 오류율을 보이며 기존의 GCV와 유사한 성능을 가지는 것을 알 수 있다.

#### 참고 문헌

- [1] S. Haykin, *Neural Networks, A comprehensive Foundation* Second Edition by Prentice-Hall, Inc, 1999.
- [2] 류재홍, 정종철, *커널 이완절차에 의한 커널 공간의 저 밀도 표현 학습*, 한국퍼지 및 지능시스템학회, 2001년도 추계 학술대회 학술 발표 논문집, 제 11권, 제 2호, pp. 60-64, 2001.12.
- [3] Eun-Mi Kim, Seong-Mi Park, Kwang-Hee Kim, Bae-Ho Lee, *An effective machine learning algorithm using momentum scheduling*, Hybrid Intelligent Systems, Japan, 2004. 12.
- [4] P. C. Hansen, *Regularization Tools, A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems*. Version 3.1 for Matlab 6.0, 2001.
- [5] 조용현, *모멘텀을 이용한 Support Vector Machines의 학습 성능 개선*, 한국정보처리학회 논문지, 제 7권 2p 5호, 2000.5.
- [6] J.C.Platt, *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*, In *Advances in Kernel Methods : Support Vector Learning*, ed., MIT Press, Cambridge, 1998.
- [7] O. L. Mangasarian and D. R. Musicant, *Active Set Support Vector Machine Classification*, Neural

Information Processing Systems 2000 (NIPS 2000),  
T. K. Lee, T. G. Dietterich and V. Tresp, editors,  
pp577-583. MIT Press, 2001.



김 은 미

2001년 여수대학교 영어영문학과졸업(컴퓨터공학 복수전공)(학사). 2003년 전남대학교 컴퓨터공학과(공학석사). 2004년~현재 전남대학교 컴퓨터정보통신공학과 박사과정 재학. 관심분야는 패턴인식, 인공지능, 영상처리, 신경망



박 성 미

1986년 전남대학교 계산통계학과(학사)  
2001년 전남대학교 컴퓨터공학과(공학석사). 2004년 전남대학교 컴퓨터공학과 박사수료. 2004년 3월~현재 정보통신연구진흥원 IT 교수요원, 전남대학교 초빙교수. 관심분야는 영상처리, 컴퓨터비전, 패턴인식

턴인식



김 광 회

1998년 전남대학교 컴퓨터공학과(학사)  
2000년 전남대학교 컴퓨터공학과(공학석사). 2000년 3월~현재 전남대학교 컴퓨터공학과 박사과정. 2002년 1월~현재 국방과학연구소 연구원. 관심분야는 영상신호처리(얼굴인식), 레이더 신호처리



이 배 호

1978년 한양대학교 공과대학 전자공학과(학사). 1980년 한국과학기술원(KAIST) 전기 및 전자공학과(공학석사). 1993년 University of Missouri, 전기 및 컴퓨터공학과(공학박사). 1980년~1983년 국방과학연구소(ADD) 연구원. 1993년~현재 전남대학교 공과대학 전자컴퓨터공학부 교수. 2004년~현재 NURI 전자정보가전인력양성사업단장. 관심분야는 인공지능, 컴퓨터 비전, 디지털영상처리, 멀티미디어, 의료영상