

모멘트와 바이어스 학습법에 의한 학습 성능

김은미*, 이배호*

*전남대학교 컴퓨터정보통신공학과
e-mail:em1342@hotmail.com

Learning performance of by the momentum and the bias learning method

Eun-Mi Kim*, Bae-Ho Lee*

*Dept. of Computer and information engineering, Chonnam
University

요 약

근원데이터나, 이원데이터를 이용한 문제를 해결하기 위해서는 많은 경우에 완전 해를 갖는 문제로 변형시키기 위해 정규화할 필요성이 있다. 본 논문에서는 이러한 정규화 인수를 찾는 문제를 기존의 GCV, L-Curve, 그리고 이원데이터를 RBF 신경회로망에 적용시킨 커널 학습법에 대한 각각의 성능을 비교실험을 통해 고찰한다. 이때 커널을 이용한 학습법의 성능을 향상하기 위해, 전체학습과 성능의 제한적 비례관계라는 설정아래, 각각의 학습에 따라 능동적으로 변화하는 동적모멘텀의 도입을 제안한다. 끝으로 제안된 동적모멘텀이 분류문제의 표준인 Iris 데이터, Singular 시스템의 대표적 모델인 가우시안 데이터, 그리고 마지막으로 1차원 이미지 복구문제인 Shaw데이터를 이용한 각각의 실험에서 분류문제와 회계문제 양쪽 모두에 있어 기존의 GCV, L-Curve와 동등하거나 우수한 성능이 있음을 보인다.

1. 서론

목적치를 아는 근원공간 혹은 이원공간에서의 데이터가 존재한다면 이는 단순한 QP문제로 재정의될 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. 그러나 이러한 방법들은 그 알고리즘과 성능의 우수함에도 불구하고 정규화 인수를 구하는 방법과 주어진 문제의 해를 구하는 방법이 이원화되었다는 결정적인 문제를 가지고 있다[1]. 이러한 이원화된 문제를 해결하기 위하여 패턴 가중치와 바이어스 항을 동시에 학습하는 커널 학습법을 이용한 새로운 정규화 인수를 찾는 방법인 바이어스 학습법으로 제안되었다. 하지만 제

안된 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.

본 논문에서는 이러한 바이어스 학습법이 가지는 문제점을 해결하기 위한 방법으로 동적모멘텀을 제안하고 이러한 동적모멘텀의 성능을 평가하기 위해 각기 다른 의미의 데이터셋을 이용 3단계에 의해 GCV, L-Curve와 함께 비교실험 하여 동적모멘텀의 적용이 분류·회계문제 양쪽 모두에 성능향상을 가져오며, GCV, L-Curve와 동등하거나 보다 나은 성능을 얻을 수 있음을 확인한다.

2. 바이어스 학습법

바이어스 항 α_0 와 정규화 인수 λ 의 관계를 정립하면 아래와 같다.

$$\begin{aligned} \mathbf{d} &= \mathbf{K} \mathbf{a} + \lambda \mathbf{a} \\ &= \mathbf{K} \mathbf{a} + \boldsymbol{\xi} \end{aligned} \quad (1)$$

$$\begin{aligned} &= \mathbf{K} \mathbf{a} + \overline{\boldsymbol{\xi}} + (\boldsymbol{\xi} - \overline{\boldsymbol{\xi}}) \\ &= \mathbf{K} \mathbf{a} + \alpha_0 + \boldsymbol{\varepsilon} \\ \boldsymbol{\xi} &\equiv \lambda \mathbf{a}, \\ \overline{\boldsymbol{\xi}} &\equiv \frac{\sum \xi_i}{N} = \frac{\lambda \sum \alpha_i}{N}, \end{aligned} \quad (2)$$

$$\begin{aligned} \alpha_0 &\equiv \overline{\boldsymbol{\xi}}, \\ \boldsymbol{\varepsilon} &\equiv (\boldsymbol{\xi} - \overline{\boldsymbol{\xi}}) \end{aligned}$$

수정된 정규화 네트워크의 해는 다음과 같다.

$$F_{\lambda}(\mathbf{y}) = \mathbf{a}_{reg}^t \mathbf{K}(:, \mathbf{y}) + \alpha_0 \quad (3)$$

여기에서 $\mathbf{a}_{reg} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{d}$ 이 된다. 실제 바이어스 학습(Bias Learning)에서는 \mathbf{a} 와 α_0 가 확장 커널을 이용하여 동시에 학습하게 된다.

패턴 행렬은 바이어스 α_0 를 학습하기 위해 아래와 같이 확장하였다.

$$\begin{aligned} \mathbf{H}^t \boldsymbol{\omega} &= \mathbf{d} \\ \mathbf{H} &= [\mathbf{K} \quad \mathbf{ones}(N, 1)]^t, \quad \boldsymbol{\omega} = \begin{bmatrix} \mathbf{a} \\ \alpha_0 \end{bmatrix} \end{aligned} \quad (4)$$

이상과 같은 방법으로 바이어스 α_0 는 패턴 가중치 \mathbf{a} 가 학습하는 동안 함께 학습함으로써 기존의 사용자 지정 방식의 정규화 인수는 패턴가중치와 함께 학습하게 된다. 이때 확장 커널 행렬 $\mathbf{H}^t \boldsymbol{\omega} = \mathbf{d}$ 는 입력공간 $\mathbf{Y}^t \mathbf{a} = \mathbf{d}$ 와 대응한다.

3. 동적모멘텀

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[4~7].

$$\begin{aligned} \mathbf{M} &= \frac{m(k+1)}{\tau} \\ \text{if } \mathbf{M} > m & \\ \mathbf{M} &= 0; \\ \tau &= \tau^2 \end{aligned} \quad (5)$$

여기서 τ 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 상한값이 된다. 결국 동적모멘트 \mathbf{M} 의 값이 모멘트의 상한값인 m 보다 클 경우 동적모멘트를 다시 0으로 하고 τ 를 τ^2 으로 다음번의 예상된 한계치를 취대로 뚫으로써 예상된 한계치 내에서 다시 모멘트를 증가하면서 학습하게 한다.

4. 실험 및 결과

실험을 위한 시스템의 구현은 Matlab 6.5을 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 이미 지복구문제인 SHAW 데이터를 이용한 3단계 실험으로 구성되며 본 논문에는 첫단계 실험만 게재하였다. Iris 데이터를 이용한 실험에서는 각각의 알고리즘의 이용하여 정규화 인수를 구하고 결정된 정규화 인수를 이용하여 바이어스와 오류율을 산출하였다. 이러한 방식으로 계산되어진 데이터를 순수한 학습에 따라 구한 데이터와 구분하기 위해 학습법의 뒤에 Inverse표기 하였다.

정규화 인수인 λ 의 크기는 KLMSBL>KLMSBLDM>GCV>L-Curve 순으로 크게 나왔으며, 각각의 알고리즘이 산출해내는 정규화 인수는 모두 다른 것으로 나타났다. 바이어스 항의 크기는 KLMSBL>KLMSBL Inverse>GCV>KLMSBLDM>L-Curve>KLMSBLDM 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. KLMSBL을 제외한 나머지 알고리즘들은 바이어스 항을 포함한 것이 바이어스 항을 포함하지 않은 것에 비해 전체적으로 정확도가 높아진 것을 알 수 있으며, 특히 KLMSBL은 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.

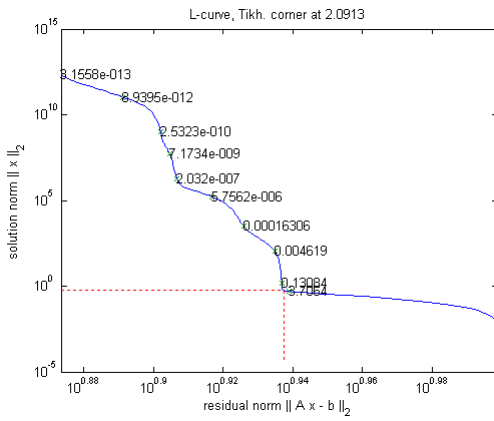


그림 1. L_curve

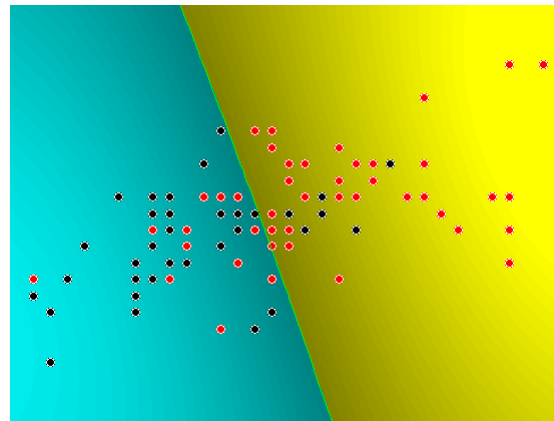


그림 3. KLMS_{DRBL}

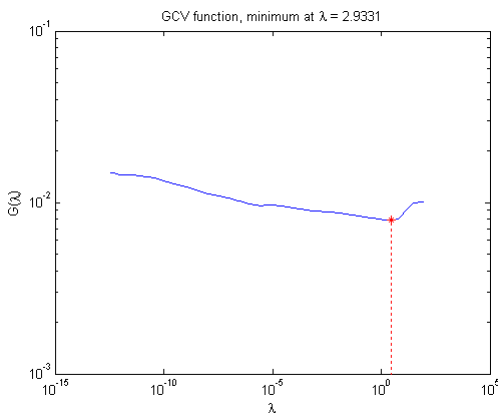


그림 2. GCV

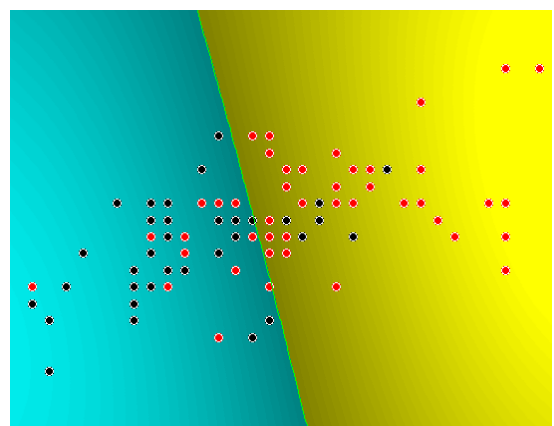


그림 4. KLMS_{DRBL} Momentum

표 1 Results of the regularization parameter using Iris data

	λ	bias	performance(%)	RMS Error	RMS Error with Bias
① L-Curve	2.091283e+000	2.077852e-003	74	8.706841e-001	8.706816e-001
② GCV	2.933131e+000	2.491655e-003	74	8.745416e-001	8.745381e-001
③ KLMS _{BL}	9.111957e+000	2.679960e-002	75	8.658558e-001	8.666892e-001
④ KLMS _{BL} Inverse	9.111957e+000	3.951047e-003	74	9.029183e-001	9.029097e-001
⑤ KLMS _{BLDM}	6.405503e-001	2.224852e-003	75	8.665070e-001	8.666279e-001
⑥ KLMS _{BLDM} Inverse	6.405503e-001	9.800551e-004	74	8.655781e-001	8.655776e-001

이것은 동적모멘텀을 적용한 KLMSBLDM과는 다른 관점으로 분류되어져야한다. KLMSBLDM은 같은 75%의 분류성능을 가지나 순수 학습과 Inverse를 이용한 방식에서 RMS오류 및 바이어스 항을 포함한 RMS오류의 실험결과 오류율에 대한 변화가 GCV 및 L-Curve와 유사한 성능을 보이며 두 두 방식에 오류율의 차이가 거의 없는 것으로 볼 때 과잉학습의 결과가 아니라는 것을 알 수 있다. 즉, 앞서 밝힌

바와 같이 KLMSBL은 신경망의 고전문제인 지역값과 과잉학습의 문제를 극복하지 못하는 것을 알 수 있으며 제안된 동적모멘텀이 이러한 문제를 효과적으로 해결할 수 있음을 실험을 통하여 알 수 있다.

5. 결론

기존의 L-Curve[2], GCV[3] 및 커널 학습법의 하나인 KLMS를 이용한 분류 및 회계 성능 분석 실험을

통해서 KLMS는 신경망의 고전적인 학습문제의 하나인 지역값과 과잉 학습에 관한 잠정적 문제점을 갖고 있으며, 이는 Iris데이터를 이용한 실험을 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.

참고문헌

- [1] 류재홍, 정종철 커널 이완절차에 의한 커널 공간의 저밀도 표현 학습, 한국퍼지 및 지능시스템학회, 2001년도 추계 학술대회 학술 발표 논문집, 제 11권, 제 2호, pp. 60-64, 2001.12
- [2] 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.
- [3] S. Haykin, Neural Networks, A comprehensive Foundation Second Edition by Prentice-Hall, Inc, 1999.
- [4] 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
- [5] 조용현, 모멘트를 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.