

## 준지도 커널능형회귀모형에 관한 연구<sup>†</sup>

석 경 하<sup>1</sup>

<sup>1</sup>인제대학교 데이터정보학과

접수 2013년 2월 21일, 수정 2013년 3월 10일, 게재확정 2013년 3월 18일

### 요약

데이터마이닝과 기계학습의 응용분야에서는 라벨 없는 자료를 이용하는 연구가 많이 진행되고 있다. 이러한 연구는 분류문제에 집중되었다가 최근에 회귀분석문제로 관심이 모아지고 있다. 본 연구에서는 커널능형회귀모형 형태의 준지도 회귀분석 방법을 제시한다. 제안된 방법은 기존의 전환적 방법과는 달리 라벨 없는 자료의 라벨을 추정하는 과정을 필요로 하지 않기 때문에 선택해야 할 모수의 수도 적고, 계산과정도 단순할 뿐 아니라 일반화에 강점이 있다. 모의실험과 실제 자료 분석을 통해 제안된 방법이 라벨 없는 자료를 잘 활용하여 라벨 있는 자료만 이용하는 방법보다 더 우수한 추정을 하는 것을 볼 수 있었다.

주요용어: 일반화 교차타당성, 준지도 커널능형회귀, 준지도 회귀분석, 최소제곱 서포터벡터회귀, 커널능형회귀.

### 1. 서론

회귀분석 (regression)과 분류 (classification)는 지도학습 (supervised learning)의 대표적인 방법이 고 군집분석 (clustering)은 자율학습 (unsupervised learning)의 대표적인 분석방법이다. 이렇게 분석 방법을 나누는 기준은 라벨 (label), 목표값 (target), 반응값 (response) 유무에 따른 것이다. 특히 분류 작업에서는 목표값을 라벨 (label)이라 부르지만 본 논문에서는 용어의 통일을 위해 회귀분석에서의 반응값도 라벨이라고 표기하겠다.

최근에 만들어지는 자료는 라벨을 만드는 것이 어렵거나, 많은 시간과 경비가 소요되거나 혹은 자료가 너무 커서 라벨을 만드는 것이 힘들기 때문에 라벨을 가지는 비율이 작은 경우가 많이 있다. 그 예로는 음성인식 (speech processing), 문자분류 (text categorization), 웹분류 (web categorization), 그리고 생명정보 (bioinformatics) 등이 있다.

이러한 자료를 분석하기 위해 준지도학습 (semi-supervised learning)이라는 새로운 영역이 개발되어 많은 연구가 이루어지고 있다 (Zhu, 2005; Chapelle 등, 2006; Xu 등, 2010; Singh 등, 2008). 준지도학습은 라벨 있는 자료와 라벨 없는 자료를 모두 사용하여 새로운 분류규칙을 만드는 방법이다. 라벨 있는 자료가 지도학습에 사용될 만큼 충분하지 않거나 라벨 없는 자료가 분류규칙을 만드는데 도움을 줄 수 있는 경우에는 준지도학습 방법을 사용한다.

최근에 준지도학습에 대해 많이 연구되고 있는데 가장 많이 사용되고 있는 알고리즘은 자기훈련 (self-training)이다. 이는 혼합모형 (mixture model)과 EM알고리즘의 한 부류로 해석되는데 이를 이용한

<sup>†</sup> 이 논문은 2012년도 정부 (교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2011-0009705).

<sup>1</sup> (621-749) 경남 김해시 어방동, 인제대학교 데이터정보학과, 통계정보연구소 교수.  
E-mail: statskh@inje.ac.kr

예로는 Rosenberg (2005)를 들 수 있다. 한번 잘못 정해진 규칙이 계속 영향을 미칠 수 있는 것이 자기 훈련의 최대 단점이다. 이를 극복하기 위해 개발된 것이 상호훈련 (co-training)이다. 이 방법은 변수들을 독립된 두 개의 집합으로 나눌 수 있고, 각 변수 집합에서 분류기 (classifier)가 충분히 훈련될 수 있다는 가정을 하고 있다. 여기에 관한 것은 Nigam과 Ghani (2000)와 Zhou와 Goldman (2004)에서 볼 수 있다.

최근에 많이 주목 받고 있는 방법으로는 Vapnik (1998)의 TSVM (transductive support vector machine)을 들 수 있다. 이 방법은 SVM의 경계최대화 (margin maximization) 개념을 이용한 것으로서 밀도가 희박한 지역 (sparse region)으로 분류선이 지나도록 설계되었다 (low density separation). 이 원리를 이용한 방법으로는 정보정칙화 (information regularization; Szummer와 Jaakkola, 2002)가 있다.

그러나 위에서 소개한 여러 가지 방법들은 해가 전역최적 (global optimum)이 아닌 국소최적 (local optimum)에 수렴할 수 있다는 결정적인 단점을 가지고 있다. 그 이유는 목적함수가 볼록함수 (convex function)가 아니기 때문이다. 또한 TSVM 처럼 많은 계산 시간을 요구하는 단점을 가진다. 이러한 단점을 극복하기 위해 그래프기반 준지도학습이 개발되었다.

그래프기반 준지도학습에는 여러 가지가 개발되어 있는데 그 중에서 가장 기본이 되는 연구는 Zhou 등 (2003)의 GRFM (Gaussian random fields model)과 Zhou 등 (2004)의 CM (consistency model)이다. 이 두 가지 방법은 그래프이론을 기반으로 자료의 다양체구조 (manifold structure)를 탐색하기 때문에 전역최적에 수렴하는 좋은 성질을 가지고 있다 (Belkin 등, 2006). 이와 더불어 준지도학습에 관한 이론을 정립하여 많은 연구의 동기를 제공한 Niyogi (2008)가 있고, 준지도학습이 지도학습보다 더 좋은 결과를 보일 수 있음을 보인 Chapelle 등 (2006), Zhu (2005) 그리고 Zhu와 Goldberg (2009) 등이 대표적인 연구로 평가된다.

준지도에 관한 대부분의 연구는 분류문제에 국한되어 있는 것이 현실이다. 그러나 Belkin 등 (2006)과 Zhu (2005)는 그래프기반 준지도 방법이 회귀분석에 응용될 수 있음을 보여 줌으로써 라벨 없는 자료가 회귀분석에서도 유익한 정보를 줄 수 있다는 것을 보였다. 그리고 Zhou와 Li (2007)는 상호훈련과 서로 다른 측도 (metric)를 기반으로 하는 상호회귀 (CoReg) 방법을 제시하여 좋은 결과를 보여 주었다. 그리고 라벨 없는 자료 중 신뢰 (confidence)가 높은 자료만 상호적으로 이용하도록 제안하였다. 하지만 사용된 회귀모형이 계산은 용이하지만 수행능력이 떨어지는 최근인접 (k-nearest neighbor) 모형이라는 것이 아쉽다는 평가다. Wang 등 (2006)은 전통적인 커널회귀모형 (kernel regression)을 기반으로 준지도회귀모형을 개발하였다. 그들이 개발한 모형은 가중인자 (weight factor)를 통해 라벨 없는 자료의 영향을 조절하였다. 가중인자 선택을 적절하게 할 수 있으면 제안된 준지도회귀분석이 커널회귀분석과 그래프기반 회귀분석보다 더 나은 결과를 가져 올 수 있음을 보였다. Cortes와 Mohri (2007)도 커널회귀모형을 기반으로 준지도회귀분석 모형을 개발하였는데 Vapnik (1998)의 오차한계 (error bound)를 사용하였다. 제안된 방법은 오직 라벨 없는 자료의 라벨 추정에만 관심 있는 전환적 (transductive) 방법인데 실험을 통해 큰 자료에서도 수행결과가 좋다는 것을 보여주었다. Xu 등 (2011)은 최소제곱 서포트벡터 회귀모형 (LS-SVR, least square support vector regression; Suykens 등, 2002)을 기반으로 하는 준지도 최소제곱 서포트벡터 회귀모형 (S2LS-SVR; semi-supervised LS-SVR)을 개발하여 실험을 통해 제안된 방법의 효율성을 입증하였다. Lafferty와 Wasserman (2008)은 그래프 라플라시안 (Laplacian)을 사용하는 정칙성 (regularization)에 기반을 둔 준지도 회귀분석이 커널회귀분석보다 더 빠른 최소최대 (minimax) 수렴속도를 가질 수 없음을 보였다. 그러나 준지도 평활성 (semi-supervised smoothness) 가정을 보완시킴으로써 더 나은 준지도 방법을 만들 수 있음을 보여주었다. 최근에 Seok (2012)은 국소 상수 (local constant) 회귀모형을 기반으로 하는 준지도 회귀모형을 개발하였다. 그리고 개발된 회귀모형의 점근 분포를 밝혀 커널회귀모형보다 더 빠른 수렴속도를 가

질 수 있는 조건을 찾았다. 그리고 모의실험을 통해 제안된 이론의 타당성을 증명하였다.

준지도 회귀분석에 관한 최근의 연구, Wang 등 (2006), Cortes와 Mohri (2007), Xu 등 (2011), Seok (2012)에서 개발된 방법은 크게 두 단계로 나누어진다. 그 첫 번째 단계는 라벨 없는 자료의 라벨을 추정하는 초기 추정 (pilot estimation) 단계이고 이렇게 추정된 자료와 라벨 있는 자료를 통합하여 최종 추정을 하는 것이 두 번째 단계이다. 이런 방법의 단점은 초기 추정값을 목표값으로 하는 전환적 방법이라는 것이다. 추정값을 목표값으로 사용한다는 것은 그 추정값을 관측값으로 대신한다는 것인데 오차에 대한 충분한 정보가 없는 상태에서는 최종 추정의 정확성을 보장할 수 없을 것이다. 그리고 최종 추정에 영향력이 큰 초기 추정치 때문에 라벨 있는 자료가 가지는 정보를 축소시킬 수 있다. 그 뿐만 아니라 선택해야 할 모수가 많아지고 계산과정이 복잡해져 계산 시간이 많이 필요할 수 있다. 이와 같은 단점을 극복하기 위해 본 연구에서는 초기 추정을 필요로 하지 않는 준지도 회귀모형을 제안한다. 제안된 방법은 기존 방법에서 요구하는 라벨 없는 자료의 라벨을 추정하는 초기 추정 과정을 필요로 하지 않기 때문에 선택해야 할 모수의 수도 적고 계산과정도 단순하다. 그리고 전환적 방법보다 일반화 (generalization)의 장점을 가진다. 모의실험과 실제 자료 분석을 통해 제안된 방법이 라벨 없는 자료를 잘 활용하여 라벨 있는 자료만 이용하는 방법보다 더 우수한 추정을 하는 것을 볼 수 있었다.

2절에서는 커널능형회귀모형 (KRR; kernel ridge regression)을 소개한다. 그리고 3절에서는 LS-SVR과 S2LS-SVR을 소개하고 4절에서는 준지도 커널능형회귀모형 (S2KRR; semi-supervised KRR)을 소개한다. 5절에서는 제안된 방법이 잘 적용됨을 실험을 통해 보일 것이다. 그리고 6절에서는 결론 및 향후연구과제에 대해 언급한다.

## 2. 커널능형회귀모형

자료가  $\{(X_L, \mathbf{y}_L)\} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$ 로 주어지고,  $\mathbf{x}_i \in R^d$ 는 입력값이고  $y_i$ 는  $\mathbf{x}_i$ 의 라벨인데 실수 값을 갖는다. 회귀분석은  $y_i$ 를 잘 설명해 주는 함수  $f$ 를 찾는 것이 목적이다. 융통성 있는 함수 집합을 정의하여 그 집합 안에서 과적합을 예방하고 예측 잘하는 함수를 찾고자 한다. 함수  $f$ 가 속하는 함수집합을 설명하기 전에 다음과 같은 특징공간으로의 사상 (mapping)  $\psi: R^d \rightarrow R^M$ 을 고려한다. 이 사상을 이용한 함수집합을 다음과 같이 정의할 수 있다.

$$H = \{f | f(\mathbf{x}) = \boldsymbol{\omega}'\psi(\mathbf{x}), \boldsymbol{\omega} \in R^M\}.$$

일반적으로 특징공간의 차원 ( $M$ )은 원변수의 차원 ( $d$ )보다 아주 크기 때문에 특징공간에서 계산하는 것은 복잡하다. 그러므로 특징공간에서 계산하는 것을 피하는 방법을 모색한다. 만약  $H$ 가 함수 추정에 충분한 집합이라면 다음과 같은 정칙화 (regularized) 오차함수를

$$\min_{f \in H} \sum_{i=1}^l (y_i - f(\mathbf{x}_i))^2 + \lambda_1 \|f\|_H^2 \quad (2.1)$$

고려할 수 있다. 여기에서  $\lambda_1$ 는 함수의 복잡도와 자료의 적합도를 조절하는 정칙 모수 (regularization parameter)이다. Kimeldorf와 Wahba (1971)의 정리 (representer theorem)에 의해  $f(\mathbf{x}) = \sum_{i=1}^l \alpha_i k(\mathbf{x}, \mathbf{x}_i)$ 로 표현이 가능하므로 함수집합은

$$H = \left\{ f | f: R^d \rightarrow R, f(\mathbf{x}) = \sum_{i=1}^l \alpha_i k(\mathbf{x}, \mathbf{x}_i) = K_{LL} \boldsymbol{\alpha}_L \right\}$$

로 정의 할 수 있다. 여기에서  $K_{LL}$  는  $l \times l$  크기의 커널행렬인데  $\mathbf{x}_i$ 와  $\mathbf{x}_j$ 의 커널함수  $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \psi(\mathbf{x}_i)' \psi(\mathbf{x}_j)$ 가  $(i, j)$ 번째 원소이고,  $\mathbf{w} = \sum_{k=1}^l \alpha_k \psi(\mathbf{x}_k)$ 와 같이 표현됨을 커널재생 (reproducing kernel) 이론에 의해 증명된다. 사용될 수 있는 여러 종류의 커널함수 중에서 대표적인 것은 다음과 같다.

$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{y}$  : 선형 (linear) 커널

$k(\mathbf{x}, \mathbf{y}) = (\tau + \mathbf{x}'\mathbf{y})^d$  : 다항 (polynomial) 커널

$k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2 / \sigma^2)$  : 가우시안 (Gaussian, radialbasis function) 커널

$k(\mathbf{x}, \mathbf{y}) = \tanh(\kappa_1 \mathbf{x}'\mathbf{y} + \kappa_2)$  : MLP (multi-layer perceptron) 커널

통상적으로 가우시안 커널을 많이 사용하는데 그 이유는 다음과 같다 (Xu 등, 2011; Lin과 Lin, 2003). 1) 선형 커널은 수행 성능이 떨어지고, 가우시안 커널의 특별한 경우이다. 2) 다항 커널은 모수의 수가 MLP 커널과 같이 2개로 많고 수치적 해를 구하는데 어려움을 가지고 있다. 3) MLP 커널은 양정치가 아닐 뿐 아니라 가우시안과 유사한 성능을 가진다. 4) 가우시안 커널은 많은 사용자가 선호하는 평활성 (smoothness)을 가진다. 그러므로 본 연구에서도 가우시안 커널을 사용한다.

목적함수 (2.1)은 행렬로 표기하면

$$\min_{\boldsymbol{\alpha}_L} (\mathbf{y}_L - K_{LL} \boldsymbol{\alpha}_L)' (\mathbf{y}_L - K_{LL} \boldsymbol{\alpha}_L) + \lambda_1 \boldsymbol{\alpha}_L' \boldsymbol{\alpha}_L \quad (2.2)$$

로 쓸 수 있는데 이것을 만족하는  $\boldsymbol{\alpha}_L$ 는

$$\hat{\boldsymbol{\alpha}}_L = (K_{LL}' K_{LL} + \lambda_1 I_l)^{-1} K_{LL}' \mathbf{y}_L$$

이다. 여기에서  $I_l$ 은 크기가  $l \times l$ 인 단위행렬이다. 그리고 새로운 입력  $\mathbf{x}_*$ 에 대한 KRR 추정값  $\hat{y}^*$ 는 다음과 같이 구해진다.

$$\hat{y}^* = \hat{f}(\mathbf{x}_*) = \sum_{i=1}^l \hat{\alpha}_i k(\mathbf{x}_*, \mathbf{x}_i) = K_{L*}' (K_{LL}' K_{LL} + \lambda_1 I_l)^{-1} K_{LL}' \mathbf{y}_L$$

여기에서  $K_{L*}$ 는 크기가  $l$ 인 행벡터인데  $i$ 번째 원소가  $k(\mathbf{x}_*, \mathbf{x}_i)$ ,  $i = 1, \dots, l$ 이다. 최근에도 KRR 추정법은 모델복잡도와 예측정확도를 잘 해결해 주는 것으로 평가되고 있다 (Hofmann 등, 2008; Exterkate 등, 2011).

### 3. 준지도 최소제곱 서포트벡터 회귀모형

#### 3.1. 최소제곱 서포트벡터 회귀모형

2절과 같이  $\psi : R^d \rightarrow R^M$ 가 입력공간을 고차원 특징공간으로 사상하는 특징함수라고 할 때, 다음과 같이  $f(\mathbf{x}) = \mathbf{w}'\psi(\mathbf{x})$ 는 원공간 (primal space)에서 다음의 최적화 문제의 해로 정의될 수 있다.

$$\min_{\mathbf{w}, e} J_P(\mathbf{w}, e) = \frac{1}{2} \mathbf{w}'\mathbf{w} + \frac{\lambda}{2} \sum_{i=1}^l e_i^2 \quad (3.1)$$

$$\text{s.t. } y_i = \mathbf{w}'\psi(\mathbf{x}_i) + b + e_i, i = 1, \dots, l.$$

여기에서  $\lambda$ 는 정칙 모수이다. (3.1)을 라그랑제 (Lagrange) 배수  $\alpha_i$ ,  $i = 1, \dots, l$ 를 사용한 라그랑제 함수로 표현하면 다음과 같다.

$$L_D(\boldsymbol{\alpha}) = J_P - \sum_{i=1}^l \alpha_i \{ \mathbf{w}'\psi(\mathbf{x}_i) + b - y_i + e_i \}.$$

최적 라그랑제 배수는 다음의 조건으로부터 구해진다.

$$\begin{aligned}\frac{\partial L_D}{\partial \mathbf{w}} = 0 &\Rightarrow \mathbf{w} = \sum_{i=1}^l \alpha_i \psi(\mathbf{x}_i), \\ \frac{\partial L_D}{\partial b} = 0 &\Rightarrow \sum_{i=1}^l \alpha_i = 0, \\ \frac{\partial L_D}{\partial e_i} = 0 &\Rightarrow \alpha_i = \lambda e_i, i = 1, \dots, l, \\ \frac{\partial L_D}{\partial \alpha_i} = 0 &\Rightarrow y_i - \mathbf{w}'\psi(\mathbf{x}_i) - b - e_i = 0, i = 1, \dots, l.\end{aligned}\quad (3.2)$$

(3.2)의 변수  $\mathbf{w}$ 와  $e$ 를 제거하여 다음과 같은 식으로부터  $\boldsymbol{\alpha}$ 와  $b$ 를 구할 수 있다.

$$\begin{pmatrix} 0 & \mathbf{1}_l' \\ \mathbf{1}_l & K + I_l/\lambda \end{pmatrix} \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{y}_L \end{pmatrix}$$

여기에서  $\mathbf{1}_a$ 는 모든 원소가 1인  $a \times 1$  크기의 벡터이다. 따라서 (3.2)와 Mercer 정리 (1909)를 이용하면 입력값이  $\mathbf{x}_*$ 인 자료의 LS-SVR 추정값은

$$\hat{f}(\mathbf{x}_*) = \sum_{i=1}^l \hat{\alpha}_i K(\mathbf{x}_*, \mathbf{x}_i) + \hat{b}$$

이다.

### 3.2. 준지도 최소제곱 서포트벡터 회귀모형

$\{(X_L, \mathbf{y}_L)\} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$ 와 더불어 라벨 없는 자료  $X_U = (\mathbf{x}_{l+1}, \dots, \mathbf{x}_n)'$ ,  $n = l + u$ 가 주어질 때 S2LS-SVR은 다음과 같이 정의된다.

$$\begin{aligned}\min_{\mathbf{w}, b, \xi, \zeta} J(\mathbf{w}, \xi, \zeta) &= \frac{1}{2} \mathbf{w}'\mathbf{w} + \frac{\lambda_1}{2} \xi'\xi + \frac{\lambda_2}{2} \zeta'\zeta \\ \text{s.t. } y_i &= \mathbf{w}'\psi(\mathbf{x}_i) + b + \xi_i, i = 1, \dots, l, \\ \hat{y}_i &= \mathbf{w}'\psi(\mathbf{x}_i) + b + \zeta_i, i = l + 1, \dots, n.\end{aligned}\quad (3.3)$$

여기에서  $\xi = (\xi_1, \dots, \xi_l)'$ ,  $\zeta = (\zeta_{l+1}, \dots, \zeta_n)'$ 는 실수값을 가지는 보조변수 (slack variable)이고  $\hat{y}_i$  ( $i = l + 1, \dots, n$ )는  $x_i$ 에 대응되는 라벨의 초기 추정값이다. 그리고  $\lambda_1, \lambda_2$ 는 정칙 모수이다. 식 (3.3)을 라그랑제 함수로 표현하면 아래와 같다.

$$\min L_{\mathbf{w}, b, \xi, \zeta} = J - \sum_{k=1}^l \alpha_k \{ \mathbf{w}'\psi(\mathbf{x}_k) + b + \xi_k - y_k \} - \sum_{k=l+1}^n \beta_k \{ \mathbf{w}'\psi(\mathbf{x}_k) + b + \zeta_k - \hat{y}_k \} \quad (3.4)$$

(3.4)로부터 다음의 조건식을 얻을 수 있다.

$$\begin{aligned}
 \frac{\partial L}{\partial \mathbf{w}} = 0 &\Rightarrow \mathbf{w} = \sum_{k=1}^l \alpha_k \psi(\mathbf{x}_k) - \sum_{k=l+1}^n \beta_k \psi(\mathbf{x}_k), \\
 \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{k=1}^l \alpha_k + \sum_{k=l+1}^n \beta_k = 0, \\
 \frac{\partial L}{\partial \xi_k} = 0 &\Rightarrow \lambda_1 \xi_k = \alpha_k, k = 1, \dots, l, \\
 \frac{\partial L}{\partial \zeta_k} = 0 &\Rightarrow \lambda_2 \zeta_k = \beta_k, k = l+1, \dots, n, \\
 \frac{\partial L_D}{\partial \alpha_k} = 0 &\Rightarrow y_k - \mathbf{w}'\psi(\mathbf{x}_k) - b - \xi_k = 0, k = 1, \dots, l, \\
 \frac{\partial L_D}{\partial \beta_k} = 0 &\Rightarrow \hat{y}_k - \mathbf{w}'\psi(\mathbf{x}_k) - b - \zeta_k = 0, k = l+1, \dots, n.
 \end{aligned} \tag{3.5}$$

LS-SVR과 같은 방법으로 (3.5)에서  $\mathbf{w}, \xi, \zeta$ 를 제거하여 다음의 식으로부터  $b, \alpha, \beta$ 를 구할 수 있다.

$$\begin{pmatrix} 0 & \mathbf{1}'_l & \mathbf{1}'_u \\ \mathbf{1}_l & K_{LL} + I_l/\lambda_1 & K_{LU} \\ \mathbf{1}_u & K_{UL} & L_{UU} + I_u/\lambda_2 \end{pmatrix} \begin{pmatrix} b \\ \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{y}_L \\ \mathbf{y}_U \end{pmatrix}$$

여기에서  $K_{LU} = (k_{ij})_{i=1, \dots, l, j=l+1, \dots, n}$ ,  $K_{UL} = (k_{ij})_{i=l+1, \dots, n, j=1, \dots, l}$ ,  $K_{UU} = (k_{ij})_{i, j=l+1, \dots, n}$  그리고  $\mathbf{y}_U = (\hat{y}_{l+1}, \dots, \hat{y}_n)'$ 이다. 새로운 입력  $\mathbf{x}_*$ 에 대한 S2LS-SVR 추정값  $\hat{y}^*$ 는 다음과 같이 구해진다.

$$\begin{aligned}
 \hat{f}(\mathbf{x}_*) &= \mathbf{w}'\psi(\mathbf{x}_*) + \hat{b} \\
 &= \sum_{i=1}^l \hat{\alpha}_i K(\mathbf{x}_*, \mathbf{x}_i) + \sum_{i=l+1}^n \hat{\beta}_i K(\mathbf{x}_*, \mathbf{x}_i) + \hat{b}
 \end{aligned}$$

#### 4. 준지도 커널능형회귀모형

커널회귀모형을 이용하는 준지도회귀분석 모형은

$$\begin{aligned}
 \hat{\mathbf{f}} &= K\alpha = \begin{pmatrix} K_{LL}K_{LU} \\ K_{UL}K_{UU} \end{pmatrix} \begin{pmatrix} \alpha_L \\ \alpha_U \end{pmatrix} \\
 &= \begin{pmatrix} K_{LL}\alpha_L + K_{LU}\alpha_U \\ K_{UL}\alpha_L + K_{UU}\alpha_U \end{pmatrix}
 \end{aligned}$$

와 같은 형태를 지닌다. 여기에서,  $\hat{\mathbf{f}} = (\hat{f}(\mathbf{x}_1), \dots, \hat{f}(\mathbf{x}_n))'$ ,  $K = [K_{LL}K_{LU}; K_{UL}K_{UU}]$ 이다. 그리고  $\alpha = [\alpha_L; \alpha_U]$ ,  $\alpha_L = (\alpha_1, \dots, \alpha_l)'$ ,  $\alpha_U = (\alpha_{l+1}, \dots, \alpha_n)'$ 이다 (편의상  $[\mathbf{x}; \mathbf{y}] = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ 로 표기 함). 별칭 항을 포함하는 목적함수  $L_F$ 은

$$\begin{aligned}
 L_F &= (\mathbf{y}_L - K_{LL}\alpha_L - K_{LU}\alpha_U)'(\mathbf{y}_L - K_{LL}\alpha_L - K_{LU}\alpha_U) \\
 &\quad + (\mathbf{y}_U - K_{UL}\alpha_L - K_{UU}\alpha_U)'(\mathbf{y}_U - K_{UL}\alpha_L - K_{UU}\alpha_U) + \lambda_2 \alpha'_L \alpha_L + \lambda_3 \alpha'_U \alpha_U \\
 &= (\mathbf{y} - K\alpha)'(\mathbf{y} - K\alpha) + \alpha' \Lambda \alpha
 \end{aligned} \tag{4.1}$$

이다. 여기에서  $\mathbf{y} = [\mathbf{y}_L; \mathbf{y}_U]$  그리고  $\Lambda = [\lambda_2 I_L 0_{l \times u}; 0_{u \times l} \lambda_3 I_u]$ 이다. (4.1)을 최소화하는  $\alpha$ 는 다음과 같이 구할 수 있다.

$$\hat{\alpha} = (K'K + \Lambda)^{-1} K' \mathbf{y}$$

그리고 새로운 입력  $\mathbf{x}_*$ 에 대한 추정값  $\hat{f}_F(\mathbf{x}_*)$ 은

$$\hat{f}_F(\mathbf{x}_*) = \sum_{i=1}^n \hat{\alpha}_i k(\mathbf{x}_*, \mathbf{x}_i) = K_*(K'K + \Lambda)^{-1} K' \mathbf{y} \quad (4.2)$$

이다. 여기에서  $K_*$ 는 크기가  $n$ 인 행벡터인데  $i$ 번째 원소가  $k(\mathbf{x}_*, \mathbf{x}_i)$ ,  $i = 1, \dots, n$ 이다. (4.2)의  $\hat{f}_F$ 을 얻기 위해서 다음과 같은 단계를 거쳐야한다.

1.  $\hat{\alpha}_L = (K'_{LL} K_{LL} + \lambda_1 I_l)^{-1} K'_{LL} \mathbf{y}_L$ , 모수  $\lambda_1$ 과  $\sigma_1$  선택
2.  $\mathbf{y}_U = K_{UL} \hat{\alpha}_L$
3.  $\hat{\alpha} = (K'K + \Lambda)^{-1} K' \mathbf{y}$ ,  $\mathbf{y} = [\mathbf{y}_L; \mathbf{y}_U]$ , 모수  $\lambda_2, \lambda_3, \sigma_2, \sigma_3, \sigma_4, \sigma_5$  선택
4.  $\hat{f}_F(\mathbf{x}_*) = K_* \hat{\alpha}$

거의 모든 준지도 회귀분석에 관한 선행연구 (Wang 등, 2006; Cortes와 Mohri, 2007; Xu 등, 2011; Seok, 2012)는 이와 유사한 단계를 거치는 추정방법을 제시했다. 이 방법의 단점은 1) 1, 2 단계에서 구한 초기 추정값  $\mathbf{y}_U = K_{UL} \hat{\alpha}_L$ 을 3 단계에서는 목표값으로 사용해야 한다. 추정값을 목표값으로 사용한다는 것은 그 추정값을 관측값으로 대신한다는 것인데 오차에 대한 충분한 정보가 없는 상태에서는 최종 추정의 정확성을 보장할 수 없다. 2)  $X_L$ 이 가지는 정보 중 부족한 부분을 보충하고자  $X_U$ 를 사용한다. 그런데  $X_U$ 와 더불어 최종 추정에 영향력이 큰  $\mathbf{y}_U$ 를 사용함으로써  $\mathbf{y}_L$ 이 가지는 정보를 왜곡시킬 수 있다. 3) 초기 추정의 정확성이 더 좋은 최종 추정을 보장하지 못한다. 4) 선택해야할 모수가 많아 계산과정이 복잡하고 많은 계산시간을 요구한다. 5) 최종 추정이 라벨 있는 자료만으로 추정하는 결과와 아주 유사하게 나올 수 있다. 6) 라벨 없는 자료의 추정에만 관심을 가지는 전환적방법이다. 이와 같은 단점을 극복하기 위해 다음과 같은 형태의 준지도 회귀모형을 제안한다.

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^l \alpha_i K(\mathbf{x}_i, \mathbf{x}) + \sum_{j=l+1}^n \alpha_j K(\mathbf{x}_j, \mathbf{x}) \quad (4.3)$$

식 (4.3)은 행렬을 이용해서 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \hat{\mathbf{f}}_L &= K_1 \alpha \\ &= K_{LL} \alpha_L + K_{LU} \alpha_U \end{aligned}$$

여기에서,  $\hat{\mathbf{f}}_L = (\hat{f}(\mathbf{x}_1), \dots, \hat{f}(\mathbf{x}_l))'$ ,  $K_1 = [K_{LL} K_{LU}]$ 이다. 목적함수는 다음과 같이 쓸 수 있다.

$$L = (\mathbf{y}_L - K_1 \alpha)' (\mathbf{y}_L - K_1 \alpha) + \alpha' \Lambda \alpha \quad (4.4)$$

여기에서  $\Lambda = [\lambda_1 I_l \times l \ 0_{l \times u}; 0_{u \times l} \ \lambda_2 I_u \times u]$ , 그리고  $\lambda_1, \lambda_2$ 는 정칙 모수이다. 목적함수 (4.4)를 최소화하는  $\hat{\alpha}$ 는

$$\hat{\alpha} = (K'_1 K_1 + \Lambda)^{-1} K'_1 \mathbf{y}_L$$

로 주어진다. 그리고 새로운 입력  $\mathbf{x}_*$ 에 대한 S2KRR 추정값  $\widehat{f}_L(\mathbf{x}_*)$ 은

$$\widehat{f}_L(\mathbf{x}_*) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_*, \mathbf{x}_i) = K_*(K_1'K_1 + \Lambda)^{-1}K_1'\mathbf{y}_L$$

인데, 여기에서  $K_*$ 는 크기가  $n$ 인 행벡터인데  $i$ 번째 원소가  $k(\mathbf{x}_*, \mathbf{x}_i)$ 이다. 추정값  $\widehat{f}_L(\mathbf{x})$ 은  $\widehat{f}_F(\mathbf{x})$ 을 구할 때 필요한 단계 1과 2를 거치지 않고 얻을 수 있기 때문에 추정에 필요한 모수도 최대 4개 ( $K_{LL}, K_{LU}, \lambda_1, \lambda_2$ )로 최대 8개가 필요한  $\widehat{f}_F(\mathbf{x})$ 보다 계산이 쉬울 뿐 아니라 계산에 필요한 시간과 공간을 줄일 수 있다. 그리고 라벨 없는 자료의 라벨 추정값을 사용하지 않기 때문에 일반화가 더 잘 될 것으로 기대된다.

## 5. 실험

제안된 방법의 타당성을 알아보기 위해 모의실험을 실행하였다. 실험에 사용된 자료는 모의 생성된 자료와 실제자료이다. 추정에 필요한 모수 선택을 위해서 여러 가지 방법을 고려할 수 있겠지만 그 중에서 이해가 빠르고 사용하기 쉬운 일반화 교차타당성 (GCV; generalized cross validation)을 사용하였는데 다음과 같은 식으로 나타낸다.

$$GCV = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \widehat{f}(\mathbf{x}_i)}{1 - \text{trace}(A)/n} \right)^2.$$

여기에서  $\widehat{\mathbf{f}}(\mathbf{x}) = (\widehat{f}(\mathbf{x}_1), \dots, \widehat{f}(\mathbf{x}_n))' = A\mathbf{y}$ 이다. 그리고 여러 방법의 예측력을 평가하기 위해 200번의 반복에 의한 평균제곱오차 (MSE; mean squared error)와 상관계수의 평균과 표준편차를 계산하였다.

### 5.1. 실험에 사용된 자료

#### 모의생성 자료

실험에 사용된 모의생성 자료는 sine 자료와 sinc 자료다. 사용된 sine 자료는  $y_i = \sin(2\pi x_i) + e_i$ 인데  $x_i \sim iid U(0, 1)$ ,  $e_i \sim iid N(0, 0.2^2)$ 분포를 따른다. 그리고 sinc 자료는  $y_i = \text{sinc}(x_i) + e_i$ 인데  $x_i \sim iid U(-3, 3)$ ,  $e_i \sim iid N(0, 0.1^2)$ 분포를 따른다. 훈련용 (training)으로는 크기가 50 (100), 시험용 (test)으로는 크기 100, 그리고 라벨 없는 자료는 크기가 300 (1,000)인 자료를 사용하였다.

#### Corn 자료

이 자료 ([http://www.eigenvector.com/Data/Data sets.html](http://www.eigenvector.com/Data/Data%20sets.html))는 3개의 분광계 (m5, mp5, mp6)로 측정된 80개의 옥수수에 관한 자료다. 파장의 범위 (독립변수)는 1100~2498nm인데 2nm 간격으로 구성되었다. 그리고 종속변수는 수분 (moisture), 기름 (oil), 단백질 (protein) 그리고 탄수화물 (starch) 등 4개다. Rosipal과 Trejo (2001)와 Xu 등 (2011)의 실험처럼, 자료에 있는 종속변수를 사용하지 않고 4개의 종속변수를 다음과 같이 생성하여 사용하였다.

$$\begin{aligned} y_1 &= \exp(\mathbf{x}'\mathbf{x}/2m) \\ y_2 &= \exp(\mathbf{x}'A^{-1}\mathbf{x}/2m_1) \\ y_3 &= (\mathbf{x}'\mathbf{x}/m)^3 \exp(\mathbf{x}'\mathbf{x}/2m) \\ y_4 &= 0.3y_1 + 0.25y_2 - 0.7y_3 \end{aligned}$$

여기에서  $A$ 는 대각성분이 1이고 그 외의 값이 0.8인 대칭행렬이고,  $m$ 과  $m_1$ 은 각각  $\{\mathbf{x}_i'\mathbf{x}_i\}_{i=1}^{80}$ 과  $\{\mathbf{x}_i'A^{-1}\mathbf{x}_i\}_{i=1}^{80}$ 의 평균이다. 위와 같이 생성된 자료에 평균이 0이고 표준편차는 종속변수의 표준편차와



의 비 (signal to noise)가 15%인 정규분포를 가지는 오차를 추가하였다. 총 80개의 관측값 중 20개는 훈련용으로, 20개는 시험용으로 그리고 나머지 40개는 라벨 없는 자료로 활용하였다.

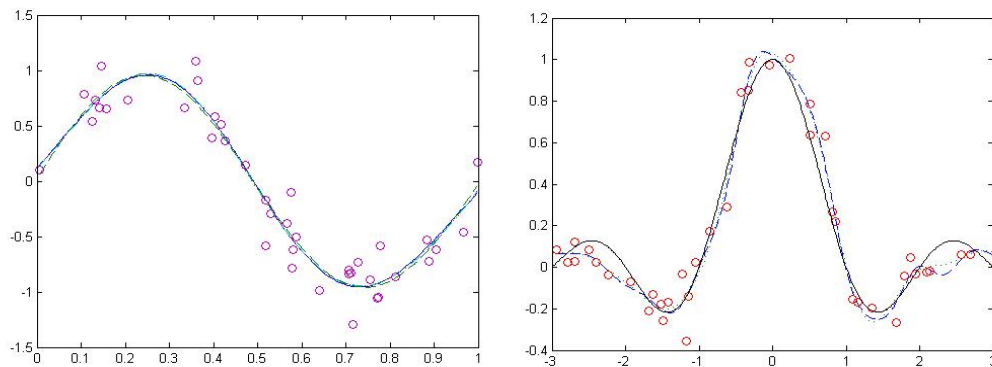
### Polymer 자료

고분자시험시설에서 얻어진 자료 (Ungar, 1995)로써 61개의 관측값, 10개의 독립변수 그리고 4개의 종속변수로 구성된 자료이다. 10개의 독립변수가 단지 온도, 공급량 등이라는 것 이외의 다른 정보는 없다. 이 자료는 불규칙적으로 분포된 자료에 대한 비선형모형의 강건성 (robustness)을 시험하기에 아주 좋은 자료로 알려져 있다. 총 61개의 관측값 중 21개는 훈련용으로, 10개는 시험용으로 그리고 나머지 30개는 라벨 없는 자료로 활용하였다.

## 5.2. 실험결과

모의 생성자료에 대한 결과가 Table 5.1과 Figure 5.1에 나타나 있다. Figure 5.1에서 볼 수 있듯이 KRR과 S2KRR은 sine 함수의 추정에는 좋은 결과를 보여준다. 그러나 sinc 함수의 추정에서는 S2KRR이 더 좋은 추정을 하고 있음을 알 수 있다. 이러한 결과는 Table 5.1에서 확인 할 수 있는데, 거의 모든 경우에서 S2KRR이 KRR보다 더 작은 평균제곱오차와 더 큰 상관계수를 가진다. 그 뿐만 아니라 S2KRR이 LS-SVR과 S2LS-SVR보다 더 좋은 결과를 보이는 것을 알 수 있다. 그러나 평균제곱오차의 표준편차는 sine 자료에서는 KRR보다 더 크고 sinc 자료에서는  $(l, u) = (100, 1000)$ 일 때 LS-SVR보다 더 큰 값을 가지는 것으로 나타난다.

Table 5.2에 Corn 자료와 Polymer 자료의 KRR, S2KRR, LS-SVR 그리고 S2LS-SVR 추정에 관한 평균제곱오차와 상관계수의 평균과 표준편차를 나타내었다. Polymer 자료에서는 종속변수가  $y_1$ 일 때 평균제곱오차와 상관계수의 평균은 S2KRR이 가장 작은 값으로 좋은 결과를 보이지만 표준편차는 LS-SVR보다 크게 나왔다. 종속변수가  $y_2, y_3$ 일 때는 S2LS-SVR, 그리고  $y_4$ 일 때는 S2KRR이 제일 좋은 수행결과를 보인다. 그러므로 이 Polymer 자료에서도 S2KRR이 더 나은 추정을 하는 것으로 평가된다. Corn 자료에서는 모든 종속변수에서 S2KRR이 제일 좋은 수행결과를 보인다. 다만 종속변수가  $y_1$ 일 때 평균제곱오차의 표준편차와 종속변수가  $y_4$ 일 때 상관계수의 표준편차가 S2LS-SVR보다 조금 더 큰 것을 알 수 있다. 그렇지만 전체적으로 평가하면 S2KRR이 가장 우수한 수행능력을 보인다고 할 수 있다. 이는 S2KRR의 일반화가 더 우수하다는 것을 확인 할 수 있는 결과이다.



**Figure 5.1** KRR estimation (dashed line) and S2KRR estimation (dotted line) of sine function (left) and sinc function (right) are imposed on the scatter plot of simulated data (circle) with true function (real line)

**Table 5.1** Mean and standard deviation of correlation coefficient and MSE of KRR, S2KRR, LS-SVR and S2LS-SVR with simulated data. The boldfaced figure signifies the smallest value.

data	$(l, u)$	estimation method	MSE		correlation coefficient	
			mean	standard deviation	mean	standard deviation
sine data	(40,200)	KRR	1.0376	<b>0.3391</b>	0.9962	0.0029
		S2KRR	<b>0.9412</b>	0.3393	<b>0.9971</b>	<b>0.0026</b>
		LS-SVR	1.0737	0.4165	0.9956	0.0040
		S2LS-SVR	1.0716	0.4163	0.9957	0.0040
	(100,1000)	KRR	0.5100	<b>0.1196</b>	0.9981	0.0014
		S2KRR	<b>0.4659</b>	0.1221	<b>0.9985</b>	<b>0.0013</b>
		LS-SVR	0.5075	0.1234	0.9981	0.0014
		S2LS-SVR	0.5073	0.1234	0.9981	0.0014
sinc data	(40,200)	KRR	1.1798	0.3105	0.9739	0.0165
		S2KRR	<b>1.0626</b>	<b>0.2700</b>	<b>0.9789</b>	<b>0.0130</b>
		LS-SVR	1.2483	0.3358	0.9703	0.0186
		S2LS-SVR	1.1826	0.3392	0.9734	0.0178
	(100,1000)	KRR	0.6677	0.1168	0.9920	0.0029
		S2KRR	<b>0.5962</b>	0.1229	<b>0.9937</b>	<b>0.0027</b>
		LS-SVR	0.6708	<b>0.1148</b>	0.9918	0.0030
		S2LS-SVR	0.6260	0.1226	0.9930	0.0029

**Table 5.2** Mean and standard deviation of correlation coefficient and MSE of KRR, S2KRR, LS-SVR and S2LS-SVR with Polymer data and Corn data. The boldfaced figure signifies the smallest value.

data	response variable	estimation method	MSE		correlation coefficient	
			mean	standard deviation	mean	standard deviation
Polymer data	$y_1$	KRR	0.1235	0.0699	0.6380	0.3446
		S2KRR	<b>0.1138</b>	0.0618	<b>0.7201</b>	0.3203
		LS-SVR	0.1142	<b>0.0578</b>	0.7181	<b>0.2550</b>
		S2LS-SVR	0.1173	0.0588	0.7075	0.2605
	$y_2$	KRR	0.0872	0.0588	0.7024	0.3560
		S2KRR	0.0830	0.0594	<b>0.7581</b>	0.2827
		LS-SVR	0.0848	0.0577	0.7371	0.2469
		S2LS-SVR	<b>0.0773</b>	<b>0.0512</b>	0.7477	<b>0.2436</b>
	$y_3$	KRR	0.0426	0.0317	0.9913	0.0133
		S2KRR	0.0414	0.0278	0.9930	0.0092
		LS-SVR	0.0552	<b>0.0166</b>	0.9956	<b>0.0045</b>
		S2LS-SVR	<b>0.0322</b>	0.0193	<b>0.9960</b>	0.0047
	$y_4$	KRR	0.0318	0.0208	0.9731	0.1406
		S2KRR	<b>0.0292</b>	<b>0.0179</b>	<b>0.9732</b>	0.1406
		LS-SVR	0.0355	0.0221	0.9704	<b>0.1404</b>
		S2LS-SVR	0.0294	0.0190	0.9730	0.1407
Corn data	$y_1$	KRR	0.2122	0.0333	0.9782	0.0089
		S2KRR	<b>0.1806</b>	0.0305	<b>0.9829</b>	<b>0.0077</b>
		LS-SVR	0.2108	0.0308	0.9784	0.0088
		S2LS-SVR	0.2003	<b>0.0254</b>	0.9796	0.0088
	$y_2$	KRR	0.3404	0.0663	0.9552	0.0224
		S2KRR	<b>0.2494</b>	<b>0.0648</b>	<b>0.9766</b>	<b>0.0133</b>
		LS-SVR	0.2826	0.0649	0.9702	0.0181
		S2LS-SVR	0.2588	0.0676	0.9760	0.0157
	$y_3$	KRR	0.2392	0.0698	0.9392	0.0193
		S2KRR	<b>0.1809</b>	<b>0.0365</b>	<b>0.9535</b>	<b>0.0185</b>
		LS-SVR	0.2669	0.0559	0.9074	0.0491
		S2LS-SVR	0.2088	0.0623	0.9461	0.0198
	$y_4$	KRR	0.3897	0.3257	0.9293	0.0647
		S2KRR	<b>0.2788</b>	<b>0.2282</b>	<b>0.9541</b>	0.0422
		LS-SVR	0.3267	0.2541	0.9474	0.0440
		S2LS-SVR	0.3150	0.2661	0.9501	<b>0.0417</b>

## 6. 결론 및 향후과제

초기 추정을 필요로 하지 않는 준지도 회귀모형을 제안한다. 제안된 방법은 기존의 연구에서 제안한 방법에서 요구하는 라벨 없는 자료의 라벨을 추정하는 초기 추정 과정을 필요로 하지 않기 때문에 선택해야 할 모수의 수도 적고 계산과정도 단순하다. 모의실험과 실제 자료 분석을 통해 제안된 방법이 라벨 없는 자료를 잘 활용하여 라벨 있는 자료만 이용하는 방법과 기존의 방법보다 더 우수한 추정을 하는 것을 볼 수 있었다. 그러나 제안된 추정량의 분산이 기대만큼 적어지지 않는 이유를 규명하여 그 문제점을 해결하는 것을 추후 과제로 남긴다. 그리고 모수 선택방법을 개발하는 것도 중요한 과제가 될 것이다.

## 참고문헌

- Belkin, M., Niyogi, P. and Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, **1**, 1-48.
- Chapelle, O., Schölkopf, B. and Zien, A. (2006). *Semi-supervised learning*, MIT Press, Cambridge, MA.
- Cortes, C. and Mohri, M. (2007). On transductive regression. In *Advances in Neural Information Processing System*, **19**, 305-312.
- Exterkate, P. (2012). *Model selection in kernel ridge regression*, CREATES Research Papers from School of Economics and Management, University of Aarhus, Aarhus, Denmark.
- Hofmann, T., Schölkopf, B. and Smola, A. J. (2008) Kernel methods in machine learning. *Annals of Statistics*, **36**, 1171-1220.
- Kimeldorf, G. S. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82-95.
- Lafferty, J. and Wasserman, L. (2008). Statistical analysis of semi-supervised regression. In *Advances in Neural Information Processing Systems*, **20**, 801-808.
- Lin, H. and Lin, C. (2003). *A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods*, Technical Report, Department of Computer Science, National Taiwan University, Taipei, Taiwan.
- Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London A*, **209**, 415-446.
- Nigam, K. and Ghani, R. (2000). Analyzing the effectiveness and applicability of co-training. *Ninth International Conference on Information and Knowledge Management*, 86-93.
- Niyogi, P. (2008). *Manifold regularization and semi-supervised learning: Some theoretical analyses*, Technical Report TR-2008-01, Computer Science Department, University of Chicago, Chicago, IL.
- Rosenberg, C., Hebert, M. and Schneiderman, H. (2005). Semi-supervised self-training of object detection models. *Seventh IEEE Workshop on Applications of Computer Vision*, **1**, 29-36.
- Rosipal, R. and Trejo, L. (2001) Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, **2**, 1667-1689.
- Seok, K. (2012). Study on semi-supervised local constant regression estimation. *Journal of the Korean Data & Information Science Society*, **23**, 579-585.
- Singh, A., Nowak, R. and Zhu, X. (2008). Unlabeled data: Now it helps, now it doesn't. In *Advances in Neural Information Processing Systems*, **21**, 1513-1520.
- Suykens, J.A.K., Gestel, T. V., Bravanter, J. D., Moore, B. D. and Vandewalle, J. (2002). *Least squares support vector machines*, World Scientific, London.
- Szummer, M. and Jaakkola, T. (2002). Information regularization with partially labeled data. In *Advances in Neural Information Processing Systems*, **15**, 1025-1032.
- Ungar, L. H. (1995). UPenn ChemData repository [Machine-readable data repository], Philadelphia, PA. Available electronically via ftp://ftp.cis.upenn.edu/pub/ungar/chemdata.
- Vapnik, V. (1998). *Statistical learning theory*, Wiley, New York.
- Wang, M., Hua, X., Song, Y., Dai, L. and Zhang, H. (2006). Semi-supervised kernel regression. In *Proceeding of the Sixth IEEE International Conference on Data Mining*, 1130-1135.
- Xu, S., An, X., Qiao, X., Zhu, L. and Li, L. (2011) Semisupervised least squares support vector regression machines. *Journal of Information & Computational Science*, **8**, 885-892.
- Xu, Z., King, I. and Lyu, M. R. (2010). *More than semi-supervised learning*, LAP LAMBERT Academic Publishing, London.

- Zhou, X., Ghahramani, Z. and Lafferty, J. (2003). Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, 912-919.
- Zhou, Y. and Goldman, S. (2004). Democratic co-learning. In *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI2004)*, 594-602.
- Zhou, Z. and Li, M. (2007). Semi-supervised regression with co-training style algorithm. *IEEE Transactions on Knowledge and Data Engineering*, **19**, 1479-1493.
- Zhu, D. (2005). *Semi-supervised learning literature survey*, Technical Report, Computer Sciences Department, University of Wisconsin, Madison, WI.
- Zhu, X. and Goldberg, A. (2009). *Introduction to semi-supervised learning*, Morgan & Claypool, London.

# A study on semi-supervised kernel ridge regression estimation<sup>†</sup>

Kyungha Seok<sup>1</sup>

<sup>1</sup>Department of Data Science, Inje University

Received 21 February 2013, revised 10 March 2013, accepted 18 March 2013

## Abstract

In many practical machine learning and data mining applications, unlabeled data are inexpensive and easy to obtain. Semi-supervised learning try to use such data to improve prediction performance. In this paper, a semi-supervised regression method, semi-supervised kernel ridge regression estimation, is proposed on the basis of kernel ridge regression model. The proposed method does not require a pilot estimation of the label of the unlabeled data. This means that the proposed method has good advantages including less number of parameters, easy computing and good generalization ability. Experiments show that the proposed method can effectively utilize unlabeled data to improve regression estimation.

*Keywords:* Generalized cross validation, kernel ridge regression, least squares support vector regression, semi-supervised kernel ridge regression, semi-supervised regression.

---

<sup>†</sup> This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0009705).  
<sup>1</sup> Professor, Institute of Statistical Information, Department of Data Science, Inje University, Kimhae 621-749, Korea. E-mail: statskh@inje.ac.kr