

방사형 기저함수 네트워크를 위한 점진적 학습 알고리즘의 수렴율

이남길*

요약. 본 논문에서는 방사형 기저함수 네트워크(RBFN)에 대해 기대위험(expected risk)의 데이터 의존적 상한(upper bound)을 분석한다. 기대위험의 상한을 네트워크 크기와 훈련 데이터셋 크기의 함수로 명시적으로 표현하므로, RBFN이 기저 함수를 점진적으로 추가하는 경우에도 적용할 수 있다. 제안된 기대위험 상한은 RBFN에 대한 점진적 학습 알고리즘의 점진적 성질의 관점에서 면밀히 분석된다. 제안된 RBFN의 특성이 비선형 회귀 문제에 대한 점진적 학습 알고리즘의 수렴율을 추정하는 데 효과적임을 보이기 위해, 시뮬레이션 데이터를 활용한 실험도 수행되었다.

1. 소개

방사형 기저함수 네트워크(RBFN)는 오랫동안 비모수적 지도학습의 틀 안에서 사용되어 왔다 [1–3]. RBFN은 실수값 함수 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ 로서, 다음과 같이 기저함수들의 선형결합으로 표현된다:

$$(1.1) \quad f(\mathbf{x}) = \sum_{j=1}^m w_j k_\sigma(\mathbf{x}, \mathbf{t}_j).$$

여기서 m 은 기저함수의 개수, $k_\sigma(\mathbf{x}, \mathbf{t}_j)$ 는 중심벡터 $\mathbf{t}_j \in \mathbb{R}^d$ 와 폭 파라미터 $\sigma > 0$ 를 갖는 j 번째 기저함수이다. 지도학습에서는 다음과 같은 기대위험(expected risk)을 최소화하는 함수 f 를 구하는 것을 목표로 한다:

$$Pl_f := E[l(f(X), Y)] = \int l(f(\mathbf{x}), y) dP(\mathbf{x}, y).$$

여기서 $l: \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ 은 손실함수이며, P 는 $\mathbb{R}^d \times \mathbb{R}$ 위에 정의된 확률 분포이다. 하지만 자료가 가지는 실제 분포 P 는 알려져 있지 않기 때문에, 기대위험은 다음과 같이 정의된 경험적 위험(empirical risk)을 통해 추정된다:

$$P_n l_f := \frac{1}{n} \sum_{i=1}^n l(f(\mathbf{x}_i), y_i).$$

여기서 $(\mathbf{x}_i, y_i), i = 1, \dots, n$,은 분포 P 로부터 추출된 훈련 샘플이다.

Vapnik [4]은 경험적 위험과 Vapnik-Chervonenkis (VC) 차원을 사용해 기대위험의 상한을 제안하였다. 하지만 VC 차원은 데이터가 생성된 분포와는 독립적이므로 보통은 기대위험에 대한 보수적인 추정을 하게 된다. 본 연구에서는 Rademacher 복잡도 [5–8]에 기반해 데이터 의존적 상한을 유도한다. Rademacher 복잡도는 데이터의 분포에 의존적이므로, 자료의 분포 변화에 민감한 추정을 수행한다.

2020 Mathematics Subject Classification: 41-04, 65D15.

Key words and phrases: Incremental learning, Rademacher complexity, Radial basis function network.

* This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (Nos. RS-2024-00336424, RS-2024-00358572).

© Kangwon National University Research Institute for Mathematics Sciences, 2025.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

그러나 Rademacher 복잡도의 계산적 어려움으로 인해 실제 실험 결과는 매우 제한적으로 수행되어 왔다. 회귀 문제에서는 Bartlett, Bousquet 및 Mendelson [5], 그리고 Bartlett 및 Mendelson [6]이 Rademacher 복잡도 추정을 위해 커널 공간에 기반한 방법들을 제안하였다. 하지만 제안된 결과들은 커널 공간의 함수들이 가지는 기저 함수의 개수를 구체적으로 명시하지 않으므로 RBFN의 모형 선택에 직접 적용하기 어렵다. 예를 들어, Niyogi 및 Girosi [3]는 기저 함수의 개수 m 을 포함하는 위험 상한을 제안함으로써 최적의 m 값이 근사적으로 $O(n^{1/3})$ 을 따른다는 것을 보였다.

본 연구에서는 convex optimizaition 문제 [9]의 풀이를 통해, 유한개의 기저함수를 가지는 네트워크에 대한 Rademacher 복잡도의 추정치를 구한다. 그리고 기대위험의 추정 상한에 기반하여, 경험적 위험의 점근적 성질을 조사하며, 기저 함수의 개수 및 훈련 샘플 크기 관점에서 수렴율을 근사적으로 구한다. 시뮬레이션 데이터에 기반한 수치 실험을 수행하여, 비선형 회귀 문제에서 RBFN의 수렴율 추정에서 제안된 점근적 성질이 매우 효과적임을 확인한다.

2. 주요 결과

여기서는 σ 는 고정된 상수임을 가정한다. $\mathbb{R}^d \times \mathbb{R}^d$ 상에서 정의된 기저함수 k_σ 는 대칭이며 양정치 함수이다. 즉, 어떠한 정수 r 및 수열 $c_i \in \mathbb{R}$ 및 $\mathbf{t}_i \in \mathbb{R}^d$, $i = 1, \dots, r$ 에 대하여도 $\sum_{i=1}^r \sum_{j=1}^r c_i c_j k_\sigma(\mathbf{t}_i, \mathbf{t}_j) \geq 0$ 이 성립한다. 어느 $0 < C < \infty$ 에 대해, 다음과 같이 \mathcal{F}_m 를 m 개의 기저함수를 가진 네트워크로 정의한다:

$$(2.1) \quad \mathcal{F}_m = \mathcal{F}_m(C) = \left\{ \mathbf{x} \mapsto \sum_{j=1}^m w_j k_\sigma(\mathbf{x}, \mathbf{t}_j) \mid \sum_{i=1}^m \sum_{j=1}^m w_i w_j k_\sigma(\mathbf{t}_i, \mathbf{t}_j) \leq C^2 \right\}.$$

어느 $0 < D < \infty$ 에 대해, $l : \mathbb{R} \times \mathbb{R} \rightarrow [0, D]$ 을 유계 손실함수로서 다음과 같이 첫 인수에 대해 L -Lipschitz임을 가정한다:

$$\forall y, \hat{y}_1, \hat{y}_2, \quad |l(\hat{y}_1, y) - l(\hat{y}_2, y)| \leq L|\hat{y}_1 - \hat{y}_2|.$$

정수 $r \leq n$ 및 $m \leq n$ 에 대해서, $r \times m$ 행렬 $\mathbf{K}_{r,m}$ 을 모든 $i = 1, \dots, r$ 및 $j = 1, \dots, m$ 에 대해 $(\mathbf{K}_{r,m})_{ij} = k_\sigma(\mathbf{x}_i, \mathbf{x}_j)$ 으로 정의한다. 본 논문의 주요 결과는 다음과 같이 기술된다:

정리2.1. 모든 $x > 0$ 에 대해, 적어도 $1 - 2e^{-x}$ 의 확률로 다음의 부등식이 성립한다:

$$(2.2) \quad \sup_{f \in \mathcal{F}_m} \{P_l f - P_n l_f\} \leq A_{m,n} + \frac{Dx}{n} + \sqrt{\frac{2Dx}{n} A_{m,n} + \left(\frac{Dx}{n}\right)^2} + D\sqrt{\frac{2x}{n}},$$

여기서

$$A_{m,n} = \frac{2CL}{n} \sqrt{\text{trace}(\mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T)}.$$

3. 최소제곱법

이번 장에서는 정리 2.1을 특정한 최소제곱 추정법에 적용한다. 이 최소제곱법은 중심 벡터 \mathbf{t}_j , $j = 1, \dots, m$ 를 훈련 데이터셋의 첫 m 개의 벡터들로 설정한다. 즉, $\mathbf{t}_j = \mathbf{x}_j$ 이다. Σ 는 $m \times m$ 행렬로서 성분 $(\Sigma)_{ij} = E[k_\sigma(X, \mathbf{x}_i) k_\sigma(X, \mathbf{x}_j)]$ 을 가지며, Θ 를 $m \times 1$ 벡터로서 성분들 $(\Theta)_{i1} = E[k_\sigma(X, \mathbf{x}_i) Y]$ 를 가진다고 하자. 정리 2.1에 주어진 기대위험의 상한과 관련된 점근적 성질을 최소제곱법에 대해 적용하여 다음과 같은 결과를 얻을 수 있다:

따름정리3.1. k_σ 는 대칭이고 양정치이며 $k_\sigma(\mathbf{x}_j, \mathbf{x}_j) \leq 1$ 을 모든 $j = 1, \dots, m$ 에 대해 만족한다고 하자. 모든 $i = 1, \dots, m$ 및 $j = 1, \dots, m$ 에 대해 $E[(k_\sigma(X, \mathbf{x}_i) k_\sigma(X, \mathbf{x}_j))^2] < \infty$ 이고 $E[(k_\sigma(X, \mathbf{x}_i) Y)^2] < \infty$ 임을 가정한다. 만일 $m < n$ 이고 $n \rightarrow \infty$ 에 대해 $n - m \rightarrow \infty$ 이고,

어떤 $p \geq 0$ 에 대해 $C = O(m^p)$ 이며, $D = O(1)$ 및 $L = O(1)$ 이라면, 다음과 같은 부등식이 성립한다:

$$(3.1) \quad \sup_{f \in \mathcal{F}_m} \{Pl_f - P_n l_f\} = O\left(\frac{m^p}{n^{1/2}}\right).$$

더 나아가, 만일 어떤 $B > 0$ 에 대해 $|Y| \leq B$ 이고, 어떤 $q \geq 0$ 에 대해 Σ 의 최소고유값이 $\lambda_{\min}^{-1}(\Sigma) = O_p(m^q)$ 을 만족한다면, 최소제곱추정 \hat{f} 은 다음 식을 만족한다:

$$(3.2) \quad Pl_{\hat{f}} - P_n l_{\hat{f}} = O\left(\frac{m^{(q+1)/4}}{n^{1/2}}\right).$$

4. 수치적 실험

본 연구에서 제안된 기대위험 상한의 효과를 입증하기 위해 시뮬레이션 데이터셋을 기반으로 실험을 수행하였다. 우리는 가우시안 커널 함수를 사용한 RBFN을 적용하였다: $k_\sigma(\mathbf{x}, \mathbf{t}_j) = \exp(-\|\mathbf{x} - \mathbf{t}_j\|^2 / (2\sigma^2))$. 여기서 중심 벡터 \mathbf{t}_j 는 훈련 데이터의 입력 벡터들 중 무작위로 선택되었다. 제곱 오차의 합의 평균을 이용해 테스트 오차와 훈련 오차를 구하였다.

시뮬레이션 데이터로는 Hastie 등 [10]의 11장에서 설명된 radial 데이터를 사용하였다. 이 radial 데이터는 10차원 입력 벡터 $\mathbf{x}_i = (x_{i1}, \dots, x_{i10})^T$ 와 출력 값 y_i 로 구성되어 있다. 입력 벡터의 각 성분 x_{ij} 는 표준 정규 분포 $N(0, 1)$ 로부터 독립적으로 생성되며, 출력 값은 다음과 같이 계산된다:

$$y_i = \prod_{j=1}^{10} \phi(x_{ij}) + \epsilon_i.$$

여기서 $\phi(t) = (2\pi)^{-1/2} \exp(-t^2/2)$ 는 표준 정규 분포의 밀도 함수이며, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ 는 평균이 0이고 분산이 σ_ϵ^2 인 노이즈 항이다. 노이즈 분산 σ_ϵ^2 은 신호 대 잡음비(signal-to-noise ratio, SNR)가 4 또는 64가 되도록 설정하였다. 즉,

$$\frac{\text{Var}(E[Y|X])}{\text{Var}(Y - E[Y|X])} = \frac{\text{Var}(f(X))}{\text{Var}(\epsilon)} = 4 \text{ 또는 } 64.$$

훈련 샘플은 다양한 크기로 설정하였으며, 테스트 샘플은 10000개로 고정하였다. 훈련 샘플 생성은 30회 반복하였고, 각 실험의 결과는 평균값을 취해 분석하였다.

그림 1 (a)는 기저 함수 수 m 이 200으로 고정된 상태에서, 훈련 샘플 크기에 따라 최소제곱 추정치 \hat{f} 의 테스트 오차와 훈련 오차 간의 차이 $Pl_{\hat{f}} - P_n l_{\hat{f}}$ 를 보여준다. 도표에서 확인할 수 있듯이, 차이에 대한 추정 상한(검정 점선)은 실제 차이(빨간 실선)보다 훨씬 크지만, 감소 속도(비율)는 유사하다. 특히 주목할 점은, 추정된 상한의 기울기가 대략 -0.5 로 일정하다는 것이다. 이는 기저 함수 수 m 이 고정된 경우, $Pl_{\hat{f}} - P_n l_{\hat{f}} = O_p(n^{-1/2})$ 이라는 이론적 결과를 뒷받침해준다. 반면, 실제 차이의 기울기는 대략 -1 근처로 변화하며, 이는 최소제곱법이 더 빠른 수렴율을 보인다는 것을 나타낸다. 그림 1 (b)는 훈련 샘플 수 n 을 1000으로 고정했을 때, 기저 함수 수 m 의 변화에 따른 테스트 오차와 훈련 오차 간의 차이를 나타낸다. 이 그림은 기저 함수 수가 증가함에 따라 실제 차이가 커지고, 추정된 상한 역시 거의 유사한 비율로 증가함을 보여준다. 그림 1(a)와 (b) 각각의 경우에 대해, 신호 대 잡음비(SNR)는 4(위 행) 또는 64(아래 행)로 설정되었고, 커널 폭 파라미터 σ 는 0.5(맨 왼쪽 열), 4(가운데 열), 16(맨 오른쪽 열)로 설정되었다.

그림 2 (a)와 (b)는 테스트 오차와 훈련 오차 간의 차이의 수렴율을 보다 면밀히 보여준다. 비교를 위해 최소 고유값 $\lambda_{\min}(\Sigma)$ 의 값도 함께 나타내었다. 그림 2 (a)는 훈련 샘플 수 n 이 증가함에 따라 $\lambda_{\min}(\Sigma)$ 의 값이 거의 일정하게 유지됨을 보여주며, 이는 $\lambda_{\min}(\Sigma)$ 가 n 에 영향을 받지 않는다는 사실을 정당화해준다. 또한, 테스트 오차와 훈련 오차 간의 실제 차이는, 추정된 상한보다 더 빠르게 감소함을 확인할 수 있다. 반면, 그림 2 (b)는 기저 함수 수 m 이 증가함에 따라, 실제 차이와 상한이 유사한 속도로 증가함을 나타낸다. 또한, $\lambda_{\min}(\Sigma)$

의 감소 기울기는 약 -4.8 , 상한의 증가 기울기는 약 1.0 인데, 이는 따름정리 3.1의 결과인 $1.0 \leq (4.8 + 1)/4$ 와 일치한다.

그림 2 (c)는 기저 함수 수 m 과 훈련 샘플 수 n 이 다양한 비율로 동시에 증가할 때, 테스트 오차와 훈련 오차 간의 실제 차이의 변화를 보여준다. 즉, m 은 고정된 상수 a 에 따라 $500 \times (n/5000)^a$ 로 설정된다. 실제로, 그림 2 (a)에서는 m 이 고정된 경우, n 이 증가함에 따라 실제 차이의 감소 속도가 $n^{-1.2}$ 로 나타났다. 따름정리 3.1에 따르면, 실제 차이의 수렴율은 $m^p/n^{1.2}$ 형태를 따르며, 그림 2 (b)의 추정 상한 기울기 결과를 기반으로 p 는 1로 설정된다. 따라서, 만약 $m \propto n^a$ 라면, 수렴율은 $n^{a \times 1.0 - 1.2}$ 로 추정된다. 그림 2 (c)에서는 a 값을 각각 $2/3, 1, 4/3, 3/2$ 로 설정하였으며, 이 경우 예상되는 수렴율은 각각 $n^{-0.53}, n^{-0.2}, n^{0.13}$, 및 $n^{0.3}$ 이다. 실제로 계산된 수렴율은 $n^{-0.5}, n^{-0.2}, n^{0.2}$, 및 $n^{0.4}$ 로, 이는 추정된 값들과 매우 유사한 결과를 보여준다.

5. 증명

$\epsilon_1, \dots, \epsilon_n$ 를 $P(\epsilon_i = 1) = P(\epsilon_i = -1) = 1/2$ 의 확률을 갖는 독립적 확률 변수라고 하자. \mathcal{F} 는 \mathbb{R}^d 로부터 \mathbb{R} 로의 어떤 함수 클래스이고, \mathcal{G} 는 $\mathbb{R}^d \times \mathbb{R}$ 로부터 \mathbb{R} 로의 어떤 함수 클래스일 때, 두 확률변수 $R_n\mathcal{F}$ 및 $R_n\mathcal{G}$ 를 다음과 같이 정의한다: $R_n\mathcal{F} = \sup_{f \in \mathcal{F}} n^{-1} \sum_{i=1}^n \epsilon_i f(\mathbf{x}_i)$ 및 $R_n\mathcal{G} = \sup_{g \in \mathcal{G}} n^{-1} \sum_{i=1}^n \epsilon_i g(\mathbf{x}_i, y_i)$. \mathcal{F} 또는 \mathcal{G} 에 대한 경험적 Rademacher 복잡도는 각각 다음과 같이 정의한다:

$$E_\epsilon R_n\mathcal{F} = E \left(\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \epsilon_i f(\mathbf{x}_i) \middle| \mathbf{x}_1, \dots, \mathbf{x}_n \right),$$

$$E_\epsilon R_n\mathcal{G} = E \left(\sup_{g \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^n \epsilon_i g(\mathbf{x}_i, y_i) \middle| (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n) \right).$$

여기서 기댓값은 $\epsilon_1, \dots, \epsilon_n$ 에 대해 취해진다. \mathcal{F} 또는 \mathcal{G} 에 대한 Rademacher 복잡도는 각각 다음과 같이 정의한다:

$$ER_n\mathcal{F} = E_{\mathbf{x}} E_\epsilon R_n\mathcal{F},$$

$$ER_n\mathcal{G} = E_{(\mathbf{x}, y)} E_\epsilon R_n\mathcal{G}$$

이며, 여기서, $E_{\mathbf{x}}$ 는 $\mathbf{x}_1, \dots, \mathbf{x}_n$ 에 대한 기댓값이며, $E_{(\mathbf{x}, y)}$ 는 $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ 에 대한 기댓값이다.

정리 2.1의 결과는 다음 보조정리들로부터 이끌어낼 수 있다:

보조정리 5.1. (McDiarmid의 부등식, [11]) Z_1, \dots, Z_n 을 집합 A 에서 값을 갖는 확률변수들 이라고 하고, $f: A^n \rightarrow \mathbb{R}$ 가 다음을 만족한다고 가정하자: 모든 $1 \leq i \leq n$ 에 대해,

$$\sup_{z_1, \dots, z_n, z'_i \in A} |f(z_1, \dots, z_n) - f(z_1, \dots, z_{i-1}, z'_i, z_{i+1}, \dots, z_n)| \leq c_i.$$

그러면 모든 $t > 0$ 에 대해,

$$P(f(Z_1, \dots, Z_n) - Ef(Z_1, \dots, Z_n) \geq t) \leq e^{-2t^2 / \sum_{i=1}^n c_i^2}$$

이다.

보조정리 5.2. (대칭화 부등식, [5, Lemma A.5], [7]) \mathcal{G} 가 $\mathbb{R}^d \times \mathbb{R}$ 로부터 \mathbb{R} 로의 함수의 집합 이라고 하자. 그러면,

$$E \sup_{g \in \mathcal{G}} \{Pg - Png\} \leq 2ER_n\mathcal{G}$$

이다.

보조정리 5.3. (Lemma A.4 in [5]) $x > 0$ 을 고정하고, \mathcal{G} 를 $[a, b]$ 내에서 치역을 갖는 함수들의 집합이라고 하자. 그러면, 적어도 $1 - e^{-x}$ 의 확률로,

$$(5.1) \quad ER_n\mathcal{G} \leq \inf_{\alpha \in (0,1)} \left(\frac{1}{1-\alpha} E_\epsilon R_n\mathcal{G} + \frac{(b-a)x}{4n\alpha(1-\alpha)} \right)$$

이다. 또한, 적어도 $1 - e^{-x}$ 의 확률로,

$$E_\epsilon R_n\mathcal{G} \leq \inf_{\alpha > 0} \left((1+\alpha)ER_n\mathcal{G} + \frac{(b-a)x}{2n} \left(\frac{1}{2\alpha} + \frac{1}{3} \right) \right)$$

이다.

정리 2.1을 증명하기에 앞서, 보조정리 5.1 및 5.2를 바탕으로 다음 명제를 증명한다:

명제 5.4. \mathcal{G} 를 $\mathbb{R}^d \times \mathbb{R}$ 로부터 $[0, D]$ 로의 함수들의 집합이라고 하자. 그러면, 모든 $x > 0$ 에 대해, 적어도 $1 - e^{-x}$ 의 확률로,

$$\sup_{g \in \mathcal{G}} \{Pg - P_n g\} \leq E \sup_{g \in \mathcal{G}} \{Pg - P_n g\} + D \sqrt{\frac{2x}{n}}$$

이다. 더 나아가, 적어도 $1 - e^{-x}$ 의 확률로,

$$(5.2) \quad \sup_{g \in \mathcal{G}} \{Pg - P_n g\} \leq 2ER_n\mathcal{G} + D \sqrt{\frac{2x}{n}}$$

이다.

증명. P_n^i 를 i 번째 표본 데이터 (\mathbf{x}_i, y_i) 를 (\mathbf{x}'_i, y'_i) 로 교체하여 수립한 경험적 분포라고 하자. $g^* = \arg \sup_{g \in \mathcal{G}} \{Pg - P_n g\}$ 로 정의하면,

$$\begin{aligned} \sup_{g \in \mathcal{G}} \{Pg - P_n g\} - \sup_{g \in \mathcal{G}} \{Pg - P_n^i g\} &\leq P g^* - P_n g^* - P g^* + P_n^i g^* \\ &= \frac{1}{n} (g^*(\mathbf{x}'_i, y'_i) - g^*(\mathbf{x}_i, y_i)) \\ &\leq \frac{2}{n} D. \end{aligned}$$

동일한 방식으로, $g^{**} = \arg \sup_{g \in \mathcal{G}} \{Pg - P_n^i g\}$ 이라고 하면,

$$\sup_{g \in \mathcal{G}} \{Pg - P_n g\} - \sup_{g \in \mathcal{G}} \{Pg - P_n^i g\} \geq -\frac{2}{n} D$$

을 이끌어낼 수 있다. 보조정리 5.1의 McDiarmid의 부등식을 함수 $\sup_{g \in \mathcal{G}} \{Pg - P_n g\}$ 에 적용하면 첫 번째 결과를 이끌어낼 수 있다. 두 번째 결과는 보조정리 5.2의 대칭화 부등식을 적용하면 이끌어 낼 수 있다. \square

추가적으로, 다음의 명제는 m 개의 기저함수를 갖는 네트워크의 클래스 \mathcal{F}_m 에 관한 것이다.

명제 5.5.

$$(5.3) \quad E_\epsilon R_n \mathcal{F}_m \leq \frac{C}{n} \sqrt{\text{trace}(\mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T)}.$$

증명. $R_n \mathcal{F}_m = \sup_{f \in \mathcal{F}_m} n^{-1} \sum_{i=1}^n \epsilon_i f(\mathbf{x}_i)$ 을 계산하기 위해, $n^{-1} \sum_{i=1}^n \epsilon_i f(\mathbf{x}_i)$ 의 최대화 문제를 푸는 것을 고려하며, 이것은 아래와 같은 최소화 문제로 기술할 수 있다:

$$(5.4) \quad \begin{aligned} \text{목적: } & \underset{\mathbf{w}}{\text{minimize}} \quad -\frac{1}{n} \boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{w} \\ \text{조건: } & \mathbf{w}^T \mathbf{K}_{mm} \mathbf{w} \leq C^2. \end{aligned}$$

위 문제로부터, Lagrange 승수 $\lambda \geq 0$ 를 도입하여, 다음과 같이 Lagrangian을 기술할 수 있다: $G(\mathbf{w}, \lambda) = -n^{-1} \boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{w} + \lambda(\mathbf{w}^T \mathbf{K}_{mm} \mathbf{w} - C^2)$. Lagrangian을 \mathbf{w} 에 관해 미분함으로써 쌍대 함수를 다음과 같이 구할 수 있다:

$$(5.5) \quad g(\lambda) = \min_{\mathbf{w} \in \mathbb{R}^m} G(\mathbf{w}, \lambda) = -(4n^2 \lambda)^{-1} \boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon} - \lambda C^2.$$

이때, 최소값은 $\mathbf{w} = (2n\lambda)^{-1} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}$ 일 때 얻어진다. 이를 바탕으로, 쌍대 문제는 다음과 같이 유도된다:

$$(5.6) \quad \begin{aligned} \text{목적: } & \underset{\lambda}{\text{maximize}} \quad -\frac{1}{4n^2 \lambda} \boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon} - \lambda C^2 \\ \text{조건: } & \lambda \geq 0. \end{aligned}$$

(5.6)에 주어진 쌍대 함수 $g(\lambda)$ 를 미분하여 쌍대 최적쌍 (λ^*, d^*) 을 구할 수 있는데, 이것은 각각 λ 의 최적점과 쌍대 함수 $g(\lambda)$ 의 최적값이다:

$$\begin{aligned} \lambda^* &= \frac{1}{2nC} \sqrt{\boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}} \\ d^* &= -\frac{C}{n} \sqrt{\boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}}. \end{aligned}$$

또한, 해당하는 원초(primal) 최적점은 $\mathbf{w}^* = (2n\lambda^*)^{-1} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}$ 이다. 따라서,

$$R_n \mathcal{F}_m = -d^* = \frac{C}{n} \sqrt{\boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}}.$$

임을 이끌어낼 수 있다. 기댓값 E_ϵ 을 취하고 쥘센의 부등식(Jensen's Inequality)을 적용하면, 다음과 같이 부등식을 이끌어낼 수 있다:

$$\begin{aligned} E_\epsilon R_n \mathcal{F}_m &= \frac{C}{n} E_\epsilon \sqrt{\boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}} \\ &\leq \frac{C}{n} \sqrt{E_\epsilon \boldsymbol{\epsilon}^T \mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \boldsymbol{\epsilon}} \\ &= \frac{C}{n} \sqrt{\text{trace}(\mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T)}. \end{aligned}$$

□

정리 2.1의 증명은 다음과 같이 주어진다:

정리 2.1의 증명. 기호 표기 단순화를 위해 \mathcal{F}_m 을 \mathcal{F} 으로 표기한다. 함수 $l_f : \mathbb{R}^d \times \mathbb{R} \rightarrow [0, D]$ 는 $l_f(\mathbf{x}, y) = l(f(\mathbf{x}), y)$ 를 나타내며, $l_{\mathcal{F}}$ 는 집합 $l_{\mathcal{F}} = \{l_f | f \in \mathcal{F}\}$ 를 나타낸다. 우선, 보조정리 5.3의 부등식 (5.1)의 우변의 최소값을 구한다. 어떤 $t_1, t_2 > 0$ 에 대해, $d/d\alpha(t_1(1-\alpha)^{-1} + t_2\alpha^{-1}(1-\alpha)^{-1}) = 0$ 으로부터 $\alpha^* = (-t_2 + \sqrt{t_2^2 + t_1 t_2})/t_1$ 이고 $0 < \alpha^* < 1$ 임을 이끌어낼 수 있다. 따라서,

$$\inf_{\alpha \in (0,1)} \left(\frac{1}{1-\alpha} E_\epsilon R_n l_{\mathcal{F}} + \frac{Dx}{4n\alpha(1-\alpha)} \right) = E_\epsilon R_n l_{\mathcal{F}} + \frac{Dx}{2n} + 2 \sqrt{\left(\frac{Dx}{4n} \right) E_\epsilon R_n l_{\mathcal{F}} + \left(\frac{Dx}{4n} \right)^2}$$

이다. 더 나아가, [12]의 식 (4.20)을 이용해 $R_n l_{\mathcal{F}} \leq L R_n \mathcal{F}$ 임을 유도할 수 있다. 다음으로, 보조정리 5.3의 결과 (5.1)과 명제 5.4의 결과 (5.2)을 $\mathcal{G} = l_{\mathcal{F}}$ 에 적용하여 다음을 이끌어낼 수 있다: 적어도 $1 - 2e^{-x}$ 의 확률로,

$$\sup_{f \in \mathcal{F}} \{Pl_f - P_n l_f\} \leq 2 \inf_{\alpha \in (0,1)} \left(\frac{1}{1-\alpha} E_{\epsilon} R_n l_{\mathcal{F}} + \frac{Dx}{4n\alpha(1-\alpha)} \right) + D\sqrt{\frac{2x}{n}}.$$

결과적으로, 적어도 확률 $1 - 2e^{-x}$ 로,

$$\begin{aligned} \sup_{f \in \mathcal{F}} \{Pl_f - P_n l_f\} &\leq 2LE_{\epsilon} R_n \mathcal{F} + \frac{Dx}{n} + 4\sqrt{\left(\frac{Dx}{4n}\right) LE_{\epsilon} R_n \mathcal{F} + \left(\frac{Dx}{4n}\right)^2} + D\sqrt{\frac{2x}{n}} \\ &\leq A_{m,n} + \frac{Dx}{n} + \sqrt{\frac{2Dx}{n} A_{m,n} + \left(\frac{Dx}{n}\right)^2} + D\sqrt{\frac{2x}{n}} \end{aligned}$$

이며, 여기서

$$A_{m,n} = \frac{2CL}{n} \sqrt{\text{trace}(\mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T)}$$

이다. 여기서, 마지막 부등식은 명제 5.5로부터 나온다. □

따름정리 3.1의 증명은 다음과 같이 주어진다:

증명. 행렬 \mathbf{K}_{nm} 을 두 블록으로 파티션한다: $\mathbf{K}_{nm}^T = (\mathbf{K}_{mm}^T, \mathbf{K}_{\Delta m}^T)$. 여기서 $\mathbf{K}_{\Delta m}$ 은 $(n-m) \times (n-m)$ 행렬로서, $(\mathbf{K}_{\Delta m})_{ij} = k_{\sigma}(\mathbf{x}_{m+i}, \mathbf{x}_j)$ 이다. $E[(k_{\sigma}(X, \mathbf{x}_i)k_{\sigma}(X, \mathbf{x}_j))^2] < \infty$ 이므로, 중심극한정리로부터, $n^{-1}\mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m}$ 의 (i, j) 원소는 다음과 같이 수렴한다:

$$\begin{aligned} \frac{1}{n} (\mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m})_{ij} &= \frac{1}{n} \sum_{r=1}^{n-m} k_{\sigma}(\mathbf{x}_{m+r}, \mathbf{x}_i) k_{\sigma}(\mathbf{x}_{m+r}, \mathbf{x}_j) \\ &= E[k_{\sigma}(X, \mathbf{x}_i)k_{\sigma}(X, \mathbf{x}_j)] + O_p(n^{-1/2}) \\ &= (\boldsymbol{\Sigma})_{ij} + O_p(n^{-1/2}). \end{aligned}$$

그리고 $m^{-1}\mathbf{K}_{mm}^2$ 의 (i, j) 원소는 다음과 같이 수렴한다:

$$\begin{aligned} \frac{1}{m} (\mathbf{K}_{mm}^2)_{ij} &= \frac{1}{m} \sum_{r=1}^m k_{\sigma}(\mathbf{x}_r, \mathbf{x}_i) k_{\sigma}(\mathbf{x}_r, \mathbf{x}_j) \\ (5.7) \quad &= E[k_{\sigma}(X, \mathbf{x}_i)k_{\sigma}(X, \mathbf{x}_j)] + O_p(m^{-1/2}) \\ &= (\boldsymbol{\Sigma})_{ij} + O_p(m^{-1/2}). \end{aligned}$$

이를 통해, 다음을 유도할 수 있다:

$$\begin{aligned} \left(\frac{1}{m} \mathbf{K}_{mm}^2\right)^{-1} \cdot \frac{1}{n} \mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m} &= (\boldsymbol{\Sigma} \cdot (1 + O_p(m^{-1/2})))^{-1} \boldsymbol{\Sigma} \cdot (1 + O_p(n^{-1/2})) \\ &= \mathbf{I}_m \cdot (1 + O_p(m^{-1/2}))^{-1} (1 + O_p(n^{-1/2})) \\ &= \mathbf{I}_m \cdot (1 + O_p(m^{-1/2})) (1 + O_p(n^{-1/2})) \\ &= \mathbf{I}_m \cdot (1 + O_p(m^{-1/2}) + O_p(n^{-1/2})). \end{aligned}$$

여기서, \mathbf{I}_m 은 $m \times m$ 항등행렬이다. $\mathbf{K}_{nm}^T \mathbf{K}_{nm} = \mathbf{K}_{mm}^2 + \mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m}$ 이므로, 대각합으로부터 다음 부등식을 유도할 수 있다:

$$\begin{aligned}
\text{trace}(\mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T) &= \text{trace}(\mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T \mathbf{K}_{nm}) \\
&= \text{trace}(\mathbf{K}_{mm} + \mathbf{K}_{mm}^{-1} \mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m}) \\
&\leq m + \text{trace}(\mathbf{K}_{mm}^{-1} \mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m}) \\
&= m + m^{-1}n \cdot \text{trace}(\mathbf{K}_{mm} \cdot (m^{-1} \mathbf{K}_{mm}^2)^{-1} \cdot n^{-1} \mathbf{K}_{\Delta m}^T \mathbf{K}_{\Delta m}) \\
&= m + m^{-1}n \cdot \text{trace}(\mathbf{K}_{mm}) \cdot (1 + O_p(m^{-1/2}) + O_p(n^{-1/2})) \\
&\leq m + n \cdot (1 + O_p(m^{-1/2}) + O_p(n^{-1/2})) \\
&= O_p(m + n).
\end{aligned}$$

이를 이용하여 다음을 이끌어낸다:

$$\begin{aligned}
(5.8) \quad A_{m,n} &= \frac{2CL}{n} \sqrt{\text{trace}(\mathbf{K}_{nm} \mathbf{K}_{mm}^{-1} \mathbf{K}_{nm}^T)} \\
&= O_p(m^p n^{-1} \sqrt{m+n}) \\
&= O_p(m^p n^{-1} (\sqrt{m} + \sqrt{n})) \\
&= O_p(m^{1/2+p} n^{-1} + m^p n^{-1/2}).
\end{aligned}$$

결과적으로, 명제 2.1의 식 (2.2)의 우변은 다음을 만족한다:

$$\begin{aligned}
(5.9) \quad A_{m,n} + \frac{Dx}{n} + \sqrt{\frac{2Dx}{n} A_{m,n} + \left(\frac{Dx}{n}\right)^2} + D\sqrt{\frac{2x}{n}} \\
&= O_p(m^{1/2+p} n^{-1} + m^p n^{-1/2}) + O_p(n^{-1}) + O_p\left(\sqrt{n^{-1} \cdot (m^{1/2+p} n^{-1} + m^p n^{-1/2})}\right) \\
&\quad + O_p(n^{-1/2}) \\
&= O_p(m^{1/2+p} n^{-1} + m^p n^{-1/2} + m^{1/4+p/2} n^{-1} + m^{p/2} n^{-3/4} + n^{-1/2}).
\end{aligned}$$

$m \leq n$ 이므로 (5.9)의 첫 두 항은 $m^{1/2+p} n^{-1} + m^p n^{-1/2} = m^p n^{-1/2} (m^{1/2} n^{-1/2} + 1) \leq 2m^p n^{-1/2}$ 을 만족하고, 그 다음의 두 항은 $m^{1/4+p/2} n^{-1} + m^{p/2} n^{-3/4} = m^{p/2} n^{-3/4} (m^{1/4} n^{-1/4} + 1) \leq 2m^{p/2} n^{-3/4}$ 을 만족한다. 그러므로, 식 (5.9)는 다음과 같이 간단히 정리할 수 있다:

$$\begin{aligned}
(5.10) \quad A_{m,n} + \frac{Dx}{n} + \sqrt{\frac{2Dx}{n} A_{m,n} + \left(\frac{Dx}{n}\right)^2} + D\sqrt{\frac{2x}{n}} \\
&= O_p(m^p n^{-1/2} + m^{p/2} n^{-3/4} + n^{-1/2}).
\end{aligned}$$

만일 $p \geq 0$ 이면 (5.10)는 $O_p(m^p n^{-1/2})$ 가 되며, 만일 $p < 0$ 이면 (5.10)는 $O_p(n^{-1/2})$ 가 된다.

최소제곱추정에 의해 결정된 $\mathbf{w} = (w_1, \dots, w_m)^T$ 추정값은 다음과 같이 표현된다:

$$\hat{\mathbf{w}} = (\mathbf{K}_{nm}^T \mathbf{K}_{nm})^{-1} \mathbf{K}_{nm}^T \mathbf{Y}.$$

식 $C = O_p(m^p)$ 에서 p 의 값을 구하기 위해, $\hat{f} \in \mathcal{F}_m$ 는 다음 식을 만족함을 이용한다:

$$(5.11) \quad \sum_{i=1}^m \sum_{j=1}^m \hat{w}_i \hat{w}_j k_\sigma(\mathbf{x}_i, \mathbf{x}_j) = \hat{\mathbf{w}}^T \mathbf{K}_{mm} \hat{\mathbf{w}} \leq C^2.$$

우선, 중심극한정리에 의해,

$$\begin{aligned}\hat{\mathbf{w}} &= \left(\frac{1}{n} \mathbf{K}_{nm}^T \mathbf{K}_{nm} \right)^{-1} \left(\frac{1}{n} \mathbf{K}_{nm}^T \mathbf{Y} \right) \\ &= \boldsymbol{\Sigma}^{-1} \boldsymbol{\Theta} + O_p(n^{-1/2}).\end{aligned}$$

임을 얻을 수 있다. 따라서, 식 (5.11)은 아래와 같이 쓸 수 있다:

$$\begin{aligned}(5.12) \quad \hat{\mathbf{w}}^T \mathbf{K}_{mm} \hat{\mathbf{w}} &= (\hat{\mathbf{w}}^T \mathbf{K}_{mm}) \mathbf{K}_{mm}^{-1} (\mathbf{K}_{mm} \hat{\mathbf{w}}) \\ &\leq \lambda_{\min}^{-1}(\mathbf{K}_{mm}) \|\mathbf{K}_{mm} \hat{\mathbf{w}}\|^2 \\ &= \lambda_{\min}^{-1}(\mathbf{K}_{mm}) \|\mathbf{K}_{mm} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Theta}\|^2 \cdot (1 + O_p(n^{-1/2})).\end{aligned}$$

식 (5.7)를 통해, $\lambda_{\min}(m^{-1} \mathbf{K}_{mm}^2) = \lambda_{\min}(\boldsymbol{\Sigma})(1 + O_p(m^{-1/2}))$ 임을 알 수 있으며, 이를 통해

$$\begin{aligned}\lambda_{\min}^{-1}(\mathbf{K}_{mm}) &= m^{-1/2} \lambda_{\min}^{-1/2}(\boldsymbol{\Sigma})(1 + O_p(m^{-1/2})) \\ &= O_p(m^{(q-1)/2})\end{aligned}$$

임을 유도할 수 있다. 더 나아가, 중심극한정리에 의해, $m^{-1} \mathbf{K}_{mm} \mathbf{Y}_m = \boldsymbol{\Theta} \cdot (1 + O_p(m^{-1/2}))$ 이므로,

$$\begin{aligned}\|\mathbf{K}_{mm} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Theta}\|^2 &= \|\mathbf{K}_{mm} (m^{-1} \mathbf{K}_{mm}^2)^{-1} (m^{-1} \mathbf{K}_{mm} \mathbf{Y}_m)\|^2 (1 + O_p(m^{-1/2})) \\ &= \|\mathbf{Y}_m\|^2 (1 + O_p(m^{-1/2})) \\ &\leq m B^2 (1 + O_p(m^{-1/2})) \\ &= O_p(m)\end{aligned}$$

임을 구할 수 있다. 따라서, 식 (5.12)는 다음 식처럼 단순화된다:

$$\hat{\mathbf{w}}^T \mathbf{K}_{mm} \hat{\mathbf{w}} = O_p(m^{(q+1)/2}).$$

따라서, 식 (5.11)로부터, 다음을 얻을 수 있다:

$$C = O_p(m^{(q+1)/4}).$$

□

6. 결론

본 논문에서는 방사형 기저함수 네트워크(RBFN)에 대한 기대위험의 상한을 도출하였다. 이 상한은 유한 개의 기저 함수를 갖는 RBFN의 공간에 대해 Rademacher 복잡도를 정확히 계산함으로써 얻어진 것이다. 도출된 상한은 행렬 \mathbf{K}_{nm} 및 \mathbf{K}_{mm} 를 통해 훈련 데이터에 의존함을 보여주었으며, 이 행렬들은 손쉽게 계산될 수 있다. 또한, 기대위험의 상한을 바탕으로 상한의 점진적 성질도 도출하였다. 이 상한은 훈련 데이터 수 n , 기저 함수 수 m , 그리고 RBFN 공간의 크기와 관련된 파라미터 p 를 바탕으로 수렴율을 제시한다. 더 나아가, 커널 행렬의 최소고유값의 감소율과 관련된 상수 q 에 대해 $p \leq (q+1)/4$ 임을 증명함으로써, 최소 제곱법을 사용할 경우, 기저 함수 수가 증가함에 따라 RBFN 공간도 확장될 필요가 있음을 보였다.

시뮬레이션 실험을 기반으로 한 결과를 통해, 제안된 기대위험의 상한이 점진적 학습 알고리즘의 수렴율을 추정하는 데에 유용하게 활용될 수 있음을 보여주었다. 이때, 기저 함수의 개수가 훈련 샘플의 크기에 따라 증가하는 경우도 포함하였다. 따라서, 본 논문에서 제시한 점진적 특성은 비선형 회귀 모델에 대한 점진적 학습 알고리즘의 수렴을 위해 필요한 최소 훈련 샘플 수를 추정하는 데 효과적으로 적용될 수 있다.

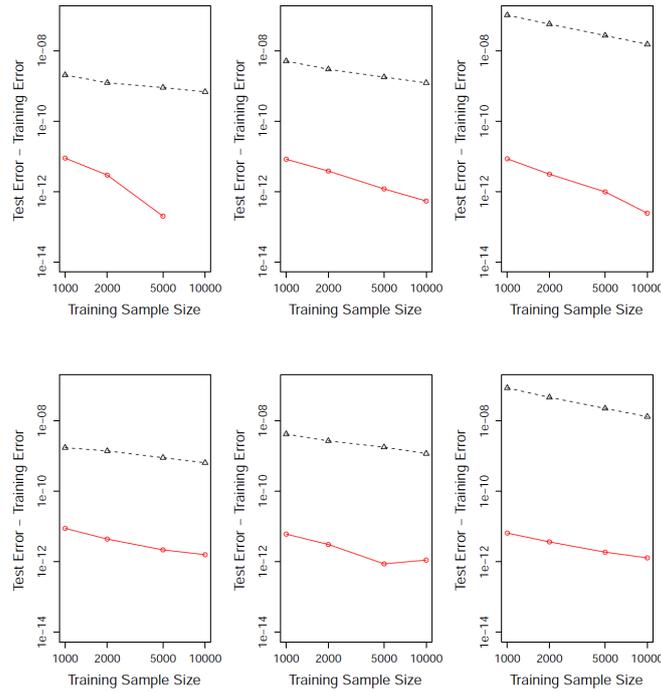
References

- [1] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edition, Prentice Hall, 1999.
- [2] T. Poggio, F. Girosi, *Networks for approximation and learning*, Proc. IEEE, **78** (1990), 1481–1497.
- [3] P. Niyogi, F. Girosi, *On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis functions*, Neural Comput., **8** (1996), 819-842.
- [4] V. N. Vapnik, *Statistical Learning Theory*, Wiley Interscience, 1998.
- [5] P. L. Bartlett, O. Bousquet, S. Mendelson, *Local Rademacher complexities*, Ann. Stat., **33** (4) (2005), 1497-1537.
- [6] P. L. Bartlett, S. Mendelson, *Rademacher and Gaussian complexities: Risk bounds and structural results*, Journal of Machine Learning Research, **3** (2002), 463-482.
- [7] O. Bousquet, *New approaches to statistical learning theory*, Ann. Inst. Stat. Math., **55** (2) (2003), 371-389.
- [8] V. Koltchinskii, *Rademacher penalties and structural risk minimization*, IEEE Trans. Inf. Theory, **47** (5) (2001), 1902-1914.
- [9] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2003.
- [10] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.
- [11] C. McDiarmid, *On the method of bounded differences*, Surveys in Combinatorics 1989, pp. 148-188, Cambridge University Press, 1989.
- [12] M. Ledoux, M. Talagrand, *Probability in Banach Spaces: Isometry and Processes*, Springer, 1991.

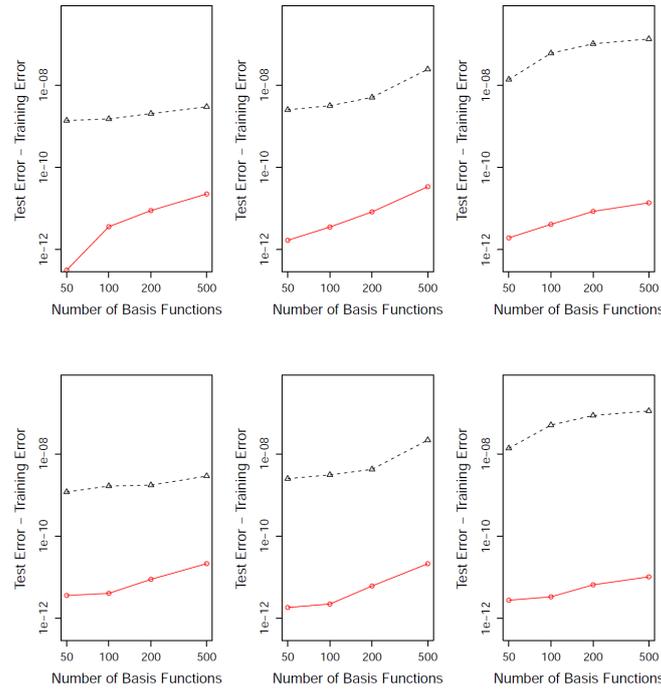
이남길

대한민국 강원특별자치도 춘천시 강원대학교 정보통계학전공 24341

E-mail: namgil.lee@kangwon.ac.kr



(a) $m = 200$



(b) $n = 1000$

그림 1. 테스트 오차와 훈련 오차 간의 차이(빨간 실선 및 원 마커)와 그에 대한 추정된 상한(검정 점선 및 삼각형 마커)을 나타낸 그래프. 가로축은 (a) 훈련 샘플 크기 또는 (b) 기저 함수의 개수를 나타낸다. (a)와 (b) 각각의 경우에 대해, 신호 대 잡음비(SNR)는 4(위 행) 또는 64(아래 행)로 설정되었고, 커널 폭 파라미터 σ 는 0.5(맨 왼쪽 열), 4(가운데 열), 16(맨 오른쪽 열)로 설정되었다.

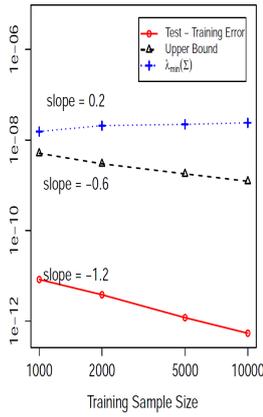
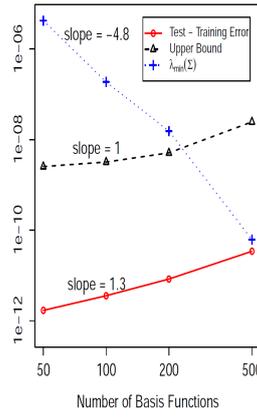
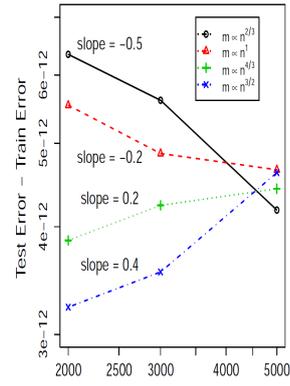
(a) $m = 200$ (b) $n = 1000$ (c) $m \propto n^a$

그림 2. 테스트 오차와 훈련 오차 간의 차이의 수렴율. (a) 기저 함수의 수 m 이 200으로 고정된 경우, (b) 훈련 샘플 수 n 이 1000으로 고정된 경우, (c) n 과 m 이 함께 증가하되, 고정된 a 값에 대해 $m \propto n^a$ 으로 증가하는 경우. 또한, (a)와 (b)에서는 최소 고유값 $\lambda_{\min}(\Sigma)$ 도 함께 나타내었다.

On the Convergence Rate of Incremental Learning Algorithms for Radial Basis Function Networks

Namgil Lee¹⁾

¹⁾Department of Information Statistics, Kangwon National University

Abstract

In this paper, we investigate a data-dependent upper bound on the expected risk of radial basis function networks (RBFNs). This risk bound is explicitly expressed as a function of both the network size and the training set size, making it applicable in scenarios where an RBFN incrementally recruits its basis functions. The proposed bound is thoroughly analyzed in terms of the asymptotic properties of incremental learning algorithms for RBFNs. Experiments on both simulated and real-world datasets are conducted to demonstrate the effectiveness of the proposed properties of RBFNs in estimating the convergence rate of incremental learning algorithms for nonlinear regression problems.

Key words: Incremental learning, Rademacher complexity, Radial basis function network

1) Dept. of Information Statistics, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do, 24341, Republic of Korea. E-mail: namgil.lee@kangwon.ac.kr