

# 적응적 성분분석 기법에 의한 RBF 신경망의 성능개선

조용현 윤중환<sup>U</sup>

대구효성가톨릭대학교 공과대학 컴퓨터정보통신공학부  
(yhcho, g9521008)@cuth.cataegu.ac.kr

## Performance Improvement of Radial Basis Function Neural Networks Using Adaptive Principal Component Analysis

Young-Hyun Cho Joong-Hwan Yun<sup>U</sup>  
School of Electronics and Information Engineering,  
Catholic University of Taegu-Hyosung

### 요 약

본 논문에서는 적응적 성분분석 기법을 이용하여 radial basis 함수 신경망의 학습시간과 분류성능을 개선한 새로운 기법을 제안하였다. 제안된 기법에서 적응적 성분분석 기법은 radial basis 함수 신경망의 은닉층 뉴런 개수와 중심값 설정을 위해 이용하였다. 제안된 기법의 radial basis 함수 신경망을 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, k-평균 군집화 알고리즘을 이용한 radial basis 함수 신경망과 비교할 때 학습시간과 시험 데이터의 분류에서 더욱 우수한 성능이 있음을 확인할 수 있었다.

### 1. 서 론

일반적으로 기존의 순차적 기법들이 가지는 문제점을 해결하기 위해서 입력과 출력간의 사상(mapping)을 가능하게 하는 신경망이 널리 연구되고 있다. 신경망을 이용하는 기법은 대규모의 병렬계산과 분산된 국부적 계산의 속성면에서는 통계적인 기법이나 계산 이론적인 방법의 문제들도 해결할 수 있다. 그러나 학습패턴과 학습알고리즘이 주어진다 고 모든 응용분야를 다 학습시킬 수는 없으며, 일반적으로 신경망은 망의 구조, 뉴런의 활성화함수, 그리고 학습규칙 등에 따라 여러 가지로 나누어진다<sup>1)~5)</sup>.

한편 입력층과 출력층사이에 은닉층을 가지는 다층전향신경망은 입력층 데이터 내에 포함된 어떤 비선형 연속함수를 근사화하거나 재구성할 수 있어 대단히 일반적이면서도 융통성 있는 속성을 가지고 있다. 이들 중 학습 알고리즘으로 역전파 알고리즘을 이용하는 다층신경망이 지금까지 가장 널리 이용되어 왔다<sup>1)~3)</sup>. 이는 MLP가 충분한 뉴런을 가지고 있을 때 어떤 임의의 함수를 근사화할 수 있다고 알려져 있으며, 지금까지 많은 시뮬레이터가 개발되어 있기 때문이다. 하지만 MLP는 프로세서의 유용한 모델을 얻기 위해서 많은 학습데이터와 시험데이터를 요구하며, 비파라미터 모델(non-parameter model)이라는 단점을 가지고 있다. 특히, 역전파 알고리즘은 학습파라미터의 설정에 따라 수렴속도와 견실성 중 하나 이상의 문제를 가지며, 전역최소점으로서의 수렴이 보장되어 있지 않다. 이러한 문제를 해결하기 위한 여러 방법들이 많은 연구자들에 의해서 연구되어 왔다<sup>1)~3)</sup>. 그 중에서 인간두뇌의 생물학적 뉴런의 기능을 모방한 국부적으로 동조된(locally-tuned) 처리장치를 은닉층의 뉴런으로 이용하는 RBF 신경망이 최근들어 많은 각광을 받고 있다<sup>6)~7)</sup>. 이는 RBF 신경망도 MLP와 같이 비선형의 다층전향신경망일 뿐만 아니라 일반적인 근사기(universal approximator)이기 때문이다. 특히 RBF 신경망은 MLP에 비해 훨씬 빠른 수렴속성과 전역최소로의 수렴이 보장된다고 알려져 있다. 그러나 RBF 신경망에서도 은닉층 뉴런의 개수와 함수의 중심(center), 평활요소(smooth factor), 그리고 연결가중치(synaptic weight) 등의 파라미터들이 설정되어야 한다. 이들 파라미터 중에서 학습 중에 은닉층 뉴런의 개수와 함수의 중심 및 평활요소는 고정된 값이며 연결가중치만이 가변되는 파라미터이다. 특히, 은닉층 뉴런의 개수와 함수의 중심값 결

정은 신경망의 구조를 결정하며, 신경망의 성능에도 많은 영향을 미친다. 이를 결정하기 위한 대부분의 기법들은 자율적인 방법으로 은닉층의 파라미터를 결정하는 다음 지도학습의 방법으로 출력 연결가중치를 결정한다. 여기서 자율적인 방법은 실제 신경망을 응용할 때 사용자나 군집화 알고리즘들 중 어느 하나에 의해서 조정되어야 될 파라미터들이 너무 많다는 단점을 가지고 있으며, 이와 함께 문제내의 국부적인 복잡성도 잘 반영하지 못하는 단점도 가지고 있다. 따라서 근본적으로 입력 데이터가 가지는 차원을 감소시킬 수 있는 새로운 기법을 도입함으로써 결정해야 될 은닉층의 파라미터들을 줄일 수 있을 것이다. 즉, 주어진 입력 데이터의 차원을 감소시켜 그에 해당하는 특징값을 RBF 신경망의 은닉층 파라미터로 사용함으로써 이들을 결정하는데 요구되는 계산량을 효과적으로 감소시킬 수 있을 것이다.

본 연구에서는 수치적인 데이터 집합의 차원을 해석하여 이를 감소시키는 것으로 널리 알려진 주요성분분석(Principal Component Analysis : PCA) 기법<sup>13)~15)</sup>을 이용하여 특징값을 추출한 다음, 추출된 특징값을 RBF 신경망의 은닉층 뉴런의 수와 중심값을 이용하는 새로운 조합형 기법을 제안한다. 제안된 기법의 RBF 신경망을 실제 암환자<sup>16)</sup> 중 200명의 환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 지금까지 널리 이용되고 있는 k-평균 군집화 알고리즘을 각각 RBF 신경망에 적용한 결과와 비교 고찰하였다.

## 2. 주요성분분석 기법에 의한 Radial Basis 함수의 중심설정

### 2.1 RBF 신경망

인식이나 분류 등의 문제들을 해결하기 위한 MLP와 RBF 신경망 사이에는 유사한 특성이 있는 반면에 은닉층의 수와 같은 구조나 각 은닉층 및 출력층 뉴런의 계산방법과 임·출력 사상에 대한 근사화 등은 아주 다른 속성을 가지고 있다. MLP 신경망에서 은닉층 뉴런과 출력층 뉴런의 활성화 함수는 단조함수이며, RBF에서 은닉층 뉴런은 radial basis 함수이고 출력층 뉴런은 단순히 가중된 합만을 출력한다. 여기서 radial basis 함수는 입력벡터  $x$ 와 중심  $c$  사이의 거리  $r = \|x - c\|$ 에 의존하는 다차원 함수이다. 비선형함

수의 근사화에 대한 가장 간단한 접근 중에 하나는 일반적으로 고정된 비선형 기초함수(basis function)의 선형결합에 의해 표현하는 것이다. 즉, 이 때 신경망의 학습목적은 계수집합인 연결가중치  $w_i$ 를 추정하는 것으로 다음과 같다. 즉,

$$F(x) = \sum_{i=1}^m w_i \varphi_i(x) \quad (1)$$

이다. radial basis 함수는 문제의 차원이 증가함에 따라 조정되는 파라미터의 확산문제를 해결할 수 있는 다차원 근사화를 위한 강력한 방법을 제공한다. 이용되고 있는 전형적인 radial basis 함수  $\varphi(x) = \mathcal{O}(\|x - c\|)$ 로는 가우스함수(Gaussian function,  $\mathcal{O}(r) = \exp(-r^2/\sigma^2)$ ), thin plate splines( $\mathcal{O}(r) = r^2 \log(r)$ ), 그리고 다중 2차 함수(multiquadratic function,  $\mathcal{O}(r) = (r^2 + \sigma^2)^{1/2}$ ) 등이 있다. 이들 식에서  $\sigma$ 는 거리의 계수(scaling) 파라미터인 평활요소이다. 일반적으로 radial basis 함수로는 중심  $c$ 에서 침투치를 가지며 그 중심으로부터 거리가 증가할수록 단조 감소하는 값을 가지는 가우스함수가 가장 널리 이용되고 있다.

한편, RBF 신경망의 radial basis 함수는 MLP의 단조 s-자형(monotonic sigmoid) 함수와 대조적으로 비단조적이다. 출력층의 뉴런은 MLP와 유사하게 입력들의 가중된 합을 수행한다. RBF 신경망은 출력층 뉴런의 출력을 얻기 위해서 radial basis 함수의 선형조합을 계산한다. 즉,

$$y^{\wedge}(t) = w_0 + \sum_{i=1}^m w_i \varphi_i(\|x(t) - c_i\|) \quad (2)$$

이다. 여기서  $t$ 는 반복수이고,  $x(k) = [x_1, x_2, \dots, x_n]^T$ 로 입력 데이터이며,  $c_i \in R^n (i=1, 2, \dots, m)$ 으로 함수의 중심이다.

radial basis 함수의 파라미터들은 3 개의 연속적인 단계로 결정될 수 있다. 첫 단계로,  $h$  개의 중심벡터  $c_i$ 가 결정되어야 한다. 이를 위해 가장 널리 이용되고 있는 방법으로는 k-평균 군집화 알고리즘이 있다. 이 알고리즘의 기본적인 생각은 입력 데이터의 밀도가 높아지면 중심의 밀도도 함께 높아진다는 것이다. 각 군집의 중심은 RBF 신경망에서  $h$  개의 은닉층 뉴런 중에 하나와 관계된다. 즉, 데이터는 그것과 가장 가까운 중심을 가진 군집에 할당되는 방법으로 분할된다. 그러나 이 알고리즘은 초기의 군집 중심을 어떻게 설정하는가에 따라 그 성능이 크게 변한다. 다음 단계로는 구해진 radial basis 함수의 중심벡터를 이용하여 평활요소  $\sigma_i$ 를 구하는 것으로 이는 P-최근접 이웃(P-nearest neighbour)로부터 결정될 수 있다. 즉,

$$\sigma_i = \frac{1}{p} \sum_{j=1}^p [\|c_i - c_j\|^2]^{1/2} \quad (3)$$

이다. 여기서  $c_j$ 는  $\sigma_i$ 의 P-최근접 이웃이다. 본 연구에서는  $\sigma_i$ 의 값에 따른 성능을 분석하기 위해서 임의의 상수값으로 설정하여 실험하였다. 마지막 단계는 설정된 radial basis 함수의 중심벡터  $c_i$ 와 평활요소  $\sigma_i$ 를 이용하여 식 (1)과 같은 선형회귀(linear regression)를 해결함으로써 연결가중치  $w_i$ 를 갱신하는 것이다. 한 예로 선형최소자승의 표준을 적용할 때 간단한 최소평균제곱(least mean square : LMS) 알고리즘을 얻을 수 있다. 즉, 연결가중치  $w_i$ 의 갱신을 이용하여 이루어진다.

$$\begin{aligned} \frac{dw_i}{dt} &= \eta_i e_i \varphi_i(x(t)) \\ &= \eta_i [y(t) - (w_0 + \sum_{i=1}^m w_i \mathcal{O}(\|x(t) - c_i\|))] \mathcal{O}(\|x(t) - c_i\|) \end{aligned} \quad (4)$$

여기서  $\eta_i$ 는 학습율로서 양수값이며,  $e_k = y(k) - y^{\wedge}(k)$ 이다. 따라서 근본적으로 입력 데이터가 가지는 차원을 감소시킬 수 있는 새로운 기법을 도입함으로써 결정해야 될 은닉층의 뉴런 수 및 그 중심값과 같은 파라미터들의 설정이 용이하게 될 것이다. 즉, 주어진 입력 데이터의 차원을 감소시켜 그에 해당하는 특징값을 RBF 신경망의 은닉층 파라미터로 사용함으로써 이들을 결정하는데 요구되는 계산량을 효과적으로 감소시킬 수 있

을 것이다.

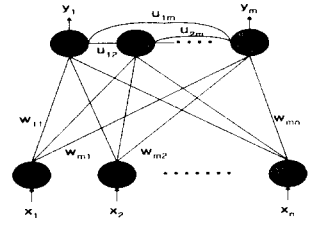
2.2 Foldiak의 학습알고리즘에 의한 적응적 특징추출

주어진 입력벡터에 대한  $m$  개의 주요특징을 추출하기 위해서는  $m$  개의 선형 처리장치로 구성되는 망이 요구되나 한 개의 선형 처리장치를 가진 망을 위한 학습알고리즘을 확장하여 여러개의 선형 처리장치를 가진 망을 학습시키는 것은 불가능하다. 따라서 주어진 입력벡터에 대한  $m$  개의 첫 번째 주요특징을 추출하기 위해서는  $m$  개의 선형 처리장치를 출력층으로 하는 단층 신경망이 구성되어야 하며, 이 망을 위한 별도의 학습알고리즘이 요구된다.

한편, 입력뉴런과 출력뉴런 사이의 연결 및 출력뉴런 상호간에도 연결을 가진 새로운 단층신경망이  $m$  개의 첫 번째 주요특징을 추출하기 위해 제안되었다<sup>[1]</sup>. 그림 1은  $n$  개의 입력뉴런과  $m$  개의 출력뉴런으로 구성된 입력과 출력뉴런간 및 출력뉴런 상호간의 측면연결을 가진 단층신경망의 구조이다. 그림에서 입력과 출력의 관계를 나타내면 다음과 같다. 즉,

$$y_i = \sum_{j=1}^n w_{ij} x_j + \sum_{k=1}^m u_{ik} y_k \quad (i=1, 2, \dots, m) \quad (5)$$

이다. 여기서  $w_{ij}$ 는 입력뉴런과 출력뉴런을 연결하는 연결가중치이고,  $u_{ik}$ 는 출력뉴런 상호간의 측면연결 가중치이다. 이때 출력뉴런 간의 상호연결을 보면 뉴런  $i$ 는 다만  $h < i$ 인 뉴런에만 연결된다.



(그림 1)  $m$  개의 첫 번째 주요특징 추출을 위한 측면연결의 단층신경망

따라서 신경망을 학습시켜 입력층 뉴런간 및 출력뉴런 상호간의 연결가중치를 각각 구함으로써 주어진  $n$  개의 입력벡터  $x$ 로부터  $m$  개의 주요특징벡터 즉 출력벡터  $y$ 를 구할 수 있다.

그림 1과 같은 단층신경망을 학습시키기 위해 먼저 입력층 뉴런간의 연결가중치  $w_{ij}$ 의 갱신규칙을 살펴보면 다음과 같다. 즉,

$$w_{ij}(t+1) = w_{ij}(t) + \eta_j [y_i(t) x_j(t) - w_{ij}(t) y_i(t)^2] \quad (i=1, 2, \dots, m, j=1, 2, \dots, n) \quad (6)$$

이다. 또한 출력뉴런 간의 측면 연결가중치  $u_{ik}$ 의 갱신규칙은 다음과 같다.

$$u_{ik}(t+1) = u_{ik}(t) + \rho y_i(t) y_k(t), \quad (i > k) \quad (7)$$

여기서  $\rho$ 는 식 (6)의  $\eta$ 와 일반적으로 동일한 값으로 설정되며, 둘 다 양의 값을 가지는 학습률로 실험에서는 고정된 값( $\rho = \eta = 0.1 / (\log_{10}(t))$ )으로 설정하였다. 따라서 식 (6)과 (7)을 이용하여 입력과 출력뉴런 간의 연결가중치 및 출력뉴런 상호간의 측면 연결가중치 갱신을 갱신시켜 식 (5)에 대입하면  $m$  개의 주요특징들을 추출할 수 있다.

결국 PCA는 높은 차원의 입력공간을 더 낮은 차원의 표현공간으로 사상시켜 입력데이터가 가지는 두드러진 특징들을 추출하는 기법이다. 이는 입력 데이터 내에 존재하는 주요특징들을 추출함으로써 입력의 개수를 감소시키는 데 이용될 수 있다. 따라서 RBF 신경망의 은닉층 뉴런 개수는 원래의 입력 데이터를 대상으로 PCA에 의해 얻어지는 감소된 주요특징의 개수로 설정하고, 은닉층 뉴런의 중심값은 얻어진 주요특징값으로 설정할 수 있을 것이다.

3. 시뮬레이션 결과 및 분석

제안된 PCA 기법을 이용한 RBF 신경망의 성능을 평가하기 위해서 3층 전향 신경망을 구성하였다. RBF 신경망에서의 은닉뉴런과 출력뉴런 사이의 초기 연결가중치와 PCA를 위한 단층신경망의 입력뉴런과 출력뉴런 및 출력 뉴런사이의 측면연결을 위한 초기 연결가중치는 각각 랜덤시드(random seed)를 이용하여 -1에서 +1 사이의 임의 값으로 설정하였다. 학습은 전체 반복회수가 20,000이상이거나 전체 오차값이 허용치 이하일 때, 또는 전체 오차함수 값의 변화가  $10^{-6}$  이하일 때 종료되도록 하였다.

실험에 이용한 암환자 분류문제의 데이터는 미국 위스콘신 대학병원에서 제공되는 "Wisconsin breast cancer databases"<sup>[8]</sup>를 대상으로 하였다. 본 실험에서는 임의의 환자 200명을 대상으로 100명은 학습을 위한 데이터로 나머지 100명은 시험을 위한 데이터로 이용하였다. 또한 실험에 이용된 데이터들에서 환자의 식별번호인 첫 번째 속성은 제거하였고, 모든 조건속성들은 정수 10으로 나누었으며, 분류속성은 2(초기)이면 0으로 4(말기)이면 1로 표현하였다. 학습이 종료된 후 출력뉴런의 값이 0.1 이하이면 0(초기)으로 하고 0.9 이상이면 1(말기)로 판정하였다. 다음의 그림 2는 5명의 환자를 대상으로 본 실험에서 사용되는 데이터를 나타내는 예이다.

1081791.	0.6, 0.2, 0.1, 0.1, 0.1, 0.1, 0.7, 0.1, 0.1, 0
1066594.	0.5, 0.4, 0.4, 0.9, 0.2, 1, 0.5, 0.6, 0.1, 1
1091232.	0.2, 0.5, 0.3, 0.3, 0.6, 0.7, 0.7, 0.5, 0.1, 1
1008300.	0.6, 0.6, 0.6, 0.9, 0.6, 0.7, 0.7, 0.8, 0.1, 0 (실함에서 제외)
1084610.	1, 0.4, 0.3, 0.1, 0.3, 0.3, 0.6, 0.5, 0.2, 1

(그림 2) 5명의 환자에 대한 속성데이터 예

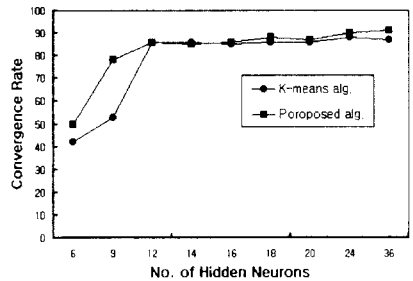
그림에서 ?는 그 속성 데이터가 알려져 있지 않아 실험에서는 학습 데이터로 이용되지 않은 경우이다.

그림 3은 평활요소  $\sigma$  ( $\sigma = 1.0$ )와 신경망의 초기 연결가중치(랜덤시드 = 0)를 고정하고 은닉층 뉴런수의 변화와 k-평균 군집화 및 제안된 알고리즘 각각에서의 수렴률과의 관계를 나타낸 것이다. 그림에서 보면, 제안된 알고리즘을 이용한 신경망의 수렴특성이 k-평균 군집화 알고리즘을 이용한 신경망에 비해 다소 우수한 수렴특성이 있음을 보여준다. 이는 주어진 문제에 대해서 제안된 알고리즘이 k-평균 군집화 알고리즘에 비하여 신경망의 은닉층 뉴런의 중심값을 더욱 잘 설정한 것으로 추측된다. 또한 두 알고리즘 모두의 경우에 은닉층 뉴런의 수가 증가할수록 그 수렴률도 함께 증가함을 알 수 있다. 그러나 은닉층 뉴런수의 증가는 계산량의 증가를 가져오며 결국 신경망을 학습시간도 오래 걸리게 된다. 따라서 은닉층 뉴런의 수는 적당하게 설정되어야 할 것이다. 본 실험에서는 은닉층 뉴런의 수를 18 개로 할 때에 학습시간과 그 수렴률 측면에서 가장 효과적임을 확인할 수 있었다.

표 1은 초기 연결가중치를 100번 변화시켜가면서 RBF 신경망을 학습시킨 결과, 학습에 소요된 반복회수  $N_k$ , CPU 시간  $t_r$ , 그리고 평균자승 오차값  $E_k$ 의 평균  $\bar{x}$ 와 표준편차  $\mu$ 를 각각 나타낸 것이다. 은닉층 뉴런의 개수와 평활요소는 두 알고리즘 모두에 대해 각각 18 개와 1.0으로 설정하였으며, 각 시도는 랜덤시드를 변화시켜 수행하였다. 여기서도  $P_c$ 은 학습 후에 100 개의 시험 데이터로 시험할 경우 올바르게 분류된 비율을 나타낸 것이다. 표에서 보는 것처럼, 제안된 알고리즘을 이용한 RBF 신경망이 k-평균 군집화 알고리즘을 이용한 신경망보다 학습시간과 데이터의 분류률면에서 더욱 우수한 성능이 있음을 알 수 있다. 특히 분류률에서는 제안된 알고리즘이 약간 우수한 성능을 보이나 학습시간면에서는 약 3배 정도 빠른을 알 수 있다. 표에서 보면 두 알고리즘을 이용하는 신경망 모두 기울기하강법을 이용하였다. 결국 이로 인해 학습률의 변화에 따라 학습성능도 달라진다. 그러나 가장 최적의 학습률을 설정할 수 있는 방법은 알려져 있지 않으며 본 실험에서는 여러 가지 학습률 값으로 실험 후에 학습률을  $0.1/(\log_{10}(N))$ 로 설정하였다.

#### 4. 결론

본 논문에서는 적응적 성분분석 기법을 이용하여 radial basis 함수



(그림 3) 은닉층 뉴런수의 변화에 따른 수렴률

<표 1> 100 개의 학습 데이터를 대상으로 100번 시도에 따른 실험 결과

	K means alg.			Proposed alg.		
	$N_k$	$t_r$	$E_k$	$N_k$	$t_r$	$E_k$
$\bar{x}$	6194.69	97.27	0.368058	1966.80	30.64	0.335882
$\mu$	786.75	26.11	0.045847	173.12	2.81	0.000326
$P_c$	86			88		

$\bar{x}$  : means,  $\mu$  : standard deviation,  $N_k$  : no. of iterations,  $t_r$  : CPU time(sec),  $P_c$  : convergence rate(%),  $E_k$  : mean square error

신경망의 학습시간과 분류성능을 개선한 새로운 기법을 제안하였다. 제안된 기법에서 적응적 성분분석 기법은 radial basis 함수 신경망의 은닉층 뉴런 개수와 중심값 설정을 위해 이용하였다. 이렇게 하면 PCA 기법이 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 RBF 신경망이 가지는 우수한 속성을 그대로 살릴 수 있다.

제안된 기법의 신경망을 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, k-평균 군집화 알고리즘을 이용한 RBF신경망에 의한 결과와 각각 비교할 때 학습시간과 시험 데이터의 분류률 면에서 더욱 우수한 성능이 있음을 확인할 수 있었다. 향후 제안된 알고리즘의 RBF 신경망을 좀 더 구체적인 문제와 통신이나 영상 인식 등과 같은 좀 더 다양한 분야에의 응용에 대한 연구가 계속 진행되어야 할 것이다.

#### 참고문헌

- [1] S. Haykin, 'Neural Networks: A Comprehensive Foundation', IEEE Press, New York, 1994
- [2] P. D. Wasserman, 'Advanced Methods in Neural Computing', Van Nostrand Reinhold, New York, 1993
- [3] A. Cichock and R. Unbehauen, 'Neural Networks for Optimization and Signal Processing', John Wiley & Sons., New York, 1993
- [4] P. Foldiak, "Adaptive Network for Optimal Linear Feature Extraction," International Joint Conference on Neural Networks, Washington D. C., vol 1, pp. 401-406, June 1989
- [5] S. Bannour, A. Mahmood, and A. Sadjadi, "Principal Component Extraction Using Recursive Least Squares Learning," IEEE Trans. on Neural Networks, vol. 6, no. 2, pp. 457-469, March 1995
- [6] K. I. Diamantaras and S. Y. Kung, 'Principal Component Neural Networks: Theory and Applications, Adaptive and Learning Systems for Signal Processing, Communications, and Control', John Wiley & Sons, Inc., 1996
- [7] J. Park and I. Sandberg, "Universal Approximation Using Radial Basis Function Networks," Neural Computatio, vol 3, no. 2, pp. 246-257, 1989
- [8] W. H. Wolberg, "Wisconsin Breast Cancer Database," Univ. of Wisconsin Hospital, July 1992(<http://mlftp.dice.unica.it/pub/mac...sin/breast-cancer-wisconsin.data>)