

신경망을 이용한 패턴인식에 대한 고찰

黃 昌 夏

대구효성가톨릭대학교 정보통계학과

On Pattern Recognition Using Neural Network

Hwang, Chang Ha

Department of Statistical Information, Catholic University of Taegu-Hyosung, Gyongsan
712-702, Korea

Abstract Machine intelligence will be a dominant technology in the 1990s. Pattern recognition techniques are often an important component of intelligent systems and are used for both data preprocessing and decision making. Broadly speaking, pattern recognition is the science that concerns the description or classification of measurements. Very many of the applications of neural networks are to classification, and so are within the field of pattern recognition. In this paper we explore how neural networks fit into the earlier framework of pattern recognition, and show that that framework can help us to make better use of neural networks for classification.

1. 패턴인식이란?

패턴인식은 정보의 복잡한 패턴들을 이용하여 불확실한 사실에 대해 의사결정하는 것이다. 그러므로, 패턴인식은 지능 정보 시스템의 구현 및 자료처리를 위해서 널리 사용된다. 패턴인식의 목적은 얼굴을 인식하는 문제, 지문 증권을 사야되

는지 또는 팔아야 되는지와 관련된 문제, 그리고 체스게임에서 다음에 어떤 수를 두어야 되는지와 같은 우리 인간에게 주어진 문제들을 해결하고자 하는 것이다. 그러나 다소 간단한 문제들은 하드웨어나 소프트웨어를 이용하여 해결될 수 있다. 패턴인식의 구체적인 예는 다음과 같다.

* 본 연구는 대구효성가톨릭대학교 97년도 특별연구비 지원에 의한 것임.

황단보도 위의 자전거와 보행자 구별하기
은하수 분류하기(나선형, 타원형 또는 기타
의 모양)

의료진단(medical diagnosis)

편지봉투의 우편번호 인식

pen-pad 컴퓨터의 필기체 문자, 숫자 인식

Tsetse 파리의 서식지 예측

Leptograpsus 계의 성(sex)과 종류의 분류

DNA 및 지문 식별

레이다 신호의 분류 및 분석

날아오는 미사일의 식별

음성인식

영상처리 및 분석

세계에 의한 옷의 손상 정도

사실, 위의 제 문제들은 패턴들의 군집(class)들이 이미 정의되어 있고 새로운 패턴이 주어지면 이를 어떤 군집으로 분류(classification)하는 문제이다. 이런 문제들을 분류 또는 감독패턴인식(supervised pattern recognition)이라 부른다. 문제해결을 위해서는 먼저 누군가 군집을 결정해야 하는데 이와 같이 패턴의 그룹 또는 군집을 찾는 것을 군집분석(cluster analysis) 또는 무감독패턴인식(unsupervised pattern recognition)이라 부른다. 군집분석 또한 패턴인식에 널리 이용되어져 왔다.

예제를 통해서 알 수 있듯이 패턴의 형태는 문제에 따라 다르다. 전문용어를 사용하여 표현하면 패턴은 특징(feature)들로 구성된다. 이런 특징들은 분류문제에 입력으로 사용되어지는 측정값이다. 예를 들어 Leptograpsus 계의 예제에서 특징은 계 껍질에 대한 네 종류의 길이와 몸의 깊이에 대한 측정값이다. 영상(image)를 다루는 패턴인식 시스템을 설계할 때 중요한 문제는 적당한 특징들을 선택하는 것이다. 올바른 특징의 선택은 특징 선택 이후에 행해지는 어떤 분석 보다 더 중요하다.

일반적으로 패턴인식의 접근방법은 통계적, 구문론적 또는 구조론적 패턴인식

및 신경망을 이용한 패턴인식 세가지로 분류된다. '통계적'이란 용어는 통계학자들이 사용한 것이 아니라 공학자들이 사용한 용어이다. 통계적 패턴인식은 통계적인 가정에 바탕을 두고 분류 업무를 수행한다. 구문론적 패턴인식은 한 패턴에 대한 의미있는 정보로 특징의 유무 또는 수치를 만들 수 있지만, 특징 간의 상호관련성 또는 상호 연결성 정보와 같은 중요한 구조적 정보도 만들 수 있다. 이러한 구조적 정보는 구조적 묘사 또는 분류를 쉽게 해준다.

신경망이란?

생물의 신경계가 정보를 어떻게 저장하고 다루는가에 관한 지식을 이용하려는 시도로 부터 신경적 계산방식이 부상하게 되었고, 신경망이라 불리는 인공신경계의 연구분야를 이끌어 내었다. 이러한 연구는 심리학, 신경과학, 인지학 및 시스템 이론과 같은 매우 다양한 분야의 연구를 혼합한 것으로 최근에 새롭게 상당한 주목을 받고 있다. 신경망은 비교적 새로운 계산방식이다. 전통적인 계산방식에 대한 신경망의 장점, 단점, 응용 그리고 그 관련성이 이해되지 않은 상태에 있으므로 신경망 영역에 대한 기대는 매우 높다. 신경망은 패턴연상(pattern association)에 의 응용에 특히 적합하다. 신경망은 비교적 간단하면서도 많은 노드들이 상호 연결된 망으로 구성되어 있으며 패턴인식을 구현하기 위한 좋은 도구로 사용될 수 있다.

패턴인식에 관한 모든 문제를 신경망으로 해결할 수 있다고는 생각되지 않는다. 패턴인식에서 신경망을 이용한 패턴인식을 새로운 개념으로 보아야 할 지, 아니면 단순히 통계적 또는 구문론적 접근방법의 구현을 위한 또 다른 기법으로 보아야 할 지는 아직 명확하지 않다. 확률적

신경망(probabilistic neural network) 또한 상당한 혼란을 야기 시킨다. 왜냐하면, 이것은 확률적이지 않기 때문이다. Kohonen의 learning vector quantization (LVQ) 방법은 통상적으로 신경망으로 간주되어지고, 가중치(weight)라 부르는 모수를 가지고 있다. 그러나, 이런 가중치들은 사실 연결강도(connection strength)를 나타내는 것이 아니다. 그리고 이방법은 오히려 고전적인 k-nearest neighbor method와 더 밀접한 관계가 있다.

패턴인식에서는 다중입력에 대해 다중출력을 제공하는 함수를 사용하기 때문에, 우리는 다층퍼셉트론이라 부르는 전방향 신경망에 대해서만 생각한다.

학 습

신경망을 이용한 패턴인식은 알고리즘적이지 못하지만 학습(learning) 또는 훈련(training)이 가능한 블랙박스 접근방법으로 볼 수 있다. 즉, 신경망 블랙박스는 각각의 학습표본들에 대해서 올바른 반응 또는 출력(예를 들면, 분류 등)을 내게끔 학습이 된다. 이러한 전략이 패턴인식 시스템 설계자의 관심을 끄는 이유는 사전 지식과 내부시스템의 동작에 대한 상세한 지식이 거의 필요 없기 때문이다. 더우기, 학습 후에 인공적으로 구현된 신경망의 내부구조는 학습표본들로부터 얻은 경험에 바탕을 두고 새로운 유사 패턴들이 주어질 경우에 그 패턴을 추정할 수 있도록 자기조직화(self-organization)하기 때문이다.

블랙박스 접근방법의 핵심적인 부분은 입력과 출력 사이의 관계를 밝혀내는 것이다. 따라서, 구문론적 패턴인식과 통계적 패턴인식에서와 마찬가지로 신경망을 이용한 패턴인식 접근방법은 학습표본들과 학습 알고리즘의 품질에 많은 영향을 받는다. 또한, 학습표본들과 학습 알고리

즘이 존재한다고 해서 주어진 신경망이 특정 응용분야에 잘 적용될 것이라는 보장은 없다. 그리고, 이러한 블랙박스 접근방법은 그 결과가 비록 정확할지라도 결과에 대한 해석이 어렵다는 문제점을 가지고 있다. 예를 들어, Tsetse 파리 서식에 대한 연구를 생각해 보자. 신경망의 결과가 분류트리의 결과 보다 더 정확하지만 관련 동물학자들은 분류트리를 선호한다. 왜냐하면, 분류트리는 해석이 용이하기 때문이다.

의료진단 예제

우리는 Cushing의 증후군을 가진 환자들이 대한 의료진단 자료를 이용하여 패턴인식 문제를 설명하고자 한다. Cushing의 증후군은 부신이 호르몬을 과분비하는 증상이다. 이 자료는 a, b, c로 세가지 형태의 증후군을 가지고 있다. 각각은 과분비의 원인을 나타내는 adenoma(선종), bilateral hyperplasia 및 carcinoma(암종)의 첫 알파벳이다.

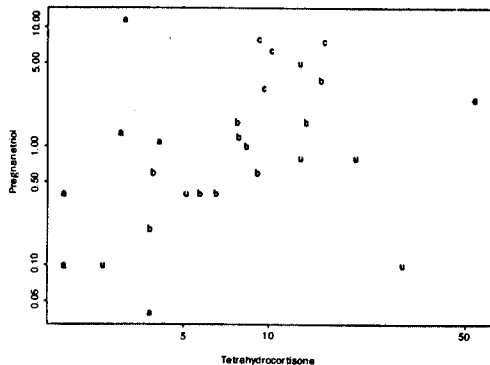


그림 1. Cushing의 증후군을 가진 환자들이 대한 두가지의 치료검진의 결과

관측치들은 두가지 종류의 부신피질 호르몬, tetrahydrocortisone과 pregnantriol의 비노 배설비율을 로그척도로 측정한다. 증후군의 형태를 모르는 환자중 한사람은 나중에 네번째 형태의 증후군으로

로 밝혀졌다. 한편 신드롬의 형태를 모르는 환자는 u 로 표현되었다. 그리고, 다른 환자는 잘못 측정되었다. 이러한 이상치(outlier)들은 패턴인식 문제에서 자주 일어난다. 그런데 이런 이상치들은 이상치로 보고되어야 하며, 어떤 한 군집으로 할당되어서도 않된다. 이 예제는 두개의 특징을 가진 매우 간단한 경우이지만 설명이 용이하기 때문에 선택되었다.

2. 통계적 패턴인식

이제 패턴인식을 위한 통계적 결정론의 기본 이론을 설명한다. 관측치들이 K 개의 군집 중 어느 하나에 속한다고 가정할 때 다음 단계의 의사결정은 각 관측치를 이들 K 개의 군집 중 어느 한 군집으로 분류하거나, 아니면 의문(doubt)시 되는 관측치들의 군집 D 또는 이상치들의 군집 O 로 분류하는 것이다. 의사결정 D 는 더 많은 관측치를 얻을 때 까지 의사결정을 연기한다는 것을 의미한다. 이때, 기계는 틀에 박힌 경우를 다루게 하고 전문지식을 필요로 하는 경우는 전문인이 다루도록 하는 것은 매우 유용한 전략이다. 예를 들면, 우편번호 판독기는 항상 어떤 사람들의 필체를 판독하는데 어려움이 있다. 그러나, 인간은 전체 문맥을 통해 숫자들을 이해할 수도 있다. 한편, 의사결정 O 는 관측치가 분명히 K 개의 군집 중 어떤 것에도 속하지 않는 것을 의미한다.

훈련표본(training sample)들의 집합 T 는 정확하게 분류된 n 개의 관측치들의 집합이다. 새로운 관측치들은 패턴 X 로 나타내고 우리는 그것의 군집을 결정하게 된다. 훈련표본들은 미래의 새로운 관측치들과 같은 모집단으로 부터 추출된 확률표본(random sample)으로 가정한다.

군집 j 에 속하는 관측치를 군집 k 로 잘못 분류할 때 생기는 비용을 C_{jk} 로 정

의한다. 서로 다른 오분류에 대해서는 매우 다른 오분류 비용을 할당할 수 있기 때문에 오분류 비용을 정의할 때는 신중을 기해야 한다. 의사결정론에서는 모든 것이 비용으로 축약될 수 있고, 합리적인 의사결정은 기대비용을 최소화하는 결정으로 가정한다.

의사결정론에서 기대비용을 최소화하는 규칙은 베이즈 규칙(Bayes rule)이다. 이 규칙은

$$\sum_j C_{jk} p(j|x)$$

를 최소화 하는 의사결정 k 가 된다. 사후 확률 $p(k|x)$ 는 이론상 중요하다. 이것은 특징벡터 x 를 갖는 관측치가 군집 k 에 속할 확률을 의미한다. 한편 사전확률은 π_k 로 나타낸다.

오분류 비용에 대한 일반적인 선택은 $C_{jj} = 0, C_{jk} = 1$ 이고 의사결정 D 에 대한 오분류 비용은 $d < 1$ 이다. 그러면, 베이즈 규칙은 다음과 같다.

$$c(x) = \begin{cases} k, & p(k|x) = \max_j p(j|x) > 1-d \\ D, & \max_j p(j|x) \leq 1-d \end{cases}$$

만일 두 군집이 모두 최대의 사후확률을 가지면, 기대비용이 같기 때문에 둘 중 어떤 군집을 택해도 무방하다.

사후확률을 알고 있다면 문제는 간단하다. 그러나 대부분의 경우에 이들은 미지이다. 따라서, 훈련표본들을 이용해서 이들을 유도해야 한다. 다음과 같이 사후확률을 유도하는 두 방법이 있다.

Sampling Paradigm

군집 k 로 부터 추출된 관측치의 확률 밀도함수를 $p_k(x)$ 로 표기하자. 그러면, 베이즈 공식에 의해 사후확률은

$$p(k|x) = \frac{p_k(x)\pi_k}{\sum_j p_j \pi_j}$$

가 된다. 따라서, 베イズ 규칙은 $p_k(x)\pi_k$ 를 최대화 하는 군집을 선택하게 된다. 이 값들은 대체로 모수를 포함하게 되며 훈련표본들로 부터 추정된다. 예를 들면, 정규분포의 경우에는 모수는 평균벡터와 공분산행렬이 된다. 이 방법은 통계학과 패턴인식에서 널리 사용되는 방법이며 대표적인 예는 선형판별분석이다. 그런데, 이 방법은 과대적합(over-fitting)을 하는 경향이 있다. 즉, 훈련표본들에 대해서는 좋은 판별결과를 주지만 새로운 관측치들인 검정표본들에 대해서는 좋은 결과를 제공하지 못한다. 한편, 특징벡터와 군집의 결합분포는 다음과 같이 주어진다.

$$p(x, k) = p(k|x)p(x) = p_k(x) \pi_k$$

여기서, $p(x) = \sum_k \pi_k p_k(x)$ 이다.

Diagnostic Paradigm

이 방법은 사후확률 $p(k|x)$ 를 훈련표본들로 부터 직접 구하는 것인데 가장 간단한 예제는 k-nearest neighbours 방법이다. k-nearest neighbours 방법은, 새로운 관측치 x 가 주어지면 훈련표본들 중에서 이 관측치에 가장 가까운 k 의 관측치를 찾고 이들이 속하는 군집들의 분포를 이용하여 $p(k|x)$ 를 구하는 방법이다. Diagnostic paradigm에서 $p(x)$ 는 분류를 위해 필요치 않기 때문에 이것을 추정하지 않는다. 그러나 이것은 이상치들을 찾는데 이용되어 진다. 이 방법에서는 통상적으로 이상치들을 무시한다.

통계학에서 diagnostic paradigm의 주 방법은 로지스틱 판별분석법이다. 먼저, 두개의 군집의 경우를 생각해 보자. 그러면, 군집 2에 대한 로그-승산은

$$\text{logit } p(2|x) = \log \frac{p(2|x)}{p(1|x)}$$

가 된다. 로짓함수는 $\text{logit } p = p/(1-p)$ 이

고, 역함수는 로지스틱함수 $\ell(x) = e^x / (1+e^x)$ 이다. 이제, 로그-승산은 다음과 같이 특징들의 선형함수라고 가정하자.

$$\begin{aligned} \text{logit } p(2|x) &= \alpha + \beta^T x, \\ p(2|x) &= \ell(\alpha + \beta^T x), p(1|x) = 1 - p(2|x) \\ &\dots\dots\dots(2.1) \end{aligned}$$

이것은 은닉층이 없고 한개의 출력노드를 갖는 전방향 신경망이다. 이것은 특징들의 선형함수와 로지스틱 출력 두 단계로 나누어 생각할 수 있다. 더욱 더 유동적(flexible)인 모형을 만들기 위해서 선형함수를 비선형함수 $f(x) = \text{logit } p(2|x)$ 로 바꿀 수 있으며, 한편 은닉층을 갖는 신경망을 이용해서 이 함수를 근사시킬 수 있다. 비록 이것이 전방향 신경망이지만 출력층에서 로지스틱함수를 사용하게 된 것은 우연의 일치이다. 전방향 신경망을 유동적인 비선형 함수로 간주하고 출력을 이용해서 무엇을 할 수 있는지를 생각하는 것은 타당하다.

이제 우편번호인식 문제와 같이 두개 이상의 군집을 갖는 경우를 생각해 보자. 자연스러운 일반화는 로그선형모형 또는 다중로지스틱 회귀모형이라 부르는

$$f_k(x) = \text{logit } p(k|x) = \alpha_k + \beta_k^T x$$

이다. 이때 사후확률은

$$p(k|x) = \frac{\exp f_k(x)}{\sum_j \exp f_j(x)} \dots\dots\dots(2.2)$$

로 주어진다. 여기서도 선형함수를 비선형함수로 바꿀수 있다. 그러면, 이것은 은닉층과 선형 출력 노드들을 가지는 신경망이 된다. 신경망 분야에서 이 방법은 softmax로 알려져 있다.

K개의 군집들 중에 어느 하나로 분류하기를 원하고 한 관측치가 어느 한 군집에 반드시 속해야 한다면 이 방법은 적당하다. 의료진단에서는 다음과 같은 문제

가 가끔 발생한다. K 종류의 질병이 있는데 환자는 병이 전혀 없거나, 하나 또는 그 이상의 많은 질병을 가질 수 있다. 이 경우 각 질병에 대해 하나의 로지스틱 모형을 설정하는 것이 자연스러운 방법이 된다. 이것은 각 군집에 대해 서로 다른 로지스틱 출력을 갖는 신경망이며 softmax 보다 더 널리 사용되고 있다. 두 방법이 큰 차이는 나지 않지만 한개 이상의 군집들이 상당한 사후확률을 가질 때는 그 차이가 유의할 수 있다. 만일 비용 C_k 가 군집내의 구조를 반영하도록 선택된다면 softmax는 상당히 정확하다.

3. 신경망 적합

우리는 지금까지 신경망이 diagnostic paradigm 방법하에서 사후확률 $p(k|x)$ 를 추정하는 로지스틱 판별법을 일반화 시킨 방법이라는 사실을 알았다. 신경망에서 모수, 즉 가중치 w 를 결정하는 것은 매우 중요한 문제이다. 가중치 벡터 w 를 사용한 신경망에 의해 적합된 함수를 표현하면 $f(x;w)=(f_k(x;w))$ 이다. 전술한 바와 같이 패턴인식에서는 $f_k(x;w)$ 는 사후확률이다. 여기서, x 는 특징벡터의 관측치이고 k 는 군집을 나타내는 변수이다.

가중치 벡터 w 를 결정하는 가장 일반적인 방법은 로그우도함수 $\sum_T \log p(k|x;w)$ 를 최대화 하는 w 를 선택하는 최우추정법이다. 여기서, T 는 훈련표본들의 집합이다. 한편, 각항 $\log p(k|x;w)$ 는 음수이기 때문에 이 방법은 $E(w)=\sum_T -\log p(k|x;w)$ 를 최소화 하는 방법과 같다. 신경망 분야에서 서로 다른 관점을 가진 여러 사람들에 의해 최우추정법이 여러번 제안되었다. 그러나 로지스틱 판별법에서는 이 방법이 이미 수십년 동안 사용되어져 왔다.

최우추정법이 이론적 장점을 가지고는

있지만 가장 널리 사용되는 방법은 아니다. 최소제곱법을 사용해서 사후확률을 적합하는 방법이 가정 널리 사용되고 있다. 두개의 군집에 대한 예를 생각하면 이 방법은 $E(w)=\sum_T [y-p(2|x;w)]^2$ 를 최소화 하는 것이다. 여기서, 군집 2가 참 군집이면 $y=1$ 이고 그렇지 않으면 $y=0$ 이다. 이 방법은 소프트웨어가 널리 이용 가능하다는 점외에는 다른 장점이 없다.

좀 복잡하기는 하지만 더 좋은 방법들이 있다. 지금까지 사용해 온 방법은 훈련표본들의 집합 T 에 사후확률들이 잘 적합되도록 신경망의 가중치 벡터 w 를 수정하는 것이다. 이 방법을 사용할 때 우리는 $p(k|x;w)$ 가 마치 참 사후확률인 것 처럼 생각하게 된다. 더 많은 은닉 노드들을 갖는 신경망을 사용하면 $f(x;w)$ 는 T 에 더 잘 적합된다. 그러나 참 사후확률들에 반드시 더 잘 적합되는 것은 아니다. 그래서 우리는 은닉 노드의 수가 증가함에 따라 새로운 관측치들에 대한 성능(performance)이 처음에는 향상되다가 나중에는 점점 나빠지는 것을 알게 된다. 이런 현상은 predictive approach를 사용하면 극복된다.

T 를 과대적합하는 것을 피할 수 있는 간단한 방법은 정칙화(regularization)이다. 이것은 굴곡이 심한 함수 $f(x;w)$ 에 벌칙을 가하는 방법으로서 주로 $E(w)+\lambda C$ 를 최소화 하는 방법이다. 이때 C 는 가중치들의 제곱합이다. 특히 $E(w)+\lambda C$ 에 최급강하법(steepest descent)을 적용하여 가중치 벡터 w 를 수정하는 방법을 가중치 감소(weight decay)법이라 한다. 이 방법의 효과를 알아 보기 위해 인공자료를 만들어 적용하였다. 이때 $\lambda=0.005$ 이다. 이 예제에는 군집이 두개이고 각 군집으로부터 125개의 관측치를 추출했다. 자세한 내용을 위해 Ripley(1994)를 참고하시오. 이것은 인공적인 예제이므로 관련된 모든 확률들을 알 수 있다. 따라서, 베이지 법

최은 8.0%의 오분류율을 가짐을 알 수 있다. 그림 2는 로지스틱 판별법이 부적당하고 신경망은 베이스 법칙과 아주 비슷한 결과를 가짐을 보여준다.

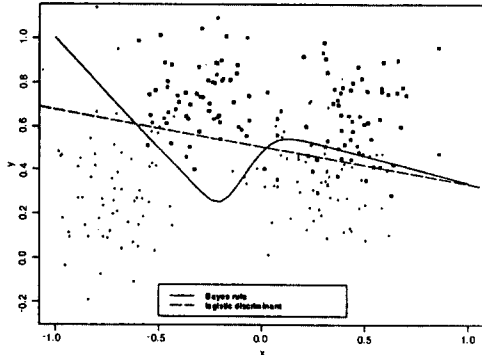


그림 2. 인공자료에 대한 예제

베이스 규칙	8.0%
로지스틱 판별법	11.4%
3개의 은닉 노드를 사용한 신경망	11.1%
3개의 은닉 노드와 가중치감소법을 사용한 신경망	9.4%
6개의 은닉 노드를 사용한 신경망	9.5%

극소점 문제

자주 간과되는 한가지 어려운 문제는 신경망 적합의 알고리즘들이 적합기준(fit criterion)의 최소점이 아닌 극소점 만을 찾는다라는 것이다. 가중치감소법이 사용될 때에도 일반적으로 여러개의 극소점들이 존재할 수 있다. 비록 많은 극소점들을 찾을 수 있다고 해도 이들 중 어떤 것을 선택해야되는지 분명치 않다. 왜냐하면 훈련표본들을 잘 적합한다고 해도 검정표본들을 잘 할당하는 것은 아니기 때문이다. Ripley(1996)는 사후확률들을 평균하는 것이 최적의 방법이라 제안한다. 이 방법은 $p(k|x) = \sum_m \alpha_m p(k|x; w_m)$ 을 사용하는 것인데 최근에 널리 이용되고 있다. 여기서, w_m 은 서로 다른 극소점에 대응되는 가중치 벡터이고 $\alpha_m \geq 0, \sum \alpha_m = 1$ 이다. 각 극소점에 대해 같은 α_m 을 사용하

는 것이 더 좋을 때가 자주 있다. 평균을 이용하는 이 방법은 신경망을 학습시키거나 검정표본들을 예측할 때 더 많은 CPU 시간을 필요로 한다. 그러나, 만일 다중 CPU가 이용가능하면 병렬처리를 할 수 있기 때문에 별 문제가 되지 않는다.

로버스트

적합 기준으로 최소제곱법을 주로 사용하기 때문에 신경망 적합의 알고리즘들은 과대오차(gross error) 또는 이상치(outlier)에 아주 민감한 경향이 있다. 따라서, Chen & Jain(1994)은 자료가 과대오차 또는 이상치에 의해 오염이 되었고 훈련표본들에 대해 과대적합하는 것을 피하기 위해서는 로버스트 통계학에서 많이 사용하는 로버스트 최소제곱법을 사용하기를 제안하였다. Liu(1994)는 회귀오차에 대한 분포로 t -분포를 사용하였다.

망복잡도(network complexity)

적합된 신경망의 복잡도(complexity)를 선택해야 한다. 이것은 은닉노드의 수와 가중치감소법의 λ 에 의해 통제된다. 사실 가중치감소법을 사용하면 은닉노드의 수는 그렇게 중요하지 않다. 사실 λ 는 베이즈 방법에 의해 결정될 수 있다. 분류를 위해서는 자주 $\lambda=0.01$ 를 선택한다. 망의 복잡도를 결정하는 문제는 최근의 연구분야이다.

복잡도를 결정하는 다른 방법은 새로운 표본들에 대한 성능(performance)을 추정하는 것에 기초를 두고있다. 이제 타당성 검정을 위한 표본들의 집합 v 를 가지고 있다고 가정한다. 그리고 여러개의 서로 다른 망들을 적합시키고 v 에 대해 그들의 성능을 평가하여 가장 좋은 것을 선택한다. 마지막으로 검정표본들의 집합 S 에 대해 선택된 망의 성능을 측정한다.

실제로 T, v, S 를 모두 같은 것은 쉽

지 않다. 그래서 교차타당성(cross-validation)을 사용한다. 이 방법을 간단히 설명하면, 먼저 T 를 V 개의 부분집합으로 나누어 각 부분집합을 차례로 v 로 두고 나머지 표본들을 사용하여 신경망을 적합시킨다. 그리고 v 에 속하는 표본들을 예측한다. 이런 일을 V 번 반복하면 모든 관측치들을 예측할 수 있으며 신경망의 성능을 평가할 수 있다. V 를 결정하는 두가지 통상적인 방법이 있다. 한 방법은 한번에 하나의 관측치를 제거하여 신경망을 n 번 훈련시킨다. 이 방법은 매우 느리고 이론적으로도 바람직하지 않다. 다른 한 방법은 $V=5$ 또는 10을 택하는 것이다.

4. 유리 자료

이 자료는 German이 214개의 유리파편에 대해 수집한 자료이다. 각 파편에 대한 자료는 굴절률과 Na, Mg, Al, Si, K, Ca, Ba 및 Fe와 같은 산화물들의 화학적 함성비를 나타내는 값들이다. 원래 파편들은 7종류로 분류되었는데 그중 한 종류는 자료에 없다. 6종류는 창문용 플로트유리(플로트법으로 제조된 판유리) (70), 창문용 비플로트유리(76), 자동차 창유리(7), 그릇(13), 식탁용 기구(9) 그리고 자동차 전조등(29)이다. 함성비의 합은 거의 100%에 가깝고 나머지는 모래이다.

이 예제에 대해 표본들을 10개의 부분으로 나누어 교차타당성법을 사용하였다. 선형 로지스틱판별법은 약 36.0%의 오분류율 가지는 반면에 가장 가까운 군집을 선택하는 방법은 약 23.4%의 오분류율을 가진다. 적당한 신경망을 선택하기 위해 10-중 교차타당성 집합을 설정하고 구해진 모든 극소점들에 대해 예측값을 평균하였다. 교차타당성법에 의한 오분류율은 다음 표에 주어진다.

λ	은닉노드의 수		
	2	4	8
0.0001	30.8	23.8	27.1
0.001	30.4	26.2	26.2
0.01	31.8	29.9	29.9

여기서, λ 는 가중치 감소법의 모수이다. 위의 결과는 교차타당성 검정표본들의 집합에 대해서 오분류율 23.8%를 가지는, 즉 4개의 은닉노드와 모수 $\lambda=0.0001$ 을 사용하는 신경망을 사용하기를 제안한다.

5. 결 론

오늘날 전방향 신경망은 패턴인식 분야에서 매우 널리 이용되고 있는 반면에 통계학 분야에서 개발된 비선형 판별법은 그렇게 많이 알려지지 않았다. 통계학, 신경망 분야 그리고 패턴인식 분야에서 개발된 많은 분류법들에 대해 장단점을 쉽게 비교할 수 있다. 한편 신경망은 유동적 비선형 회귀분석 방법에 속한다고 볼 수 있다. 모수추정, 분류기 평가 및 알고리즘 개발과 같은 많은 흥미있는 문제들이 남아 있다.

참 고 문 헌

1. Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.
2. Chen, D. S. and Jain, R. C. (1994), *A Robust Back Propagation Learning Algorithm for Function Approximation*, IEEE Transactions on Neural Networks, 5, 467-479.
3. Cheng, B. and Titterington, D. M. (1994), *Neural Network: A Review from a Statistical Perspective*. Statis-

- tical Science Vol. 9, No. 1, 2-54.
4. Cherkassky, V., Friedman, J. H. and Wechsler, H. (1994), *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, Springer-Verlag.
 5. Hertz, J., Krogh, A., and Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*, Addison-Wesley Publishing Company.
 6. Honik, K., Stinchcombe, M. and White, H. (1989), *Multilayer Feed-forward Networks are Universal Approximators*, Neural Networks, 2, 359-366.
 7. Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D. (1989), *Backpropagation Applied to Handwritten Zip Code Recognition*, Neural Computation 1, 541-551.
 8. Liu, Y. (1994), *Robust Parameter Estimation and Model Selection for Neural Network Regression*, NIPS6, 192-199.
 9. Ripley, B. D. (1994), *Neural Networks and Related Methods for Classification*, J. R. Statist. Soc. B 56, No. 3, 409-456.
 10. Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press.
 11. Schalkoff, R. (1992), *Pattern Recognition-Statistical, Structural and Neural Approaches*, John Wiley & Sons, Inc.
 12. Sethi, I. K. and Jain, A. K. (1991), *Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections*, North-Holland.
 13. Smith, M. (1993), *Neural Networks for Statistical Modeling*, Van Nostrand Reinhold.
 14. White, H. (1989), *Learning in Artificial Neural Networks: A statistical perspective*, Neural Computation, 1, 425-464.