

## 다집단 분류 인공신경망 모형의 아키텍처 튜닝\*

정철우\*\* · 민재형\*\*\*†

### Tuning the Architecture of Neural Networks for Multi-Class Classification

Chulwoo Jeong\*\* · Jae H. Min\*\*\*

#### ■ Abstract ■

The purpose of this study is to claim the validity of tuning the architecture of neural network models for multi-class classification. A neural network model for multi-class classification is basically constructed by building a series of neural network models for binary classification. Building a neural network model, we are required to set the values of parameters such as number of hidden nodes and weight decay parameter in advance, which draws special attention as the performance of the model can be quite different by the values of the parameters. For better performance of the model, it is absolutely necessary to have a prior process of tuning the parameters every time the neural network model is built. Nonetheless, previous studies have not mentioned the necessity of the tuning process or proved its validity. In this study, we claim that we should tune the parameters every time we build the neural network model for multi-class classification. Through empirical analysis using wine data, we show that the performance of the model with the tuned parameters is superior to those of untuned models.

Keyword : Multi-class Classification, Neural Networks, Model Architecture, Tuning Method

논문접수일 : 2012년 10월 09일 논문게재확정일 : 2012년 11월 09일

논문수정일(1차 : 2012년 10월 23일, 2차 : 2012년 11월 05일)

\* 이 논문은 2010년도 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(NRF-2010-327-B00154).

\*\* 한국국방연구원

\*\*\* 서강대학교 경영전문대학원

† 교신저자, jaemin@sogang.ac.kr

## 1. 서 론

인공신경망이 처음 개발되었을 당시에는 생물학적 신경의 작용 원리를 설명하기 위한 목적을 가지고 있었지만, 그 이후 지금까지 다양한 분야에서 가장 뛰어난 성과를 보이는 분류 기법 중의 하나로 자리매김하고 있다[4]. 특히, 다집단 분류 기법으로서의 인공신경망은 분류정확도에 있어 기존의 기법들에 비해 높은 성과를 보이는 장점 외에도[16, 19], 전통적인 통계기법들이 가지는 정규성, 선형성 등의 가정을 필요로 하지 않기 때문에 적용 범위가 넓다는 장점을 가진다[6]. 비록, 인공신경망 모형에 대한 몇 가지 한계점, 예를 들어, 모형이 산출하는 결과에 대한 설명력이 부족하여 예측결과의 원인을 설명하기 어렵다는 점, 소량의 데이터를 가지고는 모형의 학습이 어렵다는 점, 기존 데이터에 과적합하여 새로운 데이터에 대한 예측력이 낮아지는 문제 등이 지적되었지만[2, 15], 이러한 한계점을 극복하기 위한 연구에 힘입어 여전히 학계 및 실무에서 많이 활용되고 있는 기법이다.

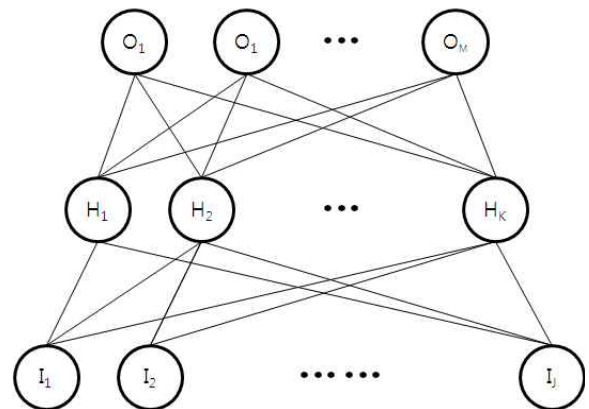
본 연구는 인공신경망 모형의 한계점을 극복하기 위한 연구의 일환으로, 다집단 분류를 위한 인공신경망 모형 구축 시 모형의 몇 가지 요소를 튜닝함으로써 모형의 복잡성을 통제하는 기법을 제시하고자 한다. 이를 통해, 인공신경망 모형이 가지는 과적합 문제를 방지하고 그 성과를 극대화할 수 있는 방안을 제시한다는 점에서 본 연구의 기여점을 찾을 수 있다.

본 연구는 다음과 같이 구성된다. 제 2장에서는 인공신경망의 일반 모형과 분리 모형의 개념, 인공신경망 아키텍처 튜닝의 의미를 소개한다. 제 3장에서는 본 연구에서 제시하는 아키텍처 튜닝 기법의 실증분석을 위한 실험 디자인을 기술한다. 제 4장에서는 실증분석의 과정과 그 결과를 보여준다. 그리고 마지막 제 5장에서는 본 연구를 요약하고 본 연구의 한계점 및 향후 연구방향을 제시한다.

## 2. 이론적 배경

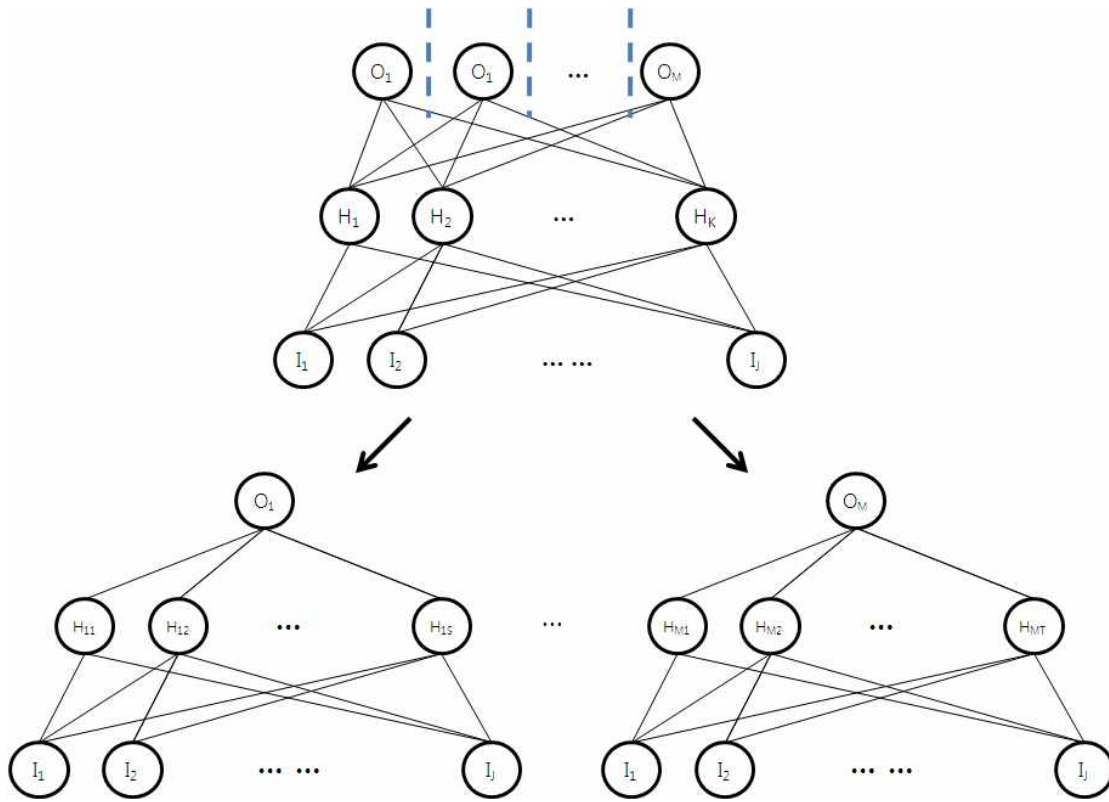
### 2.1 인공신경망의 일반 모형과 분리 모형

인공신경망 모형을 이용한 다집단 분류 방법에는 일반 모형을 이용한 방식과 분리 모형을 이용한 방식이 보고되고 있다[1, 18]. 전자의 방식은 [그림 1]에서 보는 바와 같이 입력층과 출력층, 그리고 출력층과 은닉층 사이에 연결된 가중치가 단 한 번 설정되는 방식이다. 이 경우에는 출력층의 모든 노드에 대하여 전체적인 오차의 합을 최소화하는 방향으로 가중치가 설정되기 때문에 이로 인한 정보의 손실이 발생할 수 있다.



[그림 1] 인공신경망 일반 모형

이에 반해 [그림 2]는 출력층의 노드가 여러 개인 일반 모형이 출력층 노드마다 분리되어 1개의 출력 노드만 가지는 여러 개의 모형으로 이루어진 분리 모형을 나타내고 있다. 이러한 분리 모형을 이용한 분류 방식은 근본적으로 훈련용 데이터를 이용하여 이집단 분류 인공신경망 모형을 집단의 수만큼 구축한 후 검증용 데이터에 적용함으로써 집단을 분류하는 방식이라고 할 수 있다. 다만, 모형 구축 및 적용에 있어 다집단 데이터(multi-classified data)를 어떠한 방식으로 이집단 데이터(binary classified data)로 보고 분류해 나가느냐에 따라 OPP 방법(ordinal pairwise partitioning method), 이진분류 방법(binary classi-



[그림 2] 인공신경망 분리 모형

fication method), 단순분류 방법(simple classification method)과 같은 방법으로 나누어진다.

OPP 방법은 집단의 서열에 따라 가장 높은 등급의 집단 또는 가장 낮은 등급의 집단부터 반대 등급의 집단까지 차례대로 분류해나가는 방식이다. 이진분류 방법은 중간 등급을 기준으로 분류하는 방식으로, 예를 들어, 집단이 1등급부터 5등급까지 있을 때, 1, 2, 3등급과 4, 5등급 사이, 또는 1, 2등급과 3, 4, 5등급 사이를 분류하는 것과 같이 중간 등급에서부터 나누어 모형을 구축하고 적용하는 방법이다. 단순분류 방법은 자료의 서열적 특성을 이용하지 않고 각 집단을 1로 하고 나머지 집단을 0으로 하는 분리 모형을 집단의 수만큼 구축하여 훈련용 데이터에 적용시키고, 그 결과 각 개체에 대한 산출변수 값이 가장 큰 산출변수 값에 대응하는 집단으로 분류하는 방식이다. 김다윗 외[1]는 인공신경망의 일반 모형과 함께 여러 가지 분리 모형의 성과를 비교하였는데, 검증용 데이터에 대하여 전

향적 OPP 방법(70.2%)과 단순분류 방법(69.6%)이 가장 높은 성과를 보이는 것으로 보고하였다. 그런데 세 가지 용도의 데이터(훈련용, 테스트용, 검증용)에 대한 분류정확도를 보면 단순분류 방법(72.8%, 67.7%, 69.6%)이 전향적 OPP 방법(86.7%, 83.2%, 70.2%)보다 변동성이 낮아 일관성이 더 높다고 할 수 있다. 이러한 선행 연구결과에 따라 본 연구에서는 인공신경망 분리 모형의 구축 및 적용을 위해 단순분류 방법을 이용한다.

인공신경망의 분리 모형이 일반 모형에 비해 더 높은 성과를 보이는 이유는 다음과 같이 요약할 수 있다. 첫째, 분리된 인공신경망 모형은 각 출력 노드 별로 입력층과 은닉층 사이에 유일한 가중치를 가지게 되므로 일반 모형에서 공유되는 가중치에 의한 정보의 손실을 줄여 학습을 보다 정밀하게 할 수 있다. 둘째, 인공신경망을 분리시키면 출력 노드에서 발생하는 출력값과 목표값 사이의 오차를 줄여주어 인공신경망 모형의 판별력을 향상시키게 된다.

## 2.2 복잡성 통제를 위한 인공신경망 모형의 아키텍처 튜닝

인공신경망 모형을 수리적으로 표현하면 식 (1)과 같다.

$$y_{mi} = h\left(\sum_{k=0}^K w_{km}^{(2)} \cdot g\left(\sum_{j=0}^J w_{jk}^{(1)} x_{ij}\right)\right) \quad (1)$$

여기서,  $y_{mi}$ 은  $i$ 번째 개체에 대한  $m$ 번째 출력값이다( $m = 1, 2, \dots, M$ ;  $i = 1, 2, \dots, I$ ).  $x_{ij}$ 은  $i$ 번째 개체에 대한  $j$ 번째 변수값이다.  $w_{jk}^{(1)}$ 은  $j$ 번째 입력변수에서  $k$ 번째 은닉노드로 가는 경로의 가중치이고,  $w_{km}^{(2)}$ 은  $k$ 번째 은닉노드에서  $m$ 번째 출력노드로 가는 경로의 가중치이다.  $M, I, J, K$ 는 각각 출력노드 수, 개체 수, 입력변수 수, 은닉노드 수를 가리킨다. 함수  $g(\cdot)$ 와  $h(\cdot)$ 는 각각 은닉노드와 출력노드에서의 변환함수를 가리킨다.

은닉노드와 출력노드의 변환함수로는 선형 함수, sigmoid 함수, hyperbolic tangent 함수 등이 이용될 수 있다[5]. 본 연구에서는 은닉노드와 출력노드의 변환함수로서 sigmoid 함수를 이용하였다. sigmoid 함수의 수리적 표현은 식 (2)와 같다.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

식 (1)에서 가중치  $w_{jk}^{(1)}$ 와  $w_{km}^{(2)}$ 는 최소자승법, 즉 잔차에 대한 제곱합(residual sum of squares : RSS)을 최소화하는 값으로서 추정된다. RSS에 대한 수리적 표현은 식 (3)과 같다.

$$RSS = \sum_{m=1}^M \sum_{i=1}^I (y_{mi} - \hat{y}_{mi})^2 \quad (3)$$

만일 식 (3)이 이진분류 모형에 대한 식이라면  $M$ 은 1이 된다. 이 경우, RSS에 대하여 가중치 상쇠 모수(weight decay parameter)라고 하는 경로 가중치에 대한 일종의 페널티를 부여함으로써 모형의 복잡성 정도를 통제할 수 있다. 이에 대한 모형은 식 (4)와 같다.

$$\min = \sum_{i=1}^I (y_{1i} - \hat{y}_{1i})^2 + \frac{\lambda}{2} \sum w^2 \quad (4)$$

여기서,  $\lambda$ 는 가중치 감쇠 모수인데 일반적으로 1보다 작은 양의 상수로 설정된다. 이 값은 인공신경망 모형의 경로 가중치에 대해 유의하지 않은 가중치를 0에 가깝게 만들어줌으로써 경로의 수를 줄여주는 효과를 가져오는데,  $\lambda$ 의 값이 크면 클수록 모형의 복잡성을 통제하는 정도가 크게 나타나게 된다. 인공신경망 모형에 있어 가중치 감쇠 모수의 도입 타당성은 가중치 상쇠 모수를 적용한 모형이 그렇지 않은 모형에 비해 통계적으로 유의하게 높은 성과를 보인다고 보고한 여러 실증 연구[3, 9, 12, 13, 14]를 통해 지지되고 있다.

인공신경망 모형의 복잡성 정도는 모형의 일반화 문제와 연결되어 있다[11]. 모형이 복잡할수록 주어진 데이터에 모형이 적합하는 정도는 커지지만, 새로운 데이터에 대해 모형이 적합하는 정도는 낮아지게 된다. 다시 말해, 모형이 복잡하면 복잡할수록 모형의 과적합 문제(overfitting problem), 즉 훈련용 데이터에 대해서는 분류정확도가 높지만 검증용 데이터에 대해서는 예측력이 낮아지는 문제가 발생할 가능성이 높아진다. 반면에 모형이 지나치게 단순하게 되면 과적합 문제가 발생할 가능성은 낮아지지만, 주어진 데이터에 대해 모형이 충분히 적합하지 않게 됨으로써 새로운 데이터에 대해서도 예측력이 떨어지는 결과로 이어지게 된다. 따라서 인공신경망 모형의 구축에 있어서 모형의 복잡성 정도를 적절히 통제하는 것은 모형의 예측력을 높이기 위한 매우 중요한 과제라고 할 수 있다.

인공신경망 모형의 복잡성을 통제하는 방법으로 는 앞에서 언급한 가중치 상쇠 모수의 크기 설정을 이용한 방법 외에도 투입변수의 선택[20], 은닉층 수의 설정, 각 은닉층의 은닉노드 수 설정 등을 통한 방법이 있다. 이 중, 은닉층 수에 대해서는 많은 선행 연구에서 은닉층 수가 1개인 모형과 2개인 모형의 성과는 서로 의미 있는 차이를 나타내지 않는다고 보고되고 있다[8, 10, 17]. 따라서 본 연구에

서도 은닉층 수는 1개로 설정하여 연구를 진행하였다. 다음으로 투입변수의 선택 문제는 4.2절에서 구체적으로 다룰 것이다. 나머지 은닉노드 수와 가중치 감쇠 모수 설정 문제에 대해 본 연구에서는 제 3장에서 기술하는 세 가지 방법을 통해 설정한 후 그 성과를 비교하고자 한다.

### 3. 실험 디자인

다집단 분류를 위한 인공지능망 모형의 경우, 일반 모형이든 분리 모형이든 상관없이 모형의 복잡성을 통제하기 위해 투입변수, 은닉노드 수, 가중치 감쇠 모수 등의 요소를 설정하는 것은 모형 구축에 앞서 우선적으로 이루어져야 한다. 그럼에도 불구하고 다집단 분류를 위한 인공지능망 모형에 대한 선행 연구에서는 이러한 튜닝 과정에 대한 구체적인 설명이 없는 경우가 대부분이다. 투입변수의 선택에 대해서는 비교적 자세히 다루고 있지만, 은닉노드 수나 가중치 감쇠 모수를 결정하는 문제에 대해서는 결정 과정에 대한 구체적인 설명이 없이 결과만 간단히 언급하는 경우가 대부분이다.

이에 본 연구에서는 투입변수를 통제된 상태에서 은닉노드의 수와 가중치 감쇠 모수를 결정하는 방법을 다음과 같이 세 가지 실험 디자인을 통해 수행하고 그 성과를 비교하고자 한다. 세 가지 실험 디자인을 설명하면 다음과 같다.

- **Design 1**(일반 모형에 대해 [은닉노드 수, 가중치 감쇠 모수] 1개 세트 적용) : 인공지능망 일반 모형일 경우에는 구축되는 모형의 수가 1개이므로 적용시키는 [은닉노드 수, 가중치 감쇠 모수] 세트 자체도 1개일 수밖에 없다. 이 경우, 격자탐색법을 이용하여 테스트용 데이터에 대해 가장 높은 성과를 보이는 세트를 구하여 모형을 구축하도록 한다. 즉, 훈련용 데이터를 이용하여 일반 모형 구축 시 테스트용 데이터에 대하여 가장 성과가 우수하게 나타나는 [은닉노드 수, 가중치 감쇠 모수] 세트를 이용하여 일반 모형을 구축한

후, 이를 검증용 데이터에 적용하여 분류 성과를 확인한다.

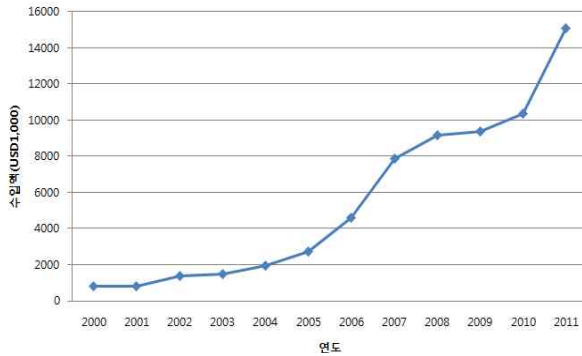
- **Design 2**(분리 모형 각각에 대하여 상이한 [은닉노드 수, 가중치 감쇠 모수] 세트 적용) : 인공지능망 분리 모형의 경우에는 이원 분류 모형이 집단의 수만큼 구축된다. 이 경우에 각 이원 분류 모형에 대해 격자탐색법을 이용하여 테스트용 데이터 기준으로 가장 높은 성과를 나타내는 [은닉노드 수, 가중치 감쇠 모수] 세트를 구하여 이를 적용하여 모형을 구축한다. 그 후, 검증용 데이터에 모형을 적용하여 분류 성과를 확인한다.
- **Design 3**(분리 모형 각각에 대하여 동일한 [은닉노드 수, 가중치 감쇠 모수] 세트 적용) : Design 2의 성과와 비교하기 위해 임의의 동일한 [은닉노드 수, 가중치 감쇠 모수] 세트를 이용하여 모든 이원 분류 모형을 구축한 후, 검증용 데이터에 모형을 적용하여 분류 성과를 확인한다.

### 4. 실증 분석

본 연구에서 제시하는 방법의 타당성을 실증적으로 검증하기 위해 와인 등급 데이터를 이용하여 제안한 모형의 성과를 분석하였다. 와인의 등급 예측은 다음과 같은 점에서 실무적 의미를 가진다.

우리나라가 칠레, 미국, EU 등의 국가와 자유무역협정을 체결하면서 국내로 수입되는 와인의 규모는 [그림 3]에서 보는 바와 같이 해마다 빠른 속도로 증가하고 있는 추세이다. 이처럼 와인 소비가 증가하는 시장 상황에서 와인 등급을 적절히 예측할 수 있는 기법이 제공되면 우선, 국내 소비자들에게 양질의 와인을 적정 가격에 공급할 수 있는 가능성을 높일 수 있다. 즉, 본 연구의 모형을 적용하여 와인의 등급을 평가하게 되면 평가 등급과 실제 등급의 차이를 알 수 있게 되어, 평가 등급에 비해 실제 등급이 현저히 낮은 경우에는 가격을 조정하는 식으로 소비자에게 적정 가격으로 와인을 공급하도록 유도할 수 있을 것이다. 여기서, 평가의 주체는 와인 수

입업자나 유통업자가 될 수도 있고, 와인 유통을 감독하는 식약청이나 소비자 단체일 수도 있다.



자료원 : 관세청 홈페이지 ‘무역통계조회’에서 자료 수집 후 작성.

[그림 3] 국내 와인 수입액 추이

또한 적정 와인 등급의 예측 필요성은 수입 와인과 경쟁하는 국내 와인 제조사에게도 관련 정보를 제공할 수 있다는 측면에서 제기할 수 있다. 즉, 본 연구의 모형을 적용하여 와인 등급의 평가에 있어 중요한 요소가 무엇인지 식별하고, 고급 와인에 대하여 각 요소의 수준을 알아냄으로써 국내 와인 제조사가 수입 와인과 경쟁할 수 있는 고급 와인을 생산해 내는데 도움을 주는 정보로 이를 활용할 수 있을 것이다.

#### 4.1 데이터 정리

본 연구에서 사용한 데이터는 Cortez et al.[7]에서 처음 만들어져 분석에 이용된 데이터이다. 이 데이터는 레드 와인 1599종과 화이트 와인 4898종을 대상으로 화학 실험 조사를 통해 얻어진 11개 항목의 수치와 함께 와인 전문가들이 이들 와인에 매긴 등급을 포함하고 있다.

<표 1>에서 X1부터 X11은 화학 실험 조사를

<표 1> 변수의 정의

	변수	변수정의
투입변수	X1	고정산(fixed acidity)
	X2	휘발성산(volatle acidity)
	X3	구연산(citric acid)
	X4	잔당(residual sugar)
	X5	염화물(chlorides)
	X6	free sulfur dioxide
	X7	total sulfur dioxide
	X8	밀도(density)
	X9	산도(pH)
	X10	황산(sulphates)
	X11	알코올(alcohol)
산출변수	X12	등급(0점~10점)

통해 구한 수치를 나타내고, X12은 3명의 와인 전문가들에게 와인의 등급을 0점(최악)에서 10점(최상) 사이에서 평가하도록 하여 나온 값들의 중앙값을 나타낸다. 여기서, 전문가 평가의 평균을 쓰지 않고 중앙값을 이용한 이유는 평균에 비해 중앙값이 분산이 낮은 이점 외에도 평균은 정수로 떨어지지 않을 수 있지만 중앙값은 자료의 수가 3개일 경우 정수로 나타나기 때문에 집단 분류를 위한 본 연구의 용도에 부합하는 자료의 형태가 되기 때문이다.

<표 2>는 와인 전문가들이 매긴 등급에 해당하는 레드 와인과 화이트 와인의 분포를 나타낸다. 레드 와인과 화이트 와인 모두 6등급에 해당하는 와인이 가장 높은 빈도를 보이고 6등급에서 멀어질수록 빈도가 낮아지는 것을 확인할 수가 있다. 그런데 이 데이터를 분석에 그대로 이용하면 문제가 발생할 수 있다. 부연하자면, 주어진 데이터는 훈련용, 테스트용, 검증용 데이터로 나뉘어져 훈련용 데이터는 모형 구축을 위해, 테스트용 데이터는 모형

<표 2> 와인 등급별 데이터 수

	등급							전체
	3	4	5	6	7	8	9	
레드 와인	10	53	681	638	199	18	0	1599
화이트 와인	20	163	1457	2198	880	175	5	4898

의 복잡성 통제를 위한 모수 설정을 위해, 검증용 데이터는 구축된 모형의 성과를 측정하기 위해 이용된다. 따라서 각 등급의 데이터는 세 가지 용도의 데이터셋에 미리 정한 비율로 배분이 될 것인데, 문제는 3, 4, 8, 9등급의 경우 주어진 데이터의 수가 너무 적어서 모형을 구축하고 검증하기가 어렵다는 것이다.

<표 3> 변환된 와인 등급별 데이터 수

	등급				전체
	4	5	6	7	
레드 와인	63	681	638	217	1599
화이트 와인	183	1457	2198	1060	4898

따라서 분석의 편의를 위해 본 연구에서는 <표 3>과 같이 3등급과 4등급을 4등급으로 합치고, 7등급, 8등급, 9등급을 7등급으로 합쳐서 결국, 전체 등급이 4, 5, 6, 7등급의 네 개 등급으로 이루어지도록 조정하였다. 그런 다음, 단순무작위 추출법을 이용하여 <표 4>와 같이 훈련용, 테스트용, 검증용 데이터의 비율이 6:2:2의 비율이 되도록 전체 데이터를 나누었다.

<표 4> 용도별 데이터 수

		레드 와인	화이트 와인
훈련용	4	36	104
	5	409	873
	6	384	1320
	7	130	642
	소계	959	2939
테스트용	4	16	46
	5	131	291
	6	131	443
	7	42	200
	소계	320	980
검증용	4	11	33
	5	141	293
	6	123	435
	7	45	218
	소계	320	979

## 4.2 투입변수의 선정

본 연구에서는 인공신경망 모형의 투입변수를 선정하기 위해 우선, 상관분석, 요인분석 등의 통계 기법을 이용하고, 최종적으로는 변수의 의미상 중요하다고 판단되는 변수를 선정하였다. 이 과정을 구체적으로 기술하면 다음과 같다.

우선, 와인 등급 변수 X12와 투입변수들 간의 스피어만 순위상관계수를 구하였다. <표 5>에서 보면 유의수준 5%에서 유의하지 않은 변수, 다시 말해 와인 등급과 선형 관계를 나타낸다고 보기 어려운 변수는 레드 와인의 경우 X4와 X9이었고, 화이트 와인의 경우에는 X3과 X6이었다. 따라서 이들 변수는 투입변수 대상에서 제외하였다.

<표 5> 와인 등급과 투입변수 간 순위상관계수

	레드 와인	화이트 와인
X1	.114**	-.084**
X2	-.381**	-.197**
X3	.213**	.018
X4	0.032	-.082**
X5	-.190**	-.314**
X6	-.057*	.024
X7	-.197**	-.197**
X8	-.177**	-.348**
X9	-0.044	.109**
X10	.377**	.033*
X11	.479**	.440**

\*유의수준 0.05에서 유의함.  
\*\*유의수준 0.01에서 유의함.

다음으로, 요인분석을 통해 11개의 투입변수들을 몇 개의 유사한 변수 집단으로 그룹화 하였다. <표 6>은 베리맥스 회전을 통해 얻어진 회전행렬이다.

<표 6>에서 레드 와인과 화이트 와인 모두 4개의 성분이 추출되었고, 각 투입변수는 추출된 4개의 성분의 선형결합으로 표현이 가능하다. 이때, 각 투입변수를 표현하는 성분 계수 중에 절대값이 가장 높은 계수가 무엇인지를 보면 각 투입변수가 속하는 변수 집단이 무엇인지 알 수가 있다. 예를 들

&lt;표 6&gt; 요인분석 결과 : 회전행렬

와인 종류	변수	성분			
		1	2	3	4
레드	X1	.912	-.127	.072	.025
	X8	.776	.105	-.437	.013
	X3	.745	.038	.444	.230
	X9	-.732	.011	-.016	-.254
	X6	-.139	.880	.034	.066
	X7	-.023	.876	-.109	.095
	X4	.345	.460	-.032	-.247
	X11	-.234	-.113	.778	-.134
	X2	-.333	-.020	-.718	-.127
	X5	.134	.001	-.273	.801
	X10	.153	.062	.270	.758
화이트	X8	.890	.221	-.136	.062
	X4	.802	.151	-.081	-.221
	X7	.766	-.091	.140	.187
	X11	-.741	-.165	.194	-.197
	X6	.628	-.199	.422	.044
	X1	.075	.794	.070	.013
	X9	-.074	-.739	.103	.230
	X3	.065	.547	.376	.379
	X2	.032	-.101	-.699	-.085
	X10	.019	-.169	.188	.669
	X5	.240	.153	-.477	.603

어, 레드 와인에서 X1, X8, X3, X9은 성분 1의 계수 절대값이 나머지 성분들의 계수 절대값에 비해 크게 나타났으므로, 이들 변수들은 비슷한 특성을 공유하고 있는 것으로 판단할 수 있다. 이와 같은 방식으로 11개의 투입변수를 4개의 집단으로 나누고, 앞서 언급한 스피어만 순위상관분석 결과 유의하지 않은 변수로 판단된 변수를 제외하고 나면, 레드 와인의 경우에는 X1, X3, X8을 한 집단, X6, X7을 한 집단, X2, X11을 한 집단, X5, X10을 한 집단으로 묶을 수 있고, 화이트 와인의 경우에는 X4, X7, X8, X11을 한 집단, X1, X9를 한 집단, X2를 한 집단, X5, X10을 한 집단으로 묶을 수가 있다.

마지막으로, 각 투입변수 집단에서 화학적으로 중요하게 판단되는 변수 1개씩을 선정하였다. 구체적으로, <표 6>에서 레드 와인의 성분 1과 화이트 와인의 성분 2, 그리고 레드 와인의 성분 2와 화이트 와인의 성분 1은 각각 비슷한 성분으로 볼 수 있다. 또한 레드 와인의 성분 3, 4와 화이트 와인의 성분 3, 4도 각각 비슷한 성분으로 판단할 수 있다. 그런데 유사한 성분으로 판단된 레드 와인의 성분 1과 화이트 와인의 성분 2에 모두 포함된 변수는 X1이고, 레드 와인의 성분 2와 화이트 와인의 성분 1에 모두 포함된 변수는 X7이다. 또한 레드 와인 성분 3과 화이트 와인 성분 3에 모두 포함된 변수는 X2이고, 레드 와인 성분 4와 화이트



트 와인 성분 4에 모두 포함된 변수는 X5와 X10이다. 여기서 Cortez et al.[7]은 X5(염화물)를 11개 투입변수 중에서 가장 상대적 중요도가 낮은 변수로 보고하고 있다. 이에 본 연구에서는 각 성분을 대표할 수 있는 변수로서 레드 와인과 화이트 와인 모두에 대하여 X1(고정산), X2(휘발성산), X7(total sulfur dioxide), X10(황산)을 투입변수로 선정하였다.

#### 4.3 분석결과와 논의

인공신경망 모형의 구축을 위한 3가지 실험 디자인에 대하여 공통적으로 설정한 몇 가지 요소를 정리하면 다음과 같다.

우선, 투입변수로는 X1, X2, X7, X10 4개의 변수를 이용하였다. 은닉층의 수는 1개로 하였고, 따라서 본 연구에 이용한 인공신경망 모형은 투입층(input layer), 은닉층(hidden layer), 출력층(output layer)의 3개 층으로 이루어졌다. 은닉층과 출력층에서 변환함수로는 sigmoid 함수를 이용하였다.

다음으로 3가지 실험 디자인에 대하여 상이하게 설정한 요소는 다음과 같다. 먼저, Design 1은 일반 모형의 경우로서 동일한 경로 가중치가 출력층의 모든 노드에 적용된다. 또한 1개 세트의 [은닉노드 수, 가중치 감쇠 모수]를 이용하여 모형을 구축한다. 여기서, 적용되는 [은닉노드 수, 가중치 감쇠 모수] 세트를 결정하기 위해 격자탐색법이 이용되었다. 즉,  $[2^1, 2^2, 2^3, 2^4, 2^5, 2^6]$ 을 은닉노드 수의 대안으로 하고,  $[2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0]$ 을 가중치 감쇠 모수의 대안으로 하여, 주어진 [은닉노드 수, 가중치 감쇠 모수] 세트 중에서 테스트용 데이터에 대하여 가장 오분류율이 낮게 나타나게 하는 세트를 최적 세트로 도출한다.

Design 2는 분리 모형으로서 여러 개의 이집단 분류 모형으로 이루어져 있고, 각 이집단 분류 모형에 대하여 구해지는 경로 가중치는 물론이고 적용되는 [은닉노드 수, 가중치 감쇠 모수] 세트도 서로 다르게 나타나게 된다. 여기서 각 이집단 분

류 모형에 적용되는 [은닉노드 수, 가중치 감쇠 모수] 세트를 결정하는 방법은 Design 1에서와 같이 격자탐색법을 이용하였다. 구체적인 방법은 앞서 기술한 바와 같다.

Design 3 역시 분리 모형이지만, Design 2와 차이점은 분리 모형 안에 있는 모든 이집단 분류 모형에 대해 임의의 동일한 [은닉노드 수, 가중치 감쇠 모수] 세트를 적용한다는 것이다. 본 연구에서는 은닉노드 수를 7로 하고, 가중치 감쇠 모수를 0.1로 하여 적용하였다.

이와 같은 설정 하에 각 [은닉노드 수, 가중치 감쇠 모수] 세트를 적용한 모형의 테스트용 데이터에 대한 오분류율을 구하여 정리하면 <표 7>과 같다.

<표 7>에 보는 바와 같이 [은닉노드 수, 가중치 감쇠 모수] 세트가 변화함에 따라 인공신경망 모형의 성과는 달라지는 것을 확인할 수 있다. 이는 최적 [은닉노드 수, 가중치 감쇠 모수] 세트를 탐색하고 구하는 것이 타당성을 가짐을 의미한다. <표 7>에서 오분류율을 최소로 하는 [은닉노드 수, 가중치 감쇠 모수] 세트를 구하면, 이것이 바로 최적 [은닉노드 수, 가중치 감쇠 모수] 세트가 된다. 이와 같이 도출한 Design 1 및 Design 2의 최적 [은닉노드 수, 가중치 감쇠 모수] 세트와 본 연구에서 임의로 설정한 Design 3의 [은닉노드 수, 가중치 감쇠 모수] 세트를 요약하면 <표 8>과 같다.

<표 8>에서 보는 바와 같이, Design 2에 대하여 레드 와인과 화이트 와인 모두 4개의 서로 다른 집단을 분리할 때 가장 성과를 크게 하는 [은닉노드 수, 가중치 감쇠 모수] 세트는 서로 상이하게 나타남을 알 수 있다. 즉, 인공신경망 분리 모형의 경우, 분리하고자 하는 집단에 따라 적용해야 하는 최적의 [은닉노드 수, 가중치 감쇠 모수] 세트는 서로 동일하지 않다는 것이다. 따라서 인공신경망의 분리 모형을 구축할 때에는 각 집단에 적합한 아키텍처를 튜닝하는 과정이 성과의 극대화를 위해 필요함을 알 수 있다.

<표 7> [은닉노드 수, 가중치 감쇠 모수] 세트에 대한 오분류율

은닉노드 수		가중치 감쇠 모수	$2^{-7}$	$2^{-6}$	$2^{-5}$	$2^{-4}$	$2^{-3}$	$2^{-2}$	$2^{-1}$	$2^0$			
레드 와인	Design 1		$2^1$	0.450	0.466	0.469	0.459	0.463	0.459	0.466	0.450		
			$2^2$	0.453	0.453	0.441	0.472	0.450	0.441	0.463	0.456		
			$2^3$	0.450	0.453	0.441	0.434	0.434	0.453	0.459	0.463		
			$2^4$	0.428	0.434	0.441	0.434	0.434	0.447	0.447	0.456		
			$2^5$	0.450	0.441	0.425	0.450	0.438	0.447	0.450	0.444		
	Design 2	4등급		$2^1$	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	
				$2^2$	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	
				$2^3$	0.044	0.041	0.041	0.047	0.041	0.066	0.041	0.041	
				$2^4$	0.047	0.044	0.041	0.050	0.056	0.050	0.063	0.038	
				$2^5$	0.044	0.044	0.053	0.041	0.041	0.050	0.056	0.053	
		5등급		$2^1$	0.050	0.047	0.044	0.050	0.050	0.047	0.044	0.059	
				$2^2$	0.284	0.284	0.406	0.281	0.356	0.375	0.284	0.300	
				$2^3$	0.278	0.294	0.322	0.288	0.344	0.306	0.284	0.344	
				$2^4$	0.284	0.334	0.353	0.288	0.284	0.294	0.278	0.266	
				$2^5$	0.328	0.309	0.309	0.369	0.319	0.306	0.372	0.278	
		6등급		$2^1$	0.272	0.306	0.294	0.275	0.291	0.297	0.278	0.319	
				$2^2$	0.281	0.288	0.294	0.306	0.319	0.244	0.272	0.313	
				$2^3$	0.422	0.422	0.422	0.422	0.422	0.422	0.422	0.422	
				$2^4$	0.397	0.422	0.422	0.422	0.422	0.422	0.453	0.425	
				$2^5$	0.419	0.422	0.422	0.409	0.434	0.419	0.400	0.419	
		7등급		$2^1$	0.441	0.403	0.416	0.378	0.397	0.388	0.397	0.388	
				$2^2$	0.409	0.419	0.388	0.444	0.428	0.450	0.406	0.416	
				$2^3$	0.406	0.409	0.425	0.409	0.409	0.406	0.431	0.419	
				$2^4$	0.131	0.131	0.131	0.131	0.131	0.131	0.131	0.131	
				$2^5$	0.141	0.131	0.131	0.131	0.131	0.131	0.128	0.131	
		화이트 와인	Design 1		$2^1$	0.122	0.131	0.144	0.144	0.134	0.138	0.134	0.131
					$2^2$	0.125	0.144	0.141	0.134	0.156	0.131	0.134	0.125
					$2^3$	0.138	0.134	0.134	0.128	0.128	0.138	0.134	0.134
$2^4$					0.128	0.131	0.131	0.122	0.125	0.138	0.122	0.119	
$2^5$					0.128	0.131	0.131	0.122	0.125	0.138	0.122	0.119	
Design 2			$2^1$	0.542	0.540	0.539	0.540	0.535	0.539	0.538	0.547		
			$2^2$	0.519	0.540	0.541	0.527	0.530	0.536	0.537	0.533		
			$2^3$	0.539	0.535	0.536	0.547	0.538	0.538	0.535	0.532		
			$2^4$	0.536	0.537	0.548	0.533	0.538	0.547	0.533	0.539		
			$2^5$	0.547	0.528	0.546	0.539	0.531	0.543	0.538	0.532		
			4등급		$2^1$	0.533	0.534	0.539	0.539	0.545	0.534	0.541	0.543
					$2^2$	0.046	0.046	0.046	0.046	0.046	0.046	0.046	0.046
					$2^3$	0.046	0.046	0.045	0.046	0.046	0.046	0.045	0.046
					$2^4$	0.046	0.046	0.045	0.046	0.046	0.044	0.046	0.044
					$2^5$	0.051	0.044	0.045	0.045	0.045	0.049	0.045	0.046
			5등급		$2^1$	0.045	0.045	0.044	0.047	0.046	0.047	0.046	0.048
					$2^2$	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288
					$2^3$	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288
					$2^4$	0.288	0.293	0.291	0.288	0.289	0.288	0.288	0.288
					$2^5$	0.287	0.288	0.298	0.288	0.296	0.289	0.289	0.293
			6등급		$2^1$	0.299	0.273	0.293	0.285	0.287	0.279	0.293	0.279
					$2^2$	0.292	0.291	0.293	0.291	0.295	0.285	0.294	0.296
					$2^3$	0.440	0.440	0.440	0.440	0.440	0.442	0.440	0.440
					$2^4$	0.440	0.442	0.442	0.442	0.440	0.441	0.440	0.440
					$2^5$	0.442	0.430	0.441	0.440	0.448	0.437	0.439	0.440
7등급			$2^1$	0.450	0.442	0.442	0.440	0.444	0.431	0.436	0.440		
			$2^2$	0.441	0.436	0.442	0.444	0.439	0.442	0.438	0.444		
			$2^3$	0.434	0.441	0.438	0.456	0.436	0.438	0.441	0.453		
		$2^4$	0.227	0.227	0.227	0.227	0.227	0.227	0.227	0.227			
		$2^5$	0.227	0.227	0.227	0.227	0.227	0.227	0.227	0.227			
7등급		$2^1$	0.227	0.227	0.227	0.227	0.227	0.227	0.227	0.227			
		$2^2$	0.227	0.227	0.227	0.227	0.227	0.227	0.227	0.227			
		$2^3$	0.227	0.227	0.227	0.230	0.227	0.226	0.227	0.227			
		$2^4$	0.227	0.226	0.223	0.227	0.227	0.227	0.229	0.227			
		$2^5$	0.227	0.227	0.227	0.227	0.226	0.224	0.232	0.234			
7등급		$2^1$	0.227	0.226	0.232	0.230	0.226	0.228	0.224	0.227			
		$2^2$	0.227	0.226	0.232	0.230	0.226	0.228	0.224	0.227			
		$2^3$	0.227	0.226	0.232	0.230	0.226	0.228	0.224	0.227			
		$2^4$	0.227	0.226	0.232	0.230	0.226	0.228	0.224	0.227			
		$2^5$	0.227	0.226	0.232	0.230	0.226	0.228	0.224	0.227			

<표 8> 실험 디자인별 최적[은닉노드 수, 가중치 감쇠 모수]

		레드 와인		화이트 와인	
		은닉노드 수	가중치 감쇠 모수	은닉노드 수	가중치 감쇠 모수
Design 1		$2^6$	$2^{-5}$	$2^2$	$2^{-7}$
Design 2	집단 4	$2^4$	$2^{-0}$	$2^4$	$2^{-0}$
	집단 5	$2^6$	$2^{-2}$	$2^5$	$2^{-6}$
	집단 6	$2^4$	$2^{-4}$	$2^3$	$2^{-6}$
	집단 7	$2^6$	$2^{-0}$	$2^4$	$2^{-5}$
Design 3		7	0.1	7	0.1

<표 9> 실험 디자인별 와인 등급 예측

데이터	예측등급	실제등급	4	5	6	7
		4	5	6	7	
레드 와인	Design 1	4	0	0	0	0
		5	8	92	44	4
		6	3	44	70	27
		7	0	5	9	14
	Design 2	4	0	1	0	0
		5	8	96	46	9
		6	3	41	70	24
		7	0	3	7	12
	Design 3	4	0	0	0	0
		5	8	95	46	6
		6	3	43	69	28
		7	0	3	8	11
화이트 와인	Design 1	4	2	5	1	0
		5	13	114	80	17
		6	16	165	330	184
		7	2	9	24	17
	Design 2	4	3	5	1	0
		5	13	114	72	19
		6	16	171	343	187
		7	1	3	19	12
	Design 3	4	1	3	1	0
		5	7	25	9	8
		6	24	261	404	193
		7	1	4	21	17

<표 9>는 <표 8>의 최적 [은닉노드 수, 가중치 감쇠 모수] 세트를 이용하여 실험을 진행한 결과이다. 실험 디자인별로 인공지능경망 모형을 구축하고, 검증용 자료에 대하여 레드 와인과 화이트 와인의 등급을 예측한 결과를 요약한 것이다. 여기서 특징적인 점을 찾아보면, 먼저 실제 4등급과 7등급 와인에 대하여 3가지 실험 디자인 모두가 중간 등급, 다시 말해 5등급이나 6등급으로 예측하려는 경향이 있음을 알 수 있다. 다음으로, 전체 오분류율은

중간 등급인 5등급과 6등급에서 얼마나 정확하게 예측이 이루어지느냐에 따라 결정된다는 것을 알 수 있다. 레드 와인의 경우 Design 2가 5등급을 정확하게 예측한 빈도가 Design 1과 3에 비해 높게 나타났다. 화이트 와인의 경우에는 Design 1과 2가 Design 3에 비해 5등급을 정확히 예측한 빈도가 훨씬 높게 나타났고, 6등급에 대해서는 Design 2가 Design 1에 비해 정확히 예측한 빈도가 비교적 높게 나타났다.

이상의 결과를 통해 각각의 경우에 대한 오분류율을 계산하면 <표 10>과 같다. <표 10>을 보면, 오분류율은 최소 44.4%에서 최대 54.3%까지 나타나고 있는데, 이 결과를 보고 오분류율이 너무 높은 수준이 아닌가라는 의문이 제기될 수도 있다. 그러나 이진 분류가 아닌 4개 등급에 대한 분류에 있어 분류 정확도는 25% 이상이면 의미가 있는 것으로 볼 수 있다. 더구나,  $\pm 1$ 등급의 허용오차를 적용하게 되면 Design 2의 경우, 오분류율은 레드 와인은 4.69%, 화이트 와인은 4.09%까지 낮아진다.

<표 10> 실험 디자인별 오분류율

	레드 와인	화이트 와인
Design 1	0.450	0.527
Design 2	0.444	0.518
Design 3	0.453	0.543

<표 10>의 결과를 보면, Design 2가 Design 1이나 3에 비해 오분류율이 낮게 나타남을 알 수 있다. 즉, 오분류율을 기준으로 볼 때, Design 2는 Design 1이나 3에 비해 높은 성과를 보이고 있다. 그러나 이러한 결과는 하나의 데이터셋에 대해 분석한 결과이므로 일반화하기에는 무리가 있다. 따

라서 무작위표본추출법을 이용해서 데이터를 6:2:2의 비율로 훈련용, 테스트용, 검증용 데이터로 나누는 과정을 10번 반복하여 동일한 실험을 수행하였다. <표 11>은 그 결과를 정리한 것이다.

<표 11>에 보듯이 10개의 서로 다른 데이터셋에 대하여 3개의 실험 디자인에 의해 구축된 인공신경망 모형의 성과를 비교해 보면, 레드 와인과 화이트 와인 모두 Design 2가 Design 1이나 3에 비해 낮은 오분류율 평균을 나타내는 것을 확인할 수 있다.

마지막으로, 이러한 실험 디자인에 따른 오분류율 평균 차이가 통계적으로 유의한지를 검정하기 위해 쌍대 t-검정을 수행하였다. <표 12>에서 보는 것처럼, 쌍대 t-검정 결과, Design 2에 의해 구축된 인공신경망 모형의 성과가 Design 1이나 3에 의해 구축된 모형의 성과에 비해 통계적으로 유의하게 우수하다는 결론을 얻을 수 있었다.

요약하면, 다집단 분류를 위한 인공신경망 모형 구축에 있어서 일반 모형보다는 분리 모형의 성과가 우수하고, 분리 모형 구축에 있어서도 은닉노드 수와 가중치 감쇠 모수를 튜닝한 모형이 튜닝하지 않은 모형보다 성과가 더 우수함을 실험을 통해 확인할 수 있었다.

<표 11> 실험 디자인별 반복측정 오분류율

데이터셋	레드 와인			화이트 와인		
	Design 1	Design 2	Design 3	Design 1	Design 2	Design 3
1	0.466	0.459	0.466	0.513	0.512	0.525
2	0.466	0.447	0.459	0.510	0.494	0.525
3	0.438	0.406	0.409	0.564	0.527	0.564
4	0.484	0.444	0.447	0.539	0.524	0.521
5	0.491	0.481	0.494	0.540	0.517	0.522
6	0.459	0.456	0.466	0.517	0.510	0.514
7	0.438	0.416	0.428	0.518	0.501	0.501
8	0.481	0.45	0.447	0.538	0.512	0.558
9	0.425	0.416	0.425	0.518	0.494	0.512
10	0.475	0.441	0.456	0.517	0.514	0.537
평균	0.462	0.442	0.450	0.527	0.510	0.528
표준편차	0.022	0.023	0.024	0.017	0.011	0.020

<표 12> 오분류를 평균 차이에 대한 쌍대 t-검정 결과

		대응차이					t-값	자유도	유의확률 (양쪽)
		평균	표준 편차	평균의 표준오차	차이의 95% 신뢰구간				
					하한	상한			
레드 와인	Design 1-2	0.021	0.013	0.004	0.011	0.030	5.004	9	.001
	Design 3-2	0.008	0.006	0.002	0.004	0.012	4.569	9	.001
화이트 와인	Design 1-2	0.017	0.011	0.004	0.009	0.025	4.785	9	.001
	Design 3-2	0.017	0.017	0.005	0.006	0.029	3.308	9	.009

## 5. 결 론

본 연구는 다집단 분류를 위한 인공신경망 모형의 아키텍처 튜닝 기법을 제시하고, 그 타당성을 확인하기 위해 세 가지 실험 디자인을 설계하고 실증분석을 통해 모형의 성과를 비교하였다. 본 연구에서 제시하는 기법은 인공신경망 모형 중에서도 분리 모형을 대상으로 격자탐색법을 이용하여 최적의 은닉노드 수와 가중치 감쇠 모수를 결정하는 방법이다. 이 기법과 함께 다른 두 가지 방법에 의해 구축된 인공신경망 모형의 성과를 비교한 결과, 본 연구에서 제시하는 기법을 이용하여 구축한 인공신경망 모형이 다른 방법에 비해 통계적으로 유의하게 우수한 분류 성과를 보이는 것을 확인하였다. 이러한 결과를 통해 다집단 분류를 위한 인공신경망 모형을 구축할 때, 본 연구에서 제시하는 아키텍처 튜닝 기법을 적용하게 되면 그렇지 않은 경우에 비해 향상된 분류 성과를 얻을 수 있을 것으로 기대한다.

이러한 기여에도 불구하고 본 연구는 몇 가지 점에서 한계를 가진다. 첫째, 본 연구는 인공신경망의 분리 모형을 적용하는 방법이 OPP 방법, 이진분류 방법, 단순분류 방법 등으로 다양하게 있지만 그 중에서 단순분류 방법만을 이용하였다. 본 연구의 목적이 모형의 아키텍처 튜닝의 당위성을 주장하는데 있다고 할 때, 기존 연구에서 가장 성과가 높다고 보고된 단순분류 방법만을 이용해도 충분하다고 판단하여 연구를 진행하였지만, 향후 연구에서는 다양한 적용 방법에 대해서도 본 연구에서 제시한 튜

닝 기법이 효과적인지 확인할 필요가 있을 것이다.

둘째, 본 연구에서 은닉노드 수와 가중치 감쇠 모수를 튜닝하는 기법으로 격자탐색법을 이용하였는데, 격자탐색법은 전역적 탐색(global search) 접근법이 아닌 지역적 탐색(local search) 접근법이라는 한계를 가진다. 향후 연구에서는 전역적 탐색을 할 수 있는 방법, 예를 들어, 분석 시간은 많이 소요되겠지만 유전 알고리즘 같은 기법을 전역적 해를 도출하는데 활용할 수 있을 것이다.

## 참 고 문 헌

- [1] 김다윗, 민성환, 한인구, “신경망 분리모형과 사례 기반추론을 이용한 기업 신용 평가”, 『Journal of Information Technology Applications and Management』, 제14권, 제2호(2007), pp.151-168.
- [2] Altman, E.I., G. Marco, and F. Varetto, “Corporate distress diagnosis comparisons using discriminant analysis and neural networks,” *Journal of Banking and Finance*, Vol.18, No.3 (1994), pp.505-529.
- [3] Bartlett, P.L., “For valid generalization, the size of the weights is more important than the size of the network,” in M.C. Mozer, M.I. Jordan and T. Petsche (Eds.), *Advances in Neural Information Processing Systems*, Vol.9, The MIT Press, Cambridge, MA, 1997.
- [4] Berry, M.J.A. and G.S. Linoff, *Data Mining Techniques : For Marketing, Sales, and Cus-*

- tomor Relationship Management*, Wiley, Indiana, 2004.
- [5] Bishop, C.M., *Neural networks for pattern recognition*, Oxford University Press, New York, 1995.
- [6] Cheng, B. and D.M. Titterington, "Neural Networks : A Review from a Statistical Perspective," *Statistical Science*, Vol.9, No.1(1994), pp.2-30.
- [7] Cortez, P., A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," *Decision Support Systems*, Vol.47, No.4 (2009), pp.547-553.
- [8] Cybenko, G., "Approximation by superpositions of a sigmoid function," *Mathematics of Control, Signals, and Systems*, Vol.2, No.4 (1989), pp.303-314.
- [9] De Veaux, R.D., J. Schumi, J. Schweinsberg, and L.H. Ungar, "Prediction intervals for neural networks via nonlinear regression," *American Statistical Association and the American Society for Quality*, Vol.40, No.44(1998), pp.273-282.
- [10] De Villiers, J. and E. Barnard, "Backpropagation neural nets with one and two hidden layers," *IEEE Transactions on Neural Networks*, Vol.4, No.1(1993), pp.136-141.
- [11] Geman, S., E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *IEEE Neural Computation*, Vol.4, No.1(1992), pp.1-58.
- [12] Hinton, G.E., "Learning translation invariant recognition in massively parallel networks," In J.W. de Bakker, A.J. Nijman and P.C. Treleaven (Eds.), *Proceedings PARLE Conference on Parallel Architectures and Languages Europe*, Springer-Verlag, Berlin, 1987.
- [13] Hush, D.R. and B.G. Horne, "Progress in Supervised Neural Networks," *IEEE Signal Processing Magazine*, Vol.10, No.1(1993), pp.8-39.
- [14] Jeong, C., J.H. Min, and M.S. Kim, "A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction," *Expert Systems with Applications*, Vol.39, No.3(2012), pp.3650-3658.
- [15] Jo, H. and I. Han, "Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction," *Expert Systems with Applications*, Vol.11, No.4(1996), pp.415-422.
- [16] Kim, J., H.R. Weistroffer, and R.T. Redmond, "Expert Systems for Bond Rating : A Comparative Analysis of Statistical, Rule-based and Neural Network Systems," *Expert Systems*, Vol.10, No.3(1993), pp.167-172.
- [17] Masters, T., *Practical Neural Network Recipes in C++*, Academic Press, Boston, 1993.
- [18] Ou, G. and Y.L. Murphey, "Multi-class pattern classification using neural networks," *Pattern Recognition*, Vol.40, No.1(2007), pp. 4-18.
- [19] Singleton, J.C. and A.J. Surkan, "Neural Networks for Bond Rating Improved by Multiple Hidden Layers," *Proceedings of the IEEE International Conference on Neural Networks*, Vol.2(1990), pp.163-168.
- [20] Zhao, H., A. Sinha, and W. Ge, "Effects of feature construction on classification performance : An empirical study in bank failure prediction," *Expert Systems with Applications*, Vol.36, No.2(2009), pp.2633-2644.