

# R의 neuralnet을 활용한 신경망분석

백재욱  
한국방송통신대학교

## Neural network analysis using neuralnet in R

Jaiwook Baik

Department of Statistics·Data Science, Korea National Open University

요 약 본 연구는 다층 퍼셉트론과 지도형 학습알고리즘에 대해 알아보았고, 아울러 neuralnet이라는 패키지를 사용하여 공변수들과 반응변수 간의 함수적 관계를 어떻게 모델링하는지 살펴보았다. 본 연구에서 적용된 알고리즘은 반응변수 값의 실제치와 예측치 간의 비교에 근거한 오차함수의 최소화를 위한 모수인 가중치들의 지속적인 조정을 특징으로 한다. 본 연구에서 설명하는 neuralnet 패키지는 활성화함수와 오차함수를 주어진 상황에 맞게 적절히 선택하고 나머지 매개변수들은 기본값으로 둘 수 있다. 본 연구에서 살펴본 불임 데이터에 대해 neuralnet 패키지를 활용한 결과 4개의 독립변수 중에서 age는 불임에 영향력이 거의 없음을 파악할 수 있었다. 아울러 신경망의 가중치는 -751.6부터 7.25에 이르기까지 다양한 값을 취하며, 첫 번째 은닉층의 절편은 -92.6과 7.25이며, 첫 번째 은닉뉴런으로 가는 공변수 age, parity, induced, spontaneous에 대한 가중치는 각각 3.17, -5.20, -36.82, -751.6임을 파악했다.

주제어 다층 퍼셉트론, 지도학습, 역전파, 뉴럴넷

**Abstract** We investigated multi-layer perceptrons and supervised learning algorithms, and also examined how to model functional relationships between covariates and response variables using a package called neuralnet. The algorithm applied in this paper is characterized by continuous adjustment of the weights, which are parameters to minimize the error function based on the comparison between the actual and predicted values of the response variable. In the neuralnet package, the activation and error functions can be appropriately selected according to the given situation, and the remaining parameters can be set as default values. As a result of using the neuralnet package for the infertility data, we found that age has little influence on infertility among the four independent variables. In addition, the weight of the neural network takes various values from -751.6 to 7.25, and the intercepts of the first hidden layer are -92.6 and 7.25, and the weights for the covariates age, parity, induced, and spontaneous to the first hidden neuron are identified as 3.17, -5.20, -36.82, and -751.6.

**Key Words** Multi-layer perceptron, Supervised learning, Back propagation, Neuralnet

Received 04 Jan 2021, Revised 14 Jan 2021

Accepted 22 Jan 2021

Corresponding Author : Jaiwook Baik  
(Korea National Open University)

Email: jbaik@knou.ac.kr

ISSN: 2466-1139(Print)

ISSN: 2714-013X(Online)

© Industrial Promotion Institute. All rights reserved. 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.

## 1. 서론

요즘과 같은 불확실성의 시대에서는 여러 상황에서 통상 입력변수로 알려진 공변량들과 출력변수로 알려진 반응변수 간에 함수적 관계에 대해 큰 관심을 갖는다. 예를 들어 복잡한 질병에 대해 모델링할 때 질병에 대한 잠재 위험요인들과 이들의 영향에 대해 살펴보는 데, 이는 해당 질병에 대한 예방 또는 개입정책의 입안에 필요하다. 이때 인공신경망을 이용하면 두 변수들간의 어떤 복잡한 함수관계도 대략적으로 설명할 수 있다.

인공신경망은 일반화선형모델(Generalized Linear model, GLM)[1]과 달리 공변수들과 반응변수 간 특정한 관계식(예를 들어 선형결합식)을 미리 지정할 필요가 없으므로 더욱 유용한 통계도구이다. 인공신경망은 특히 GLM의 연장선상에 있는 도구이므로 이와 비슷한 방식으로 활용될 수 있다. 인공신경망에 대한 전반적인 개관은 여러 연구에서 설명되어 있다[2, 3, 4].

인공신경망에서 관측 데이터는 신경망을 훈련시키는데 사용되며, 신경망은 반복적으로 모수들을 변화시켜 두 변수들 간의 관계를 학습한다. 특히, neuralnet이란 패키지[5]는 위와 같은 상황에서 두 변수들 간 관계를 대략적으로 표현할 수 있는, 전향적 신경망을 훈련하는 매우 유연한 함수를 갖고 있다.

현재 R로 인공신경망을 다룰 수 있는 패키지로는 nnet와 AMORE가 있다[6, 7]. nnet는 통상적인 역전파로 전진형 신경망을 훈련시킬 수 있다. AMORE에는 TAO robust 신경망 알고리즘이 구현되어 있다. 한편, neuralnet은 회귀분석과 같은 상황에서 적용할 수 있도록 인공신경망을 훈련시킨다[8]. 또한 neuralnet에서는 탄력적 역전파가 사용되는데, 그 이유는 탄력적 역전파가 신경망 분석에서 가장 빠른 알고리즘 중의 하나이기 때문이다[9, 10]. neuralnet이란 패키지는 복잡한 질병들을 성공적으로 모델링하는데 사용되기도 했다[11].

이에 본 연구에서는 인공신경망을 새로 접하는 사용자에게 R의 neuralnet이란 패키지를 잘 사용할 수 있도록 설명하고자 한다. 이를 위해 2절에서는 신경망을 소개하는 차원에서 다층 퍼셉트론에 대해 설명하고, 3절에서는 다층 퍼셉트론의 모수인 가중치를 효율적으로 추정할 수 있는 역전파 알고리즘을 이용하여 지도학습은 어떻게 이루어지는지 알아본다. 4절에서는 neuralnet이란 패키

지를 이용하여 주어진 데이터에 대해 신경망 모델을 어떻게 적합시키는지 알아보고 5절에서 결론을 맺는다.

## 2. 다층 퍼셉트론

neuralnet 패키지는 변수들 간 함수적 관계를 모델링할 때 잘 적용될 수 있는 다층 퍼셉트론(Multi-layer perceptron, MLP)에 초점을 두었다[12]. MLP의 하부구조는 방향성이 있는 그래프로, 뉴런이라고 불리는 점들과 방향성이 있는 간선으로 구성되어 있다. neuralnet에서 뉴런은 다음 층에만 연결되어 있다. 입력층에 있는 뉴런에는 공변수가 들어가고 출력층에 있는 뉴런에는 반응변수가 들어간다. 입력층과 출력층 사이에 있는 층은 은닉층(hidden layer)이라고 불리는데, 그 이유는 이들 층은 직접적으로 관측할 수 없기 때문이다. 입력층과 은닉층에는 절편에 해당하는 상수 뉴런(어떤 공변수에도 직접 영향을 받지 않는 뉴런임)이 포함된다. Figure 1은 입력층에는 두 개의 뉴런이 있고, 하나의 은닉층에는 3개의 은닉뉴런이 있고, 출력층에는 출력뉴런이 하나로 이것이 곧 반응변수 Y인 신경망의 예를 보여준다.

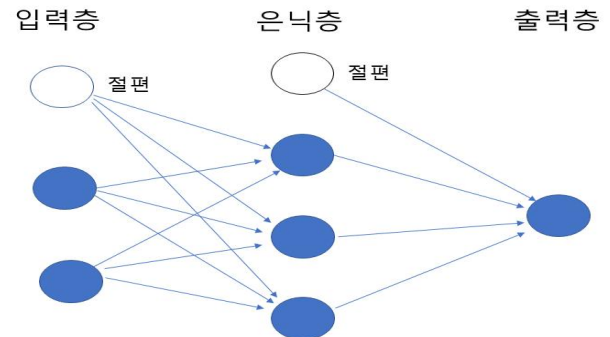


Figure 1. 입력뉴런이 2개, 은닉뉴런이 3개, 출력뉴런이 1개인 신경망

가장 간단한 형태의 다층 퍼셉트론은 은닉층이 없이  $n$ 개의 공변수를 가진 입력뉴런과 한 개의 출력뉴런만 갖는 퍼셉트론이다. 이 퍼셉트론은 다음과 같은 함수식으로 표현할 수 있다.

$$o(\mathbf{x}) = f(w_0 + \sum_{i=1}^n w_i x_i) = f(w_0 + \mathbf{W}^T \mathbf{x})$$

여기서  $w_0$ 는 절편,  $\mathbf{w} = (w_1, \dots, w_n)$ 는 각 입력뉴런에 붙는 가중치,  $\mathbf{x} = (x_1, \dots, x_n)$ 는 입력변수에 해당하며, 함수는 수학적으로 링크함수(link function)가  $f^{-1}$ 인 GLM의 함수로 볼 수 있다. 따라서 신경망에서 구해지는 가중치인  $w_i$ 는 GLM에서의 회귀모수와 같다고 볼 수 있다.

모델링을 유연하여 복잡한 문제를 해결하기 위해 은닉층을 추가할 수 있다. 하지만 은닉층을 하나만 추가해도 piecewise 연속적인 함수를 모델링하는 데에는 충분하다는 연구결과가 있다[13]. 은닉층이 하나만 있으며, 은닉층에 J 개의 은닉뉴런이 있는 경우 다층 퍼셉트론은 다음과 같이 함수값을 계산한다.

$$\begin{aligned} o(\mathbf{x}) &= f\left(w_0 + \sum_{j=1}^J w_j \times f\left(w_{0j} + \sum_{i=1}^n w_{ij}x_i\right)\right) \\ &= f\left(w_0 + \sum_{j=1}^J w_j \times f\left(w_{0j} + \mathbf{w}_j^T \mathbf{x}\right)\right) \end{aligned}$$

여기서  $w_0$ 는 출력뉴런의 절편,  $w_{0j}$ 는 j번째 은닉뉴런의 절편,  $w_j$ 는 은닉층의 j번째 은닉뉴런 가중치,  $\mathbf{w}_j = (w_{1j}, \dots, w_{nj})$ 는 입력층에서 은닉층 j번째 뉴런으로의 가중치,  $\mathbf{x} = (x_1, \dots, x_n)$ 는 입력변수들을 의미한다.

신경망에서 계산되는 식을 요약하면, 은닉층이나 출력층에 있는 모든 뉴런은 앞의 뉴런들  $z_0, z_1, \dots, z_k$ 로부터  $f(g(z_0, z_1, \dots, z_k))$ 를 구한다(뉴런  $z_0 \equiv 1$ 은 절편에 해당하는 상수임). 여기에서  $g: R^{k+1} \rightarrow R$ 는 적분함수를 나타내고  $f: R \rightarrow R$ 는 활성화함수를 나타낸다.

한편, 적분함수  $g(z_0, z_1, \dots, z_k) = w_0 z_0 + \sum_{i=1}^k w_i z_i = w_0 + \mathbf{w}^T \mathbf{z}$ 와 같이 통상 선형결합의 형태로 나타낸다. 활성화함수  $f$ 는 보통 로지스틱 함수( $f(u) = \frac{1}{1 + e^{-u}}$ )나 hyperbolic tangent 함수와 같이 유계(bounded) 비감소(nondecreasing) 비선형

이면서 미분 가능한 함수로 취한다.  $f$ 는 GLM의 경우에서와 같이 반응변수가 무엇인가에 따라 선택된다.

### 3. 연구방법론

신경망은 훈련과정 동안 학습 알고리즘에 의해 데이터를 모델에 적합시켜 가중치를 추정하게 된다. neuralnet 패키지는 지도형(supervised) 학습 알고리즘에 초점을 맞춘다. 이 알고리즘은 반응변수 값의 실제치와 예측치 간의 비교에 근거한 손실함수의 최소화를 위한 모수(가중치)들의 지속적인 조정을 특징으로 한다. 하지만 초기의 가중치들은 통상적으로 표준정규분포로부터 랜덤하게 추출된 값들로 한다. 이후 학습 과정 동안 다음 절차를 반복한다.

○ 신경망은 입력변수값  $\mathbf{x}$ 와 현재의 가중치에 근거하여 출력물  $o(\mathbf{x})$ 를 구한다. 만약 훈련과정이 아직 완료되지 않았으면 예측결과  $o$ 는 실제 관측된 결과  $y$ 와 차이가 날 것이다.

○ 잔차제곱합(Sum of Squared Errors, SSE)과 같은

오차함수 
$$E = \frac{1}{2} \sum_{l=1}^L \sum_{h=1}^H (o_{lh} - y_{lh})^2$$
 또는

cross-entropy

$$E = - \sum_{l=1}^L \sum_{h=1}^H (y_{lh} \log(o_{lh}) + (1 - y_{lh}) \log(1 - o_{lh}))$$

를 구해 실제 관측치와 예측치 간의 차이를 구한다. 여기서  $l = 1, \dots, L$ 은 관측치를 나타내고  $h = 1, \dots, H$ 는 출력노드를 나타낸다.

○ 모든 가중치들은 학습 알고리즘의 규칙에 따라 조정된다.

학습은 각 가중치에 대한 오차함수의 편미분 절대값이 일정한 임계값보다 작아질 때까지 계속된다. 이때 많이 쓰이는 학습 알고리즘은 탄력적 역전파 알고리즘이다. 탄력적 역전파 알고리즘은 기존의 역전파 알고리즘과 같이 오차함수의 지역적 최솟값을 구하기 위해 신경망의 가중치를 조정한다. 따라서 해(root)를 찾기 위해 각 가중치에 대한 오차함수의 기울기( $dE/dw$ )를 구한다. 특히, 가중치는 지역적 최솟값을 찾을 때까지 편미분

값의 반대방향으로 계속 수정한다. 따라서 만약 편미분 값이 음수이면 가중치를 증가시키고, 반대로 편미분 값이 양수이면 가중치를 감소시켜 지역적 최솟값에 도달하도록 한다. 여기에서 편미분은 chain rule을 활용하는데, 그 이유는 신경망에서의 함수는 입력층에서 은닉층, 은닉층에서 출력층으로 흘러가면서 기본적으로 총합(integration)과 활성화 함수들을 연속적으로 적용시킨 것이기 때문이다[14].

neuralnet에서는 역전파(backpropagation), weight backtracking이 있는 탄력적 역전파, weight backtracking이 없는 탄력적 역전파, modified globally convergent version 중에서 어떤 것이든 선택할 수 있다. 어떤 알고리즘이든 기울기의 반대 방향으로 학습률을 가중치에 더함으로써 오차함수를 최소화한다. 하지만 통상적인 역전파 알고리즘과 달리 탄력적 역전파에서는 각 가중치에 대해 별개의 학습률  $\eta_k$ (훈련과정에 변하는 값임)이 사용된다. 이렇게 함으로써 모든 훈련과정 및 전체 네트워크상에서 전반적으로 똑같은 학습률을 적용하는 문제에서 벗어날 수 있다. 추가적으로 편미분값의 크기 대신 그 기호만 가중치를 업데이트하는데 사용되며, 이렇게 함으로써 네트워크 전반에 걸쳐서 학습률이 동일하도록 한다. 따라서 가중치는 통상적인 역전파에서 이루어지듯

$$w_k^{(t+1)} = w_k^{(t)} - \eta \times \frac{\partial E^{(t)}}{\partial w_k^{(t)}}$$

와 같이 업데이트되지 않고 다음과 같이 업데이트된다( $t$ 는 반복횟수를 말하며  $k$ 는 가중치임).

$$w_k^{(t+1)} = w_k^{(t)} - \eta_k^{(t)} \times \text{sign} \left( \frac{\partial E^{(t)}}{\partial w_k^{(t)}} \right)$$

수렴의 속도를 높이기 위해 학습률  $\eta_k$ 의 값은 편미분 값의 기호가 그대로 유지되면 증가시키고, 편미분 값의 기호가 바뀌면 감소시킨다. 왜냐하면 부호가 바뀌었다는 것은 학습률이 너무 크기 때문에 최솟값을 지나쳤다는 것을 의미하기 때문이다. weight backtracking이란 바로 이전에 수행한 실행을 취소하고 다음 단계에서 가중치로 이전보다 더 작은 값을 넣는 것을 말한다. 이와 같은 weight backtracking을 사용하지 않는다면 알고리즘은 최솟값을 몇 번이고 넘을 것이다.

#### 4. neuralnet의 사용방법

neuralnet은 lm이나 glm과 같은 회귀분석에서 다루는 함수의 사용방법과 매우 흡사하다. 필수적인 인수로서 ‘반응변수~공변수의 합’, 공변수와 반응변수를 포함하는 ‘데이터셋’이 지정되어야 하고, 나머지 인수들은 모두 기본값으로 정의된다. 여기에서는 datasets라는 패키지가 제공하는 infert라는 데이터를 사용하여 신경망 분석을 어떻게 실시하는지 알아본다[15]. 이 데이터는 자연불임이나 유도불임의 경험이 있는 여성의 불임에 대해 살펴보는 환자-대조군 연구 데이터인데, 이 데이터를 살펴보는 이유는 통상적인 인공지능 방법론에 많이 활용되므로, R에서는 이런 데이터를 어떻게 분석하는지 보여주기 위함이다. 이 데이터는 불임인 여성 83명, 불임이 아닌 여성 165명의 총 248명에 대한 age(연령), parity(출산수), induced(유도불임 횟수), spontaneous(자연불임 횟수)에 대한 정보가 담겨있다. parity, induced 및 spontaneous는 0, 1, 2 등의 값을 갖는다.

신경망 훈련에 사용되는 neuralnet 함수는 복잡한 문제라면 그에 맞게 은닉층의 개수와 은닉뉴런의 개수를 많게 정할 수도 있다. 물론 은닉층이나 은닉뉴런을 추가하면 추가할수록 계산되는 함수는 더욱 복잡해진다. neuralnet에서는 사용자가 따로 정하지 않으면 은닉층은 하나이고 은닉뉴런 또한 한 개다. neuralnet 함수에서 중요한 인수는 다음과 같다.

- formula: ‘response variables ~ sum of covariates’의 형태로 기본값은 없다.
- data: formula에 지정된 변수를 포함하는 데이터프레임으로 기본값은 없다.
- hidden: 은닉층과 은닉뉴런의 수를 지정하는 벡터이다. 예를 들어 (3, 2, 1)은 3개의 은닉층으로 첫 번째 은닉층에는 3개의 뉴런이 있고, 두 번째 은닉층에는 2개의 뉴런이, 세 번째 은닉층에는 1개의 뉴런이 있는 경우를 나타낸다. 기본값은 1이다.
- threshold: 오차함수의 편미분 값이 일정 수준 이하일 때 반복수행을 중단하는 임계치로서 기본값은 0.01이다.
- rep: 훈련 절차를 수행하는 횟수로 기본값은 1이다.
- startweights: 미리 지정한 가중치의 초기값 벡터로서 기본값은 표준정규분포에서 랜덤하게 생성된 값들이다.
- algorithm: 알고리즘의 유형을 나타내는 문자열로

‘backprop’, ‘rprop+’, ‘rprop-’, ‘sag’ 및 ‘slr’이 있다. ‘backprop’는 통상적인 역전파를 나타내고, ‘rprop+’와 ‘rprop-’는 weight backtracking이 있는 경우와 없는 경우의 탄력적 역전파를 나타낸다. ‘sag’와 ‘slr’은 modified globally convergent algorithm으로 학습률이 다른 모든 것들(‘sag’는 최소 절대 도함수, ‘slr’은 최소 학습률)에 따라 변하는 것을 말한다. 기본값은 ‘rprop+’이다.

- err.fct: 미분 가능한 오차함수로, 문자열 ‘sse’와 ‘ce’가 사용된다. ‘sse’는 sum of squared errors(잔차 제곱합)이고 ce는 cross entropy를 나타낸다. 기본값은 ‘sse’이다.
- act.fct: 미분 가능한 활성화함수로서 문자열 ‘logistic’과 ‘tanh’은 각각 logistic 함수와 tangent hyperbolic을 나타낸다. 기본값은 ‘logistic’이다.
- linear.output: 논리형으로 만약 act.fct이 출력뉴런에 적용되지 않으면 linear.output은 TRUE로 놓는다. 기본값은 TRUE이다.
- likelihood: 논리형으로 만일 오차함수가 negative log-likelihood 함수이면 likelihood는 TRUE이어야 한다. 그러면 Akaike의 Information criterion과 Bayes Information criterion이 구해진다. 기본값은 FALSE이다.

neuralnet의 사용방법은 환자-대조군 상태의 반응변수와 age, parity, induced와 spontaneous의 네 공변수들 간의 관계를 모델링하면서 설명한다. 반응변수가 이진형이므로 활성화함수는 기본값인 logistic 함수를, 오차함수는 cross entropy를 선택한다. 또한 linear.output 항목은 FALSE로 두어 출력값이 활성화함수인 logistic 함수에 의해 [0, 1] 사이의 값이 나오도록 한다. 은닉뉴런의 숫자는 문제의 복잡도에 따라 결정되어야 한다. 예를 들어, 은닉층이 하나이고 은닉뉴런이 두 개인 신경망은 다음과 같이 훈련시킨다.

```
> library(neuralnet)
> str(infert)nn <- neuralnet(case~age+parity+
  induced+spontaneous, data=infert, hidden=2,
  err.fct="ce", linear.output=FALSE)
> summary(nn)
```

	Length	Class	Mode
call	6	-none-	call
response	248	-none-	numeric
covariate	992	-none-	numeric
model.list	2	-none-	list
err.fct	1	-none-	function
act.fct	1	-none-	function
linear.output	1	-none-	logical
data	8	data.frame	list
exclude	0	-none-	NULL
net.result	1	-none-	list
weights	1	-none-	list
generalized.weights	1	-none-	list
startweights	1	-none-	list
result.matrix	16	-none-	numeric

신경망 훈련과정과 훈련된 신경망에 대한 기본정보는 nn에 저장되어 있다. 예를 들어 nn에는 최초 가중치를 포함하여, 분석결과를 재현할 수 있는 정보를 포함한다. 그 중 중요한 값들은 다음과 같다.

- net.result: 전반적인 결과값을 포함한 리스트로서, 각 반복에서 신경망을 돌린 결과를 포함한다.
- weights: 각 반복마다 신경망에서 추정된 가중치를 포함하는 리스트이다.
- generalized.weights: 각 반복마다 신경망의 일반화된 가중치를 포함하는 리스트이다.
- result.matrix: 매 반복마다 오차, 실제 임계값, 필요 횟수, AIC와 BIC의 값(likelihood=TRUE라고 명시한 경우 계산됨)으로, 추정된 가중치를 포함하는 행렬이다. 매 반복마다 해당 값들이 열에 나타난다.
- startweights: 매 반복마다 초기 가중치를 포함하는 리스트이다.

신경망 분석 결과물의 요약은 R에서 nn\$result.matrix를 입력하면 나온다.

```
> nn$result.matrix
error 1.198159e+02
reached.threshold 9.626526e-03
steps 2.095900e+04
Intercept.to.1layhid1 -9.261958e+01
age.to.1layhid1 3.172570e+00
parity.to.1layhid1 -5.204143e+00
induced.to.1layhid1 -3.682237e+01
spontaneous.to.1layhid1 -7.515960e+02
Intercept.to.1layhid2 7.250137e+00
age.to.1layhid2 -1.425187e-01
parity.to.1layhid2 1.519384e+00
induced.to.1layhid2 -1.891283e+00
spontaneous.to.1layhid2 -2.862148e+00
Intercept.to.case 7.020969e+00
1layhid1.to.case -2.121874e+00
1layhid2.to.case -8.369042e+00
```

오차함수에 대한 모든 편미분 값의 절대치가 0.01(기본 임계값임)보다 작아질 때까지 훈련과정은 20959회 반복되었다. 추정된 가중치는 -751.6부터 7.25에 이르기까지 다양한 값을 취한다. 예를 들어 첫 번째 은닉층

의 절편은 -92.6과 7.25이며, 첫 번째 은닉뉴런으로 가는 공변수 age, parity, induced, spontaneous에 대한 가중치는 각각 3.17, -5.20, -36.82, -751.6이다. 원 데이터는 nn\$covariates와 nn\$response은 물론 nn\$data에도 저장되어 있다. 신경망분석의 결과물인 예측치  $o(x)$ 는 nn\$net.result에 나타난다. 이 사례에서 객체 nn\$net.result는 하나의 열에 나타나는데, 이는 반복횟수가 1번이기 때문이다. 만약 여러 번 반복했다면 여러 개의 열에 예측치가 나타날 것이다.

neuralnet은 일반화 가중치를 구해준다[16]. 일반화 가중치  $\tilde{w}_i$ 는  $i$ 번째 공변수의 변화가 로그오즈(log-odds)의 변화에 기여하는 크기로 정의된다.

$$\tilde{w}_i = \frac{\partial \log\left(\frac{o(x)}{1-o(x)}\right)}{\partial x_i}$$

일반화 가중치는 각 공변수  $x_i$ 의 영향력을 나타내므로 회귀모형에서  $i$ 번째 회귀모수와 같이 해석할 수 있다. 하지만 일반화 가중치는 다른 모든 공변수에 의해서도 영향을 받는다. 이들 가중치는 분산이 작다는 것은 선형효과가 있음을 의미하므로 해당 공변수의 영향력이 선형인지 보여준다. 이들 일반화 가중치는 nn\$generalized.weights에 저장되어 있다.

## 5. 결론

이 논문은 다층 퍼셉트론과 지도형 학습알고리즘에 대해 살펴보았다. 아울러 neuralnet이라는 패키지를 사용하여 공변수들과 반응변수 간의 함수적 관계를 어떻게 모델링하는지 살펴보았다. 본 논문에서는 독립변수는 물론 반응변수도 있는 경우에 적용되는 지도형(supervised) 학습 알고리즘에 초점을 맞추었다. 이 알고리즘은 반응변수 값의 실제치와 예측치 간의 비교에 근거한 오차함수의 최소화를 위한 모수(가중치)들의 계속적인 조정을 특징으로 한다. 이때 역전과 알고리즘을 이용해 오차함수의 가중치 편미분 값이 음수이면 가중치를 증가시키고, 가중치 편미분 값이 양수이면 가중치를 감소시키는 작업을, 각 가중치에 대한 오차함수의 편미분 절대값이 일정한 임계값보다 작아질 때까지 계속한다.

neuralnet 패키지에는 회귀분석에서와 같이 주어진 데이터에 대해 다층 퍼셉트론 모델을 훈련시키는 유연한 함수가 있다. neuralnet 패키지는 대부분의 매개변수들을 쉽게 조정할 수 있기 때문에 매우 유연한 패키지이다. 예를 들어, 활성화함수와 오차함수를 주어진 상황에 맞게 적절히 선택하고 나머지 매개변수들은 기본값으로 둘 수 있다.

본 논문에서는 자연불임이나 유도불임의 경험이 있는 여성의 불임에 대해 살펴보는 데이터에서 독립변수인 age(연령), parity(출산수), induced(유도불임 횟수), spontaneous(자연불임 횟수)가 불임에 미치는 영향력을 은닉층이 하나이고 은닉뉴런이 2개인 신경망분석을 통해 살펴보았다. 그 결과 신경망의 가중치들을 추정할 수 있었으며, 각 공변수의 영향력을 살펴볼 수 있었다.

지금까지는 주어진 데이터에 대해 신경망 모델을 추정하고, 어떤 공변수가 반응변수에 영향력이 있는지 살펴보았다. 추후 연구에서는 새로운 공변수 조합에 대해 반응변수를 어떻게 예측하는지 살펴보고자 한다. 아울러 지금까지의 연구에서는 각 가중치에 대한 점추정치를 구했는데, 다음에는 각 가중치에 대한 신뢰구간은 어떻게 구하는지 살펴보고자 한다. 마지막으로 본 연구는 지도형 학습모델에 대해 살펴보았는데 추후 연구에서는 비지도형의 경우 어떤 학습방법이 있는지 알아보고자 한다.

## References

- [1] McCullagh, P. and Nelder, J. (1983). Generalized Linear Models. Chapman and Hall, London.
- [2] Schmidhuber, J. (2015). "Deep learning in neural networks: An overview", Neural networks, 61, 85-117.
- [3] Emmert-Streib et al. (2020). "An introductory review of deep learning for prediction models with big data", Frontiers in Artificial Intelligence, 28.
- [4] Abiodun et al. (2018). "State-of-the art in artificial neural network applications: A survey", Heliyon, 4, 1-41.
- [5] Fritsch, S. and Günther, F. (2008). neuralnet:

- Training of Neural Networks. R Foundation for Statistical Computing, R package version 1.2.
- [6] Venables, W. and Ripley, B. (2002). Modern Applied Statistics with S. Springer, New York, fourth edition.
- [7] Limas, M. C. et al. (2007). AMORE: A MORE Flexible Neural Network Package.
- [8] Günther, F. and Fritsch, S. (2010). “neuralnet: Training of Neural Networks”, The R Journal, 2, 30-38.
- [9] Schiffmann, W., Joost, M. and Werner, R. (1994). Optimization of the backpropagation algorithm for training multilayer perceptrons. Technical report, University of Koblenz, Institute of Physics.
- [10] Kumar, A. and Zhang, D. (2006). “Personal recognition using hand shape and texture”, IEEE Transactions on Image Processing, 15, 2454 - 2461.
- [11] Günther, F., Wawro, N. and Bammann, K. (2009). “Neural networks for modeling gene-gene interactions in association studies”, BMC Genetics, 10:87, 1-14.
- [12] Bishop, C. (1995). Neural networks for pattern recognition. Oxford University Press, New York.
- [13] Hornik, K., Stichcombe, M. and White, H. (1989). “Multilayer feedforward networks are universal approximators”, Neural Networks, 2, 359 - 366.
- [14] Rojas, R. (1996). Neural Networks. Springer-Verlag, Berlin.
- [15] Trichopoulos et al. (1976). “Induced abortion and secondary infertility”, British Journal of Obstetrics and Gynaecology, 83, 645 - 650.
- [16] Intrator, O. and Intrator, N. (2001). “Interpreting neural network results: a simulation study”, Computational Statistics & Data Analysis, 37, 373 - 393.

백 재 욱(Baik, Jai Wook)



- 중앙대학교 응용통계학과 학사
- 미국 Virginia Polytechnic Institut and State University 통계학박사
- 현재 : 한국방송통신대학교 통계·데이터정보학과 교수
- 관심분야 : 통계학, 생산관리
- E-Mail : jbaik@knou.ac.kr