

신경망 모형의 초기가중치 최적화 방법에 관한 연구

조용준¹⁾, 이용구²⁾

요 약

신경망은 적용 다양성과 제약조건의 최소화성, 강력한 예측성, 범용성, 근사성 등 많은 장점을 지니고 있으나 초기 가중치의 할당에 따라 모델 생성의 Performance와 예측의 결과가 달라지게 되는 단점을 지니고 있다.

이런 신경망의 초기 가중치에 따른 단점을 보완하기 위해 통계적 알고리즘의 접목을 통해 Hybrid된 신경망 보완 알고리즘을 제시하고자 하였다. 논문을 위한 기본 가정으로 신경망의 가장 기본인 SLP 알고리즘을 바탕으로 활성화함수에 가장 일반적으로 사용되는 Sigmoid 활성화함수를 이용하였을 때, 초기 가중치로 기존의 임의의 난수 생성 방식이 아닌 통계적 로지스틱 회귀분석의 계수값(mle)을 제시하여 이를 초기치로 사용한 경우와 그렇지 않은 경우의 예측 정확성과 수렴의 Performance정도를 비교하여 가장 효과적인 초기치 방법을 제시하고자 하였다.

주요용어 : 신경망, 초기치, 국소최적해, 로지스틱, 데이터 마이닝

1. 서론

통계적 관점에서 보았을 때, 신경망 분석을 평가하는 중요 요소는 모델의 예측력과 모델 산출의 Performance라고 할 수 있다. 하지만, 신경망 알고리즘은 초기치의 할당에 따라 이 두 가지 요소가 달라지는 단점을 지니고 있다. 따라서 본 논문에서는 이런 단점을 보완하고자 신경망의 가장 기본인 퍼셉트론(Perceptron) 이론에 근거하여 보완알고리즘을 제시하고자 한다.

퍼셉트론 알고리즘은 중간단계에 활성화함수를 사용하게 되는데, 여기에 각 입력값에 초기의 임의의 가중치(Initial random value)를 할당하게 된다. 여기에서 임의의 초기 가중치가 어떻게 할당되는가에 따라 모델을 생성하게 되는 시간과 예측의 결과가 달라지게 됨으로 1회의 임의의 가중치 부여로는 완벽한 모델의 생성이 힘들게 된다. 이는 예측력과 관련된 국소 최적해(Local Optimum)와 대역해(Global Solution)의 문제와, 다른 하나는 모델 생성 과정과 관련된 수렴의 Performance 문제가 발생하게 된다.

따라서 본 논문에서 신경망에서 사용되는 초기치에 random value가 아닌 통계적 초기치 추정을 하고자 한다. 이를 위해 활성화함수를 사용하여야 하는데, 활성화함수 중에서 가장 일반적으로 사용되는 Sigmoid function(이하 : ft.)을 본 논문에서 사용하도록 하였다. 또한 이를 증명하기 위하여 신경망의 가장 근간을 이루는 퍼셉트론 이론 중에서 본 논문의 취지에 적합하도록 가장 기초적인 Single Layer Perceptron에 적용시키도록 하였다. 신경망의 초기치의 최적화를 위하여 데이터 타입에 따라 Logistic Regression 또는 Regression을 통해 추정된 독립변수의 계수 값(mle)을 초기치로 사용함으로써 신경망의 모델 산출 과정을 최소화시켜 Performance의 향상과 국소 최적해의 문제를 탈피하여 대역해에 가깝게 도달 할 수 있음을 제시하고자 한다. 또한 대

1) 중앙대학교 통계학과 박사과정, 한국능률협회 컨설팅

2) 중앙대학교 통계학과 교수

용량 데이터의 경우에 전체 데이터로 mle를 추정할 경우, mle의 산출과정에 많은 노력과 시간, 자원을 소모하게 되기 때문에 오히려 역효과를 낼 수 있으므로 데이터의 Sampling을 통하여서도 2가지 문제를 해결할 수 있는지를 실험을 통해 살펴보았다. 한마디로 요약하자면, 신경망에 대한 연구를 통해 신경망의 단점을 보완하고 기존 통계적 알고리즘의 접목을 통해 Hybrid된 신경망 보완 알고리즘을 제시하고자 한다.

2. 퍼셉트론의 개요

퍼셉트론의 정의는 학습을 통한 패턴인식이라고 명할 수 있다. 퍼셉트론은 본래 1950년대에 심리학자 F. Rosenblatt가 초기 퍼셉트론의 이론을 제안하였다. 따라서 일반적으로 퍼셉트론이라 함은 Single Layer Perceptron(SLP)이라고 본 논문에서는 말하도록 한다. 퍼셉트론에 대한 관심의 주된 이유는 어떤 타입의 패턴이 입력층에 주어졌을 때 이모델이 반응하게 하는 연결강도의 집합을 스스로 발견하는 자동적인 절차에 있다. 학습은 현재 주어진 입력 행렬에 대하여 현재의 각 연결강도를 조정함으로써 얻어질 수 있다. 퍼셉트론은 연결강도를 조정하기 위하여 반복적 학습방법을 사용한다. SLP[ROS58a, ROS58b, ROS61]는 위드로우(B. Widrow)의 ADALINE과 MADLINE등의 모델과 더불어 1960년대 초에 집중적으로 연구되었으며 이 모델들은 지금의 패턴인식 연구의 바탕이 되었다.

SLP의 구조는 입력층으로부터 잡음이 있는 변수들을 받아 일종의 선형결합을 통해 이때 결합에 사용되는 계수(coefficient)들을 연결강도(synaptic weights)라고 부르며, 비선형 활성화함수를 거쳐 출력층으로 결과를 산출하는 구조를 지니고 있다. 활성화함수는 입력값을 변환하고 이를 입력으로 사용하는 다른 마디로 출력하는 함수를 말한다. 다음의 그림은 SLP의 구조이다.

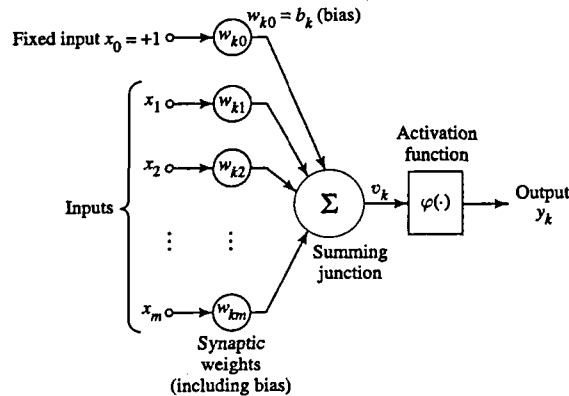


그림1 단층 퍼셉트론의 구조

2.1 활성화함수

그림1에서 보는 바와 같이 m 개의 입력을 받아 m 개의 연결강도의 벡터들과 곱해서 특정한 활성화함수(또는 전이함수(transfer function))를 거쳐 출력을 내게 된다. 즉, 활성화함수는 Output 결과를 결정하는 비선형 함수이다. 비선형 함수를 사용함으로써 결과산출을 더욱 Powerful하게 할 수 있다.

대표적인 비선형 함수로는 균등함수(Unique ft.), 임계논리 함수(Threshold logic ft.), 그리고 S자 형태의 탄젠트함수(Tangent ft.)와 시그모이드 함수(Sigmoid ft.), 가우스함수(Gaussian ft.)

등이 있다. 그 중에서 가장 많이 쓰이는 것은 시그모이드 함수(Farago 및 Lugosi, 1993; Lugosi 및 Zeger 1995; Devroye, Gyorf, 및 Lugosi, 1996)이다. 다음은 활성화함수 종류에 대한 각 그림이다.

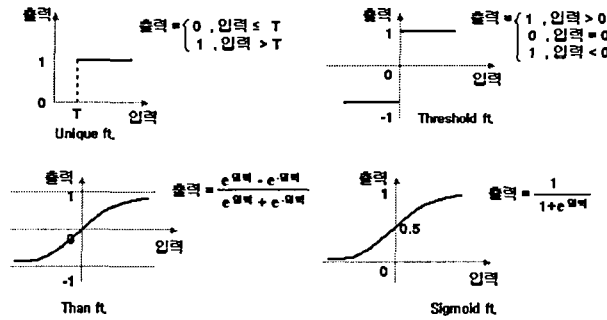


그림2 대표적 활성화함수의 종류

본 논문에서는 가장 대표적인 활성화함수인 시그모이드 함수(Sigmoid ft.)을 사용하여 데이터를 시뮬레이션 하였다. 시그모이드 함수(Sigmoid ft.)의 함수식은 다음과 같다.

$$\varphi(v) = f(v) = \frac{1}{1 + \exp(-av)} \quad a = \text{slope parameter}$$

$$V = v_k = \sum_{j=1}^m w_{kj} x_j + b_k \quad \begin{array}{l} w_{kj} \text{ 가중치} \\ b_k \text{ bias} \end{array}$$

식1 Sigmoid 함수식

위의 식은 통계학에서 흔히 말하는 로지스틱 함수임을 알 수 있다.

2.2 SLP 학습과정

연결강도를 조정하는 최초의 퍼셉트론 학습과정은 다음과 같은 학습과정을 거친다.

[단계 1] 연결강도들과 임계값을 초기화한다.

(식3.1)의 w_{1j} , b_k 에 임의의 값을 할당한다. ($1 \leq j < m$)

[단계 2] $v_i = \sum_{j=1}^m w_{ij} x_j$ 에 대입 시켜 실제의 출력값을 계산한다.

[단계 3] 실제 y 값과 $f(v_i)$ 값을 비교한다.

$$y - f(v_i) = \eta(w_{ij})$$

[단계 4] $\Delta w_{ij} = \sum_{j=1}^m \eta(w_{ij}) x_j$ 식을 통해 Δw_{ij} 를 구한다

[단계 5] 연결강도를 조정한다

$$w_{2j} = w_{1j} + \Delta w_{ij}$$

[단계 6] <단계2>로 가서 반복 수행하여 최상의 w_{ij} 를 구한다.

3. 로지스틱 회귀모형의 사용

신경망에서 로지스틱 회귀모형은 Sigmoid ft.의 하나로 S자 곡선을 그리는 비선형 함수이다. 본 논문에서 기본 가정으로 사용하는 활성화함수가 로지스틱 함수이다. 따라서 연결강도(synaptic weights)와 통계적 분석에서 사용되는 로지스틱 회귀모형의 추정된 계수값($\hat{\beta}$)과는 유사한 형태를 취하고 있다. 신경망에서는 연결강도의 초기치를 임의의 수(Random number)로 산출(Generate)하여 반복된 학습을 통해 최적의 연결강도를 산출하게 된다. 하지만 로지스틱 회귀분석을 통해 산출된 계수값을 초기치로 사용한다면, 그 초기치가 최대우도 추정량(Maximum likelihood estimate : 이하 mle)으로 볼 수 있다. 특히 mle의 특성 중에 하나가 Asymptotic unbiased와 invariant이기 때문에 반복하는 학습의 절대량을 감소 시켜 수렴의 Performance 정도를 월등히 향상시킬 수 있을 것이며 국소 최적해의 문제를 해결할 수 있는 대안으로 제시할 수 있을 것이다.

4. MLE와 난수와의 Simulation 비교

실험(Simulation)은 mle를 입력하여 테스트한 경우 1회와 대용량의 자료의 경우를 고려해 데이터 샘플링을 통해 산출된 mle를 입력하여 테스트 1회, 그리고 임의의 난수 생성의 경우 25회를 각각 시뮬레이션하여 비교 분석하였다. 난수는 [-100.00, 100.00] 사이의 소수점 두자리 값으로 생성하였으며, 실험의 효과를 위해 영단위, 일단위, 십단위의 난수를 적절히 배합하는 조작적 난수 배열을 사용하였다.

프로그램은 SQL을 사용하여 SLP 학습과정을 반복 학습하도록 프로그래밍 하였으며, 각 입력부의 독립변수들에 대한 데이터를 MS-SQL DB에 저장하였다. 초기치는 각 시뮬레이션마다 입력하여 결과를 산출하도록 구현되었다.

4.1 자료설명

자료 설명 : 쇼핑관련 구매분석 데이터

독립변수 : 구매금액, 구매수량

종속변수 : 구매여부(Y, N)

데이터 수 : 9230

Stopping Rule : 루핑 시 변화의 정도가 완전 0(Zero)이거나 루핑이 20000번 초과 시 정지 후 모델 산출

종속변수 특성 :

① 전체 데이터 : 비구매(43.6%), 구매(56.4%)

② 샘플링(30%) 데이터 : 비구매(52%), 구매(48%)

4.2 로지스틱 결과

로지스틱 회귀분석은 다음을 고려하여 분석을 실시하였다. 먼저 전체 데이터에 대한 계수값을 추정하여 이를 초기치로 선택하였으며, 향후 대용량의 데이터를 고려하였을 때, 전체 데이터에 대한 로지스틱 계수값을 계산하기 힘들기 때문에, 30%의 샘플링을 통한 계수값을 추정하여 이를 초기치로 사용하였을 때의 결과를 비교하도록 하였다.

① 전체 데이터에 대한 로지스틱 결과

다음의 도표는 전체 데이터에 대한 로지스틱 결과이다.

발점식에 포함된 변수							
		B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계(a)	X1	6.76	1.598	17.898	1	0	862.341
	X2	37.318	1.456	657.031	1	0	1.6108E+16
	상수	-2.008	0.05	1621.989	1	0	0.134

표1 전체 데이터에 대한 로지스틱 결과

위의 결과 x1, x2의 계수값(mle)으로 각각 6.76(β_1), 37.318(β_2) 임을 알 수 있으며, 상수(β_0)로 -2.008를 초기치로 사용하도록 한다.

② 샘플링(30%)에 대한 로지스틱 결과

다음의 도표는 샘플링(30%)에 대한 로지스틱 결과이다.

발점식에 포함된 변수							
		B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계(a)	X1	9.51	2.759	11.883	1	0.001	13495.572
	X2	28.734	2.406	142.592	1	0	3.0136E+12
	상수	-1.827	0.087	444.383	1	0	0.161

표2 샘플링(30%)에 대한 로지스틱 결과

위의 결과 x1, x2의 계수값(mle)으로 각각 9.51(β_1), 28.734(β_2) 임을 알 수 있으며, 상수(β_0)로 -1.827를 초기치로 사용하도록 한다.

4.3 실험결과

다음의 표3은 실험에 대한 결과 요약표이다.

	예측률	투표	초기치		
			w1	w2	blas
mle	74.84%	761	6.76	37.318	-2.008
mle(30%)	74.49%	566	9.51	28.734	-1.827
Random1	46.48%	20000	5.78	25.03	38.14
Random2	59.73%	20000	27.09	55.24	-59.03
Random3	73.53%	6471	-0.3	-6.75	-2.8
Random4	73.53%	2381	-0.23	4.78	-0.5
Random5	45.33%	20000	94.77	96.66	80.17
Random6	73.40%	6028	14.33	55.9	-4.77
Random7	73.92%	15738	22.36	-25.54	-1.24
Random8	45.59%	3635	11.56	19.24	2.25
Random9	66.44%	13172	-19.94	35.12	-0.25
Random10	48.39%	5729	9.32	-23.41	0.05
Random11	43.02%	5423	27.65	9.85	3.77
Random12	40.75%	20000	-4.65	2.38	32.30
Random13	48.40%	1653	0.56	0.45	0.23
Random14	48.40%	2063	0.13	-0.22	0.01
Random15	73.13%	9547	-45.66	11.27	-5.71
Random16	68.70%	2444	0.38	-0.54	-0.25
Random17	48.40%	1684	0.08	-0.73	0.31
Random18	71.43%	2537	0.4	-0.33	-0.33
Random19	48.88%	2101	-0.37	0.02	-0.03
Random20	41.44%	8905	5.41	5.98	9.28
Random22	51.39%	20000	22.33	64.58	78.74
Random22	51.39%	18857	4.31	2.84	19.98
Random23	51.11%	3496	13.21	-2.98	2.72
Random24	48.40%	8448	7.786	6.954	-4.256
Random25	51.39%	4347	-5.858	8.382	4.048

표3 실험1 종합 분석 결과표

표3에서 mle와 mle(30%)를 초기치로 사용한 경우가 금번 실험에서 임의의 난수를 사용한 것

에 비해 예측률이 가장 높은 것을 알 수 있다. 이는 국소 최적해의 문제를 단 1회의 시뮬레이션으로 해결할 수 있는 좋은 방법이라고 판단할 수 있을 것이다. 수렴의 정도에 있어서 역시, mle의 사용이 기타 임의의 난수를 사용하는 것에 비해 월등히 적은 횟수만에 수렴하는 것을 알 수 있다. 이는 mle의 방식이 일반적으로 모델 생성의 Performance를 가장 향상시키는 Powerful한 모델이라고 할 수 있을 것이다. (※ 루핑이 20000번 이상의 경우는 20000으로 표시)

5. 결론

위에서 살펴본 바와 같이 제한적인 실험이지만 로지스틱 회귀모형을 통한 가중치 추정의 방법은 위의 시뮬레이션 결과에 보았듯이 예측력과 관련된 국소 최적해(Local optimum)의 문제에 있어서 가장 높은 예측률을 보였으며, 이는 일반적으로 mle의 방식이 국소 최적해의 문제를 해결할 수 있는 근거를 제시한다고 할 수 있다. 또한 수렴의 Performance 정도에서의 결과에서도 보았듯이 임의의 난수에 비해서 가장 적은 횟수만에 수렴을 하는 것으로 나타났다. 이는 mle의 방식이 모델 생성의 Performance를 가장 향상시키는 Powerful한 모델이라는 근거를 제시한다고 할 수 있다. 또한 대용량 데이터의 경우에 전체 데이터로 mle를 추정할 경우, mle의 산출과정에 많은 노력과 시간, 자원을 소모하게 되기 때문에 오히려 역효과를 낼 수도 있으나 금번 실험에서 데이터 샘플링을 통해 산출된 mle의 추정값을 사용할 경우에도 임의의 난수를 생성하는 것에 비해 국소 최적해의 문제와 수렴의 Performance 결과를 충족할 수 있는 것으로 나타나 이를 보완 알고리즘으로 사용할 수 있는 근거를 제시하였다. 따라서 본 논문에서 제안하는 방식이 신경망에 있어서 사용되는 초기 가중치를 임의의 난수를 통하여 사용할 때, 여러 번 모델을 만들고 비교해보는 불편과 시간의 낭비, 예측률의 불안정성 등을 해결할 수 있는 방법의 근거를 제시하는 바이다. 따라서 향후 다른 데이터와 보편화 시킬 수 있는 수준의 랜덤 샘플링을 통해 시뮬레이션을 한다면 mle가 신경망의 가장 적합한 대역해와 비슷하게 갈 수 있음을 보임과 더불어 획기적인 모델링 산출 시간 및 메모리 단축에 기여 할 수 있음을 제시할 수 있을 것으로 예측된다. 이는 데이터 마이닝에서 가장 중요하게 생각하는 예측 정확도와 Performance 둘 다를 만족시키는 가장 적합한 Solution일 수 있다 라고 말할 수 있을 것이다.

참고문헌

- [1] Simon Haykin (McMaster Univ.) (1999), "Neural Networks: A Comprehensive Foundation", Prentice Hall, 10-18,
- [2] Alex Berson (1999), "Building Data Mining Applications for CRM", McGraw-Hill
- [3] 조용준, 허준, 최인규 (1999) 『데이터 마이닝 신경망 분석』, SPSS 아카데미
- [4] Michael J.A. Berry and Gordon Linoff (1997), "Data Mining Techniques: For MKT, Sales, and customer Support", Wiley Computer Publishing
- [5] Bishop, C.M. (1995), Neural Networks for Pattern Recognition, Oxford: Oxford University Press.
- [6] 김대수 (1992), 「신경망 이론과 응용」, 하이테크정보