

III. 머신 러닝을 활용한 개인 맞춤형 금융상품 추천 서비스

권병근^{1,*}, 윤창운², 김종원³

¹부산대학교 수학과, 산업수학 소프트웨어 연계전공, 산업수학센터 학부연구생

²부산대학교 수학과

³부산대학교 통계학과

이메일 : *house9895@pusan.ac.kr

1. Abstract

이 프로젝트는 인공지능망의 기본 구성 요소, 응용 분야 및 실용적 영향에 관해 연구를 진행했다. 가장 간단한 형태의 인공지능망인 퍼셉트론부터 시작하여 가중치 입력, 계단 함수 등에 대해 자세히 알아보았다. 다음으로, 다층 퍼셉트론의 구성과 은닉층을 활용한 예측과 오차 측정 과정을 탐구했다.

모델의 성능 향상을 위한 스케일링 기술에 대해서 알아보았고, ReLU 및 Hyperbolic Tangent와 같은 활성화 함수는 정확도와 손실에 미치는 영향을 평가했다. 과적합을 방지하고, 훈련을 간소화하는 기술로서 조기 중단의 개념을 설명하였다.

구체적인 응용 사례로, 사용자 입력을 기반으로 맞춤형 제품 추천을 제공하는 웹사이트를 개발했다.

2. Introduction

데이터 분석을 공부하면서 머신러닝을 활용한 데이터 분석 및 예측에 대해 배웠다. 부산대학교 산업수학센터에서 출판한 "고등 산업수학"이라는 책을 접하게 되었다. 이 책에서는 은행 데이터와 금융상품 추천에 관한 애용을 다루고 있었고 이 부분을 공부하고 실제로 개발해보고자 프로젝트를 진행하게 되었다.

머신러닝 기술을 활용하여 많은 양의 데이터를 분석하면, 새로운 고객의 정보를 입력하면 그에 맞는 금융상품을 추천할 수 있다. 이렇게 추천된 상품은 고객에게 맞춤형으로 제공되어 고객 만족도를 높일 수 있을 뿐만 아니라, 금융 회사에서도

맞춤형 상품을 제공하여 수익을 증대시킬 수 있을 것이다.

3. Main

(1) 사용한 데이터

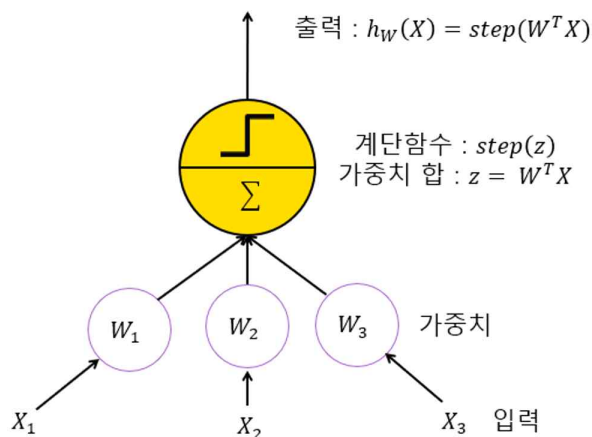
부산대학교 산업수학센터에서 출판한 "고등 산업수학"내에서 제공하는 데이터를 github에서 받아서 사용하였다. 이 데이터에는 특별 계좌, 온라인 계좌, 체크카드 로 Labeling이 되있는 데이터이다.

(2) ANN

1) 퍼셉트론(Perceptron)

퍼셉트론은 가장 간단한 인공신경망 구조로, 입력과 출력이 어떤 숫자이고, 각각의 입력 연결은 가중치와 연관이 있다. 입력의 가중치 합을 계산한 뒤 계산된 합에 계단 함수(step function)를 적용하여 결과를 출력한다. 가장 널리 사용되는 계단 함수는 헤비사이드 함수(Heaviside function)이다.

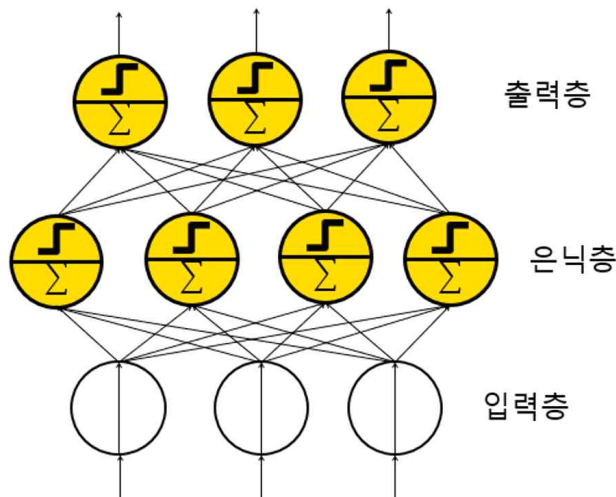
$$heaviside(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$



2) 다중 퍼셉트론(MLP, Multi-Layer Perceptron)

다중 퍼셉트론은 퍼셉트론을 여러 개 쌓아 올려 일부 제약을 줄일 수 있는 인공신경망이다. 입력층 하나와 은닉층(hidden layer)이라 불리는 하나 이상의 TLU 층과 마지막 출력층으로 구성된다.

다중 퍼셉트론의 진행 과정은 다음과 같다. 먼저 각 훈련 샘플에 대해 알고리즘이 먼저 예측을 만들고(정방향 계산) 오차를 측정한다. 다음 역방향으로 각 층을 거치면서 각 연결이 기여한 정도를 측정한다. (역방향 계산)이 오차가 감소하도록 가중치를 조정한다. (경사 하강법)



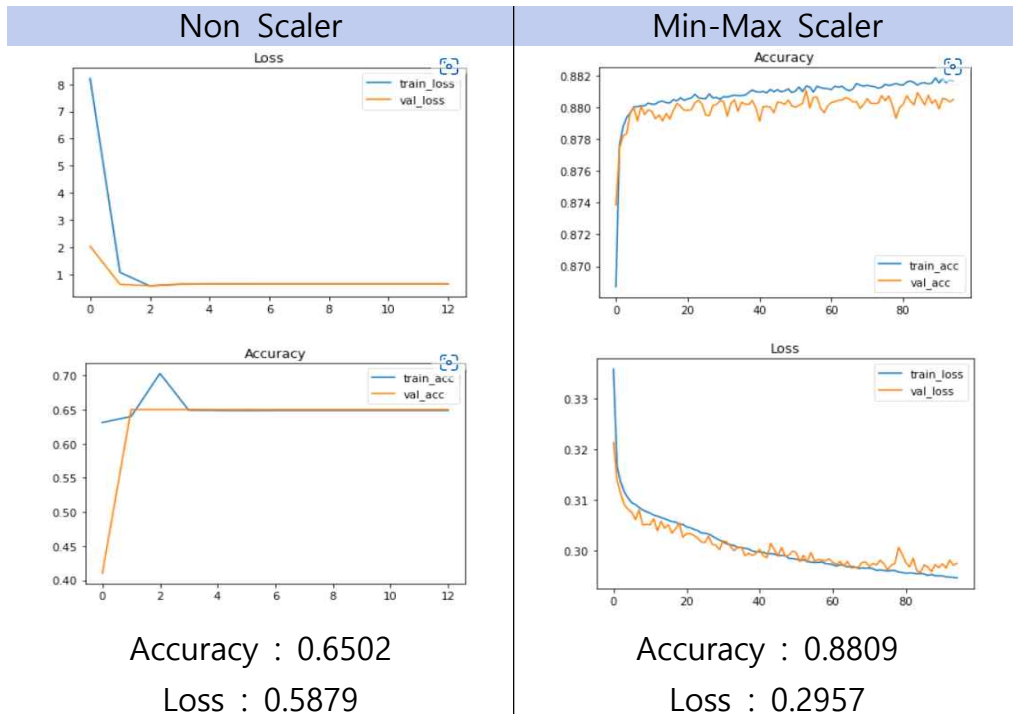
3) 인공신경망 (ANN, Artificial Neuron Network)

인공신경망은 뇌에 있는 생물학적 뉴런의 네트워크에서 영감을 받은 머신러닝 모델이다. 이 신경망은 이미지 분류, 음성 인식, 추천 알고리즘 등 다양한 분야에서 활용된다. 현재 사회에는 신경망을 훈련하기 위한 데이터가 많아졌고, 컴퓨터 하드웨어가 크게 발전했으며, 훈련 알고리즘이 향상되었기 때문에 앞으로 더욱더 우리의 생활에 큰 영향을 줄 것이다.

4) Scaling

스케일링이란 데이터의 값의 범위를 조정하여 모델의 성능을 개선하는 일종의 전처리 방법의 하나이다.

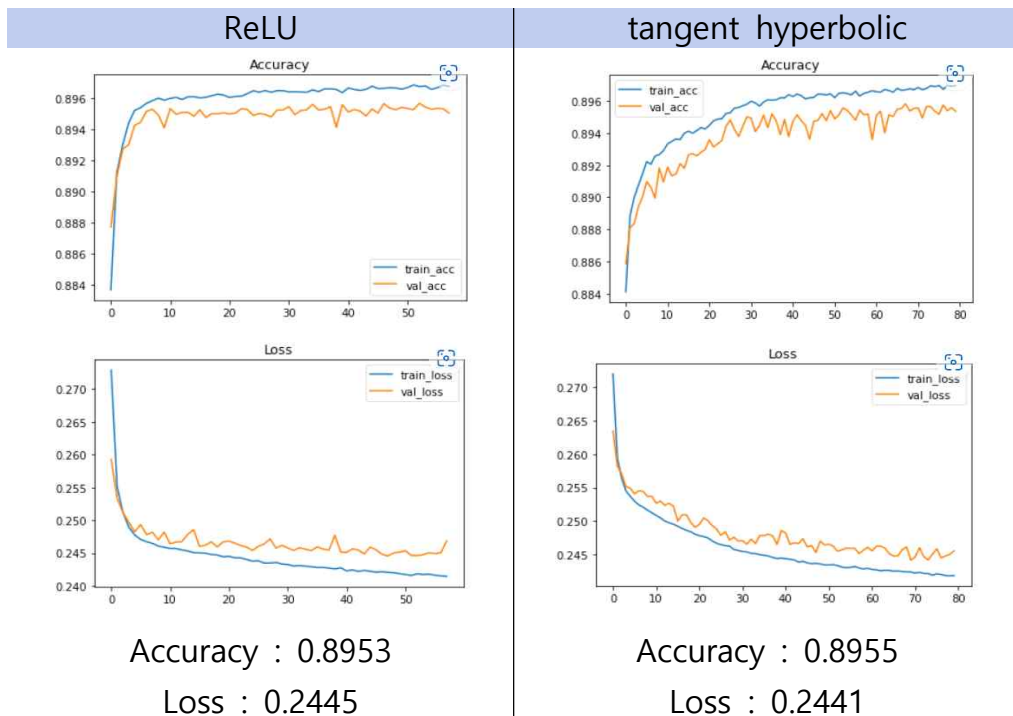
Min-Max Scaler는 데이터의 범위를 [0, 1]로 조정하는 스케일링 방법이다. 데이터의 상대적인 크기 차이가 클 때 유용하고, 이상치가 있는 경우에 유용하다는 장점이 있다. 하지만 모든 값이 0보다 큰 데이터에는 적용하기 어렵고, 최솟값과 최댓값이 조금만 변화해도 스케일링 결과가 크게 달라질 수 있다는 단점이 있다.



Min-Max Scaler 기법을 통해 학습하고 Accuracy와 Loss를 비교 해 보았다. 스케일링하지 않았을 때 비해서 성능이 매우 좋아졌다.

5) 활성화 함수(Activation function)

선형 변환은 여러 개를 연결하여도 선형 변환이다. 따라서 층 사이에 비선형성을 추가하지 않으면 아무리 층을 많이 쌓아도 하나의 층과 같아진다. 그래서 더 복잡한 문제를 풀기 위해서 사용한다. 따라서 ANN 신경망의 층을 쌓는 과정에서 Relu function과 tangent hyperbolic 함수를 사용하며 성능을 비교해보았다. Relu function은 계산 속도가 빠르다는 장점이 있고, 출력에 최댓값이 없다는 점이 경사 하강법에 있는 문제를 해결해준다. tangent hyperbolic 함수는 훈련 초기에 각 층의 출력을 원점 근처로 모으는 경향이 있다는 장점이 있다.



두 가지 활성화 함수를 비교해보았을 때 accuracy는 tangent hyperbolic이 loss는 ReLU 함수가 좋은 성능을 보여주었다. 금융상품 추천 서비스는 정확하게 추천해주는 것도 중요하지만 잘못 추천해주는 것을 줄이는데 더 중요하다고 판단하여 ReLU 함수를 활성화 함수로 선정하였다.

(3) 조기 종료(early Stopping)

조기 종료는 머신러닝 모델의 학습을 일찍 멈추는 기술이다. 일반적으로 모델 학습 시, 일정한 epoch 수만큼 학습하고 끝내게 되는데, 조기 종료 기술을 적용하면 validation loss가 더 이상 개선되지 않을 때 학습을 종료시킨다.

모델의 학습 시간을 단축할 수 있고, 모델이 과대 적합 (overfitting) 되지 않도록 방지할 수 있다는 장점이 있다.

```
from keras.callbacks import EarlyStopping

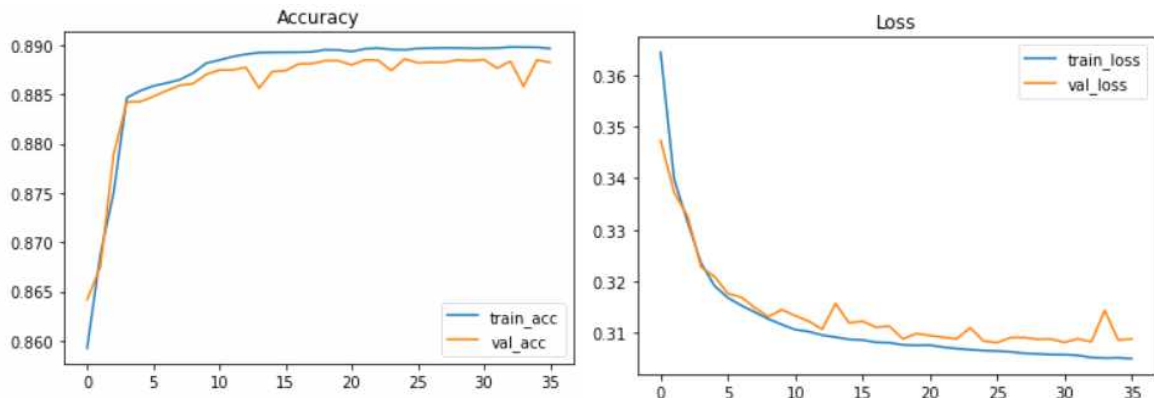
# EarlyStopping 콜백 정의
early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)

# 모델 컴파일
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# 모델 학습
history = model.fit(X_train_scaled, y_train, validation_data=(X_test_scaled, y_test),
                    epochs=100, batch_size=64, callbacks=[early_stopping])
```

(4) 최종 모델

Layer(type)	Params #
Dense(64, relu)	1856
Dense(32, relu)	2080
Dense(32, relu)	1056
Dense(1, sigmoid)	33
Total Params : 5025	



최종적으로 accuracy : 0.8887, loss : 0.3039로 모델링을 완료하였다.

(5) 웹 구현

제작한 모델을 바탕으로 실제로 값을 입력하였을 때 상품을 추천해주는 웹사이트를 구현하게 되었다. 해외 거주 여부(No : 0, Yes : 1), 신규 고객(No : 0, Yes : 1), 고객 활성화 상태(No : 0, Yes : 1), 성별(여자 : 0, 남자 : 1), 거래 누적 기간(0 ~ 249개월), 가구소득(1,300 ~ 10,000,000 천원), 나이(19~120), 지역 코드(0~51), 고객 구분(0~2)을 입력하면 체크카드, 온라인 계좌, 특별 계좌를 추천하는지를 출력하게 했다.

수학과랑통계과는 가족이조

Personalized financial product recommendation project using ANN

해외 거주 여부(Yes : 1, No : 0)	신규고객(yes:1, no:0)	고객 활성화 상태(yes:1, no:0)	
0 - +	1 - +	0 - +	
성별(남자:1, 여자:0)	거래누적기간(0 ~ 249)	가구 소득(1,300 ~ 10,000,000)	
1 - +	0 - +	50000 - +	
나이(19 ~ 120)	지역코드(0 ~ 51)	고객 구분(0 ~ 2)	
28 - +	28 - +	1 - +	
추천 여부	체크 카드	온라인 계좌	특별 계좌
	추천	추천	x

4. Summary & Conclusion

이 프로젝트는 인공지능망(ANN)에 관해 연구를 진행하였으며, 퍼셉트론과 같은 기초 개념부터 다층 퍼셉트론과 같은 상위 개념까지 연구하였다. 스케일링, 활성화 함수 및 조기 중단 기술을 적용하여 실용적인 모델을 구축하였다. 이는 데이터 활용 가능성, 하드웨어 개선 및 향상된 학습 알고리즘에 의해 널리 사용될 것이다.

결론적으로 이 프로젝트는 알고리즘과 실제 적용의 결합으로써, 인공지능망은 혁신적인 기술적 발전을 이룰 것이다. 상품 추천 웹사이트를 통한 실용적인 구현은 이러한 연구 결과의 실제적인 적용을 보여준다.

5. Utilization Plan & Expected Effect

신규 고객 등록 때, 머신러닝 기술을 활용하여 개인적인 주관에 개입되지 않고 정확한 예측을 제공할 수 있다. 이러한 기술을 활용하면 주관적인 판단이 개입되지 않아 더욱 효과적인 상품 추천이 가능해지며, 직원들은 개별적인 고민이나 검색에 걸리는 시간을 줄일 수 있을 것이다. 이를 통해 고객에게 정확하고 신속한 서비스를 제공하고, 금융 회사는 맞춤형 상품을 효과적으로 제공하여 수익을 극대화할 수 있을 것이다.

또한 신규 고객 등록할 때마다 데이터가 추가되면, 머신러닝 모델의 정확도가 점차 개선되어 시간이 지날수록 더욱 정교한 예측이 가능해질 것으로 기대된다. 이러한 업데이트는 금융 회사에서 맞춤형 상품을 제공하는데 매우 중요하며, 고객이 요구하는 다양한 상황에 대응하기 위해 필수적인 과정이 될 것이다.

6. Future Work

프로젝트 기획 단계에서 고안했던 일부 결과물을 얻는 데 성공하였지만, 여전히 과적합 문제를 완벽하게 해결하지 못하고 있다. 추가적인 공부와 모델 개선, 그리고 다른 머신러닝, 딥러닝 모델을 활용하여 더 나은 결과물을 만들 수 있다고 기대되며, 이를 통해 많은 사람이 이해할 수 있는 결과물을 창출할 것이다.

7. Github

https://github.com/KwonBK0223/Personalized_financial_product_recommendation_project_using_machine_learning

8. Reference

- * <https://github.com/Industrial-Mathematics/Bank-Textbook>
- * 김상일, (고등학교)산업 수학, 천재교과서