

인공신경망 실습 보고서

Y2023011 윤요섭

I. 자료 및 분석 방법

본 보고서는 bankruptcy.xlsx 데이터를 이용하여 Neuroshell2® R4.2를 통해 실습과제를 수행하였다. bankruptcy.xlsx 데이터 변수와 이에 대한 설명은 다음과 같다.

변수명	변수설명
CLASS	데이터셋 구분 (1=학습, 0=검증)
X13_1	금융비용대부채비율
X18_1	매출원가비율
X36_1	자기자본비율
X42_1	금융비용부담율증가분
X43_1	매입채무회전율
X74_1	지급여력도
X76_1	분식계수2
X98_1	기업경상이익율
X107_1	현금흐름7대전기총부채
X118_1	총자산변동계수
X129_1	자산대비금융비용증가율
X130_1	자산대비영업외비용증가율
X133_1	매출원가율*매출원가평균증가비
X149_1	매출대비재고자산증가율
X157_1	총자본회전율*매출액증가비
Y	부실여부(1=부실, 0=건전)

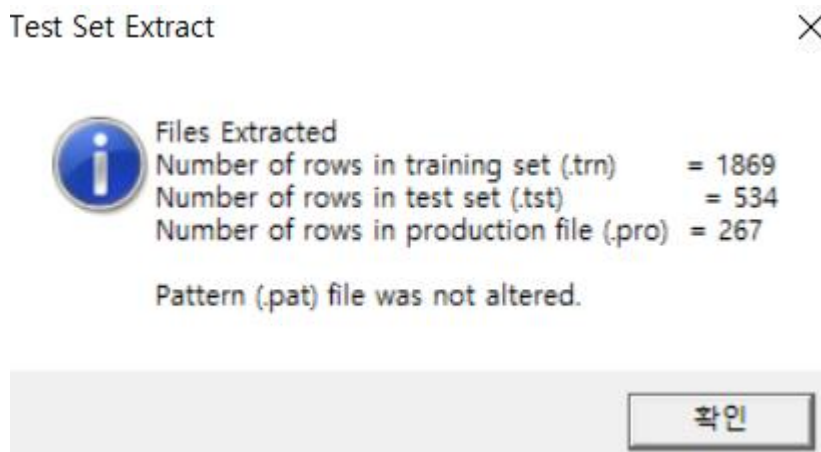
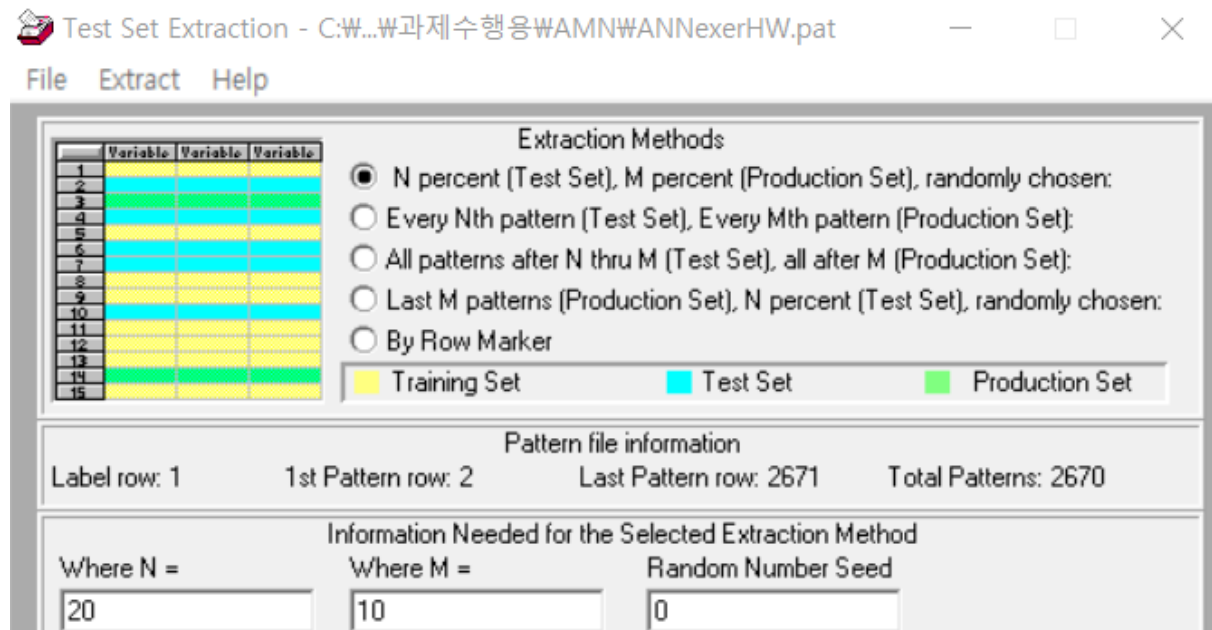
bankruptcy.xlsx 데이터를 통해 인공신경망 실습 실시하였으며, 이 때, CLASS 변수가 1인 경우 학습용 데이터로 0인 경우 검증용 데이터로 활용하였다. 종속변수로는 Y(부실여부)로 설정하였으며, CLASS변수와 Y변수 외의 변수를 독립변수로 설정한 뒤, 학습을 진행하였다.

II. 분석 과정과 결과 및 해석

본격적인 실습에 앞서서 학습용 데이터(T), 검증용 데이터(P), 테스트용 데이터(V)로 나누고 실습을 진행하였다.

Test Set Extraction의 경우 CLASS 변수를 기준으로 학습용 데이터와 테스트용 데이터로 나누려고 하였으나, By Row Marker로 설정 후, 진행 시 테스트 파일이 생기지 않는 오류로 인해서, 20%를

테스트 셋으로 랜덤으로 선택하였다. 검증용 데이터의 경우, 진행하여 학습용데이터의 과대 적합 문제를 방지하기 위해 10%로 설정하였다.



1869 Row의 학습용 데이터와 267 Row의 검증용 데이터, 534 Row의 테스트 데이터를 설정 한 뒤, Architecture and Parameters의 경우 3 Layer BackProp Network를 선택하였다.

Slab 1, Slab 2, Slab 3의 뉴런 수의 경우, Slab 1은 15개의 Neurons로, Slab 3은 1개의 Neurons으로 결정했으며, Activation function은 logistic을 설정하였다.

Slab 2(은닉층)의 경우 은닉층 노드의 최적 개수는 경험적으로 결정하며 일반적으로 $n/2$ 에서 $2n$ 사이(n 은 입력층 노드 수 + 출력층 노드 수)로 결정했다. 구체적인 뉴런의 수는 여러가지 실험을 통해 구체적으로 정하는 것이 좋지만, chatGpt가 추천해준 8개의 뉴런을 사용하기로 결정하였다.

따라서, 뉴런 수를 다음과 같이 설정해보십시오:

- 입력 층: 15개의 뉴런
- Slab 1: 15개의 뉴런
- Slab 2: 8개의 뉴런
- Slab 3: 출력 카테고리 수와 동일한 수의 뉴런

최적의 뉴런 수를 결정하고, 학습률을 높이기 위해서는 Learning rate, Weight Updates에서 Momentum 설정, epoch 횟수 등 다양한 실험을 시도해보고, Gradient Vanishing Problem, Overfitting 발생을 막기 위해서 Activation Function 설정부터, Batch Normalization을 고려하면서 진행해야 하나, 이번 실습에는 기본값으로 접근하기로 결정했다.

기본적인 세팅을 끝낸 후, 학습 결과는 다음과 같았다.

Learning: C:\w...WBITW1-1학기\비즈니스애널리틱스\과제수행용\WAMN\WANNex... — □ ×

File Run Options Help

Training Graphics

Training Set Average Error

Epochs Elapsed

Test Set Average Error

Intervals Elapsed

Error Factor Ranges

Training Set Patterns

Error Factor Ranges

Test Set Patterns

There are 1869 training patterns.

☐ learning events: 226400

☒ learning epochs: 121

☒ last average error: 0.0642819

☒ minimum average error: 0.0620040

☐ epochs since min. avg. error: 48

There are 534 test patterns.

Calibration interval (events): 200

☐ last average (internal) error: 0.0716146

☐ minimum average error: 0.0690944

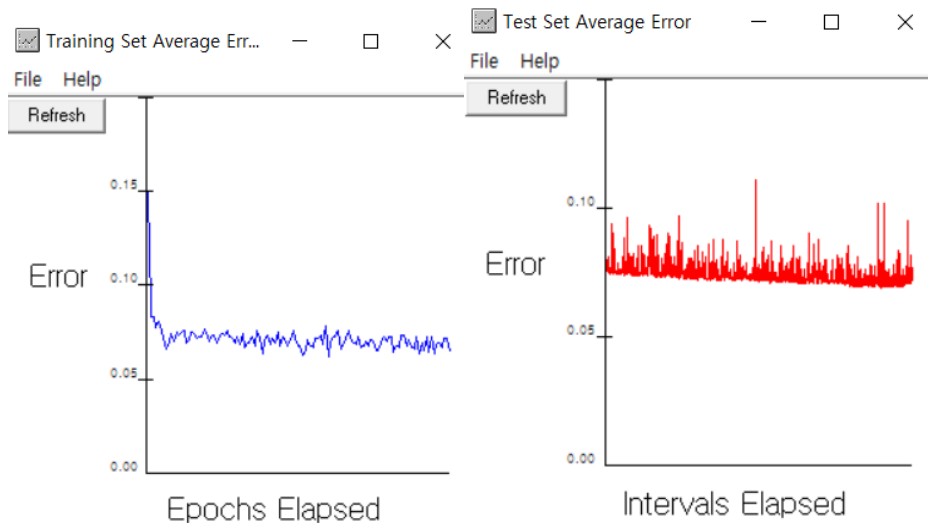
☐ events since min. avg. error: 20000

Automatically Save Training on

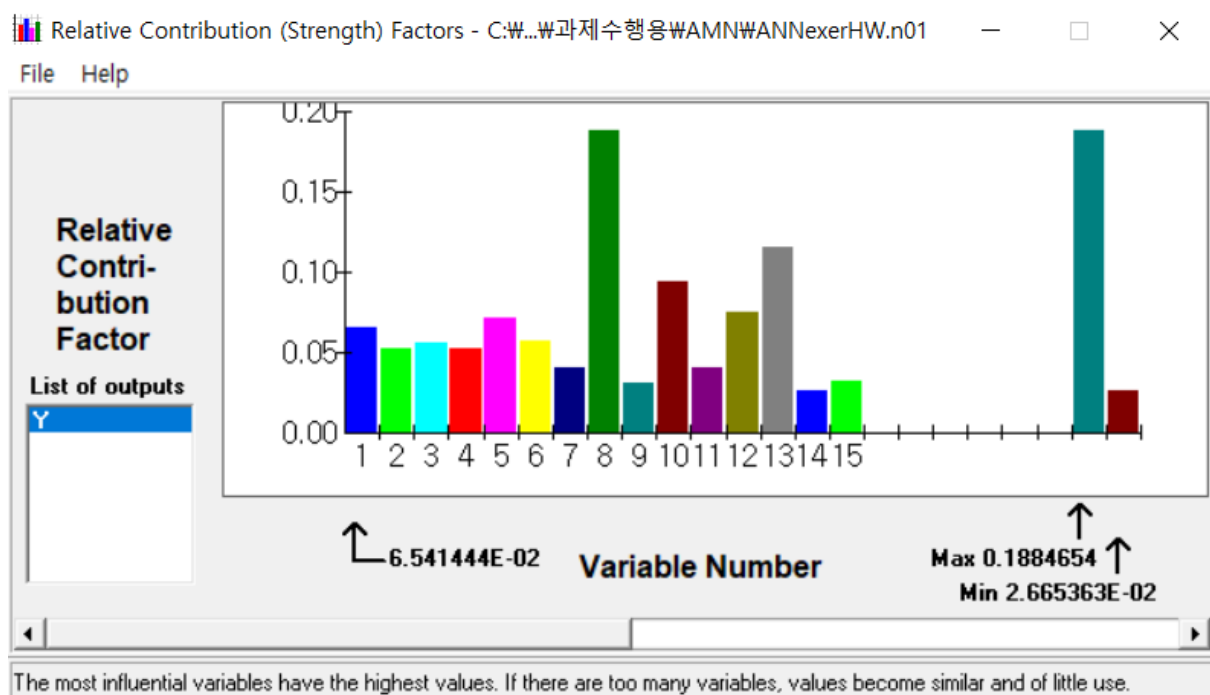
☐ best training set ☒ best test set ☐ no auto save

Training Time (hhh:mm:ss)

000:00:03



121번의 학습 과정을 거친 결과 Training Set Average Error의 경우 처음에는 학습이 반복되면서 Error가 줄어 들었지만, 그 이후 큰 차이를 보이지 않았고, Test Set Average Error의 경우 미세하게 감소하는 것 같지만 큰 차이는 없는 것으로 나타났다.



독립변수의 영향력 분석 결과 X98_1(기업경상이익율), X_133_1(매출원가율*매출원가평균증가비)의 영향력이 큰 것처럼 나타났다.

Network Processing C:\...₩과제수행용₩AMN₩ANNexe... — □ ×
File Run Help

☒ **Compute R squared, etc. (actual outputs must be in the file)**
☒ **Include actuals in .OUT file (actual outputs must be in the file)**
☒ **Include in .OUT file actuals minus network outputs**
☐ **Write neuron activations to file for slab number:**
☐ **Set highest output to 1, others to 0 (use when outputs are categories)**

Input file name: C:\...₩과제수행용₩AMN₩ANNexerHW.pat
Patterns processed:

Output:	C1
R squared:	0.5855
r squared:	0.5858
Mean squared error:	0.104
Mean absolute error:	0.208
Min. absolute error:	0
Max. absolute error:	1.000
Correlation coefficient r:	0.7654

OUT 결과 파일을 해석하기 위해, Apply to File을 실행한 결과, R squared가 높지 않은 것으로 나타났다.