

# SW특기자전형 포트폴리오

소프트웨어특기자전형을 위해 제출하는 포트폴리오는 그 동안의 다양한 교육활동을 통해 소프트웨어와 관련하여 자신이 성취한 성과와 지식, 경험, 능력을 잘 나타내줄 수 있는 실적을 세 가지로 한하여 기술합니다 (실적당 서식4-1, 4-2 각 1매씩 작성, 양면출력 금지, 글자크기 11pt, 글자체 : 휴먼명조 ).

학교명	명호고등학교	성명	신 병 근	실적번호	3
실적명	Multiple Regression을 이용한 생활 형태별 성적 예측 인공 신경망				

## ● 동기

머신러닝 원리의 이해에 필요한 수학적 지식들을 배운 고등학교 3학년부터 본격적인 머신러닝 공부를 시작할 수 있었는데, 공부를 하던 중 회귀분석 분야에 관심이 생겨 직접 어떤 사례에서 선형 관계를 모델링해보고 싶었습니다. 그런데 그때 마침 시험기간이 끝나 시험성과와 생활 태도에 대한 좋은 설문 데이터를 얻을 수 있었고 파이썬도 충분히 활용할 수 있을 만큼 적응되어 이 프로젝트를 시작하게 되었습니다.

## ● 주요 기능 및 활용도구

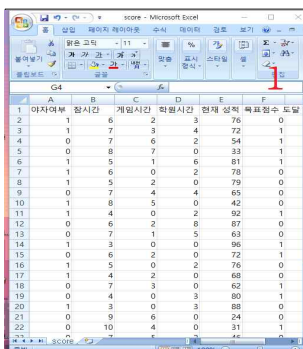
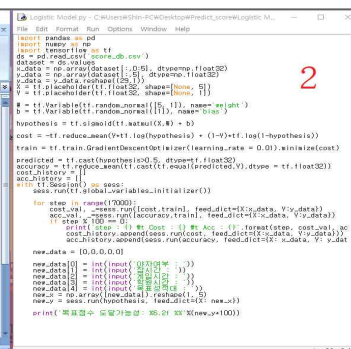
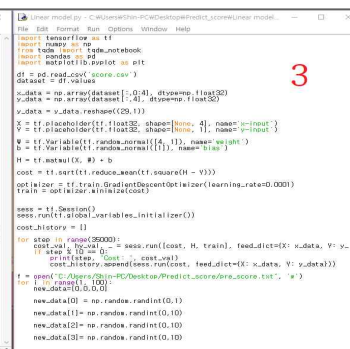
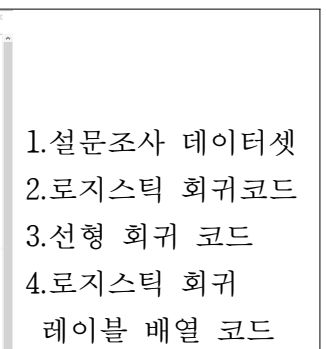
1. 친구들 30명을 대상으로 개인별 생활 형태와 성적을 설문하고 데이터 통계 라이브러리를 이용해 더 나은 특성을 추출해서 데이터셋을 구성하였고, 특정 성적대의 과적합을 방지하기 위해 상위권, 중위권, 하위권을 골고루 섞었습니다.
2. RMSE(평균 제곱근 오차), 경사하강법을 활용해 생활태도에 따른 점수를 예측하는 다중 선형 회귀를 구현하였고 오차값을 기준으로 정확도를 산출하였습니다.
3. 만들어진 선형 모델로 각 개인별 생활 형태에 따른 목표 점수 도달률을 재설문하고 데이터셋을 확장시켜, 부가적으로 자신이 목표한 성적대에 도달할 수 있을지 가능성도 계산해주는 로지스틱 회귀 모델을 구현하였습니다.

모든 코드는 Python을 이용하여 작성하였고 데이터 통계 라이브러리인 Pandas, Matplotlib, 수치 연산 모듈인 Numpy, 머신러닝 모듈인 Tensorflow 오픈소스 라이브러리를 사용하였습니다.

# <https://pandas.pydata.org/>, <https://matplotlib.org/>, <http://www.numpy.org/>

# <https://www.tensorflow.org/>

## ● 최종 산출물

1. 설문조사 데이터셋

2. 로지스틱 회귀코드

3. 선형 회귀 코드

4. 로지스틱 회귀 레이블 배열 코드

5. 3d그래프 코드

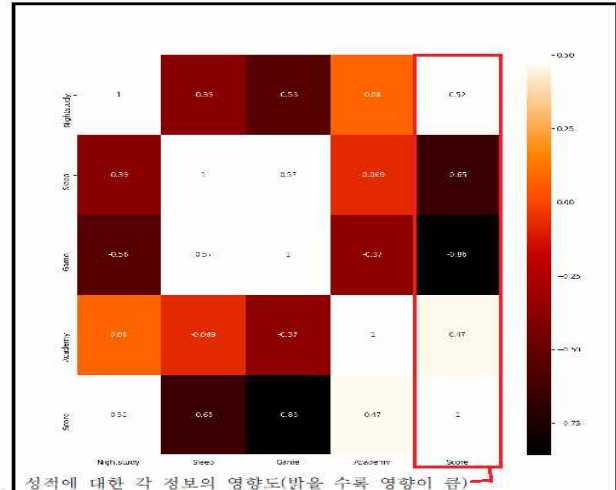
6. 데이터 분석코드

## ○ 1차 선형 회귀 모델

실문 데이터셋 [야자 여부, 잠, 게임, 학원, 성적]

	Nightstudy	Sleep	Game	Academy	Score
count	30.000000	30.000000	30.000000	30.000000	30.000000
mean	0.500000	5.200000	2.866667	2.033333	63.200000
std	0.500000	1.936337	2.161311	2.108621	20.323226
min	0.000000	3.000000	0.000000	0.000000	15.000000
25%	0.000000	5.000000	1.250000	0.000000	52.500000
50%	0.500000	5.000000	2.500000	2.000000	66.500000
75%	1.000000	7.000000	4.750000	3.000000	77.500000
max	1.000000	10.000000	7.000000	8.000000	96.000000

Seaborn, Matplotlib 통계 라이브러리를 사용해 각 정보의 상관관계 표시



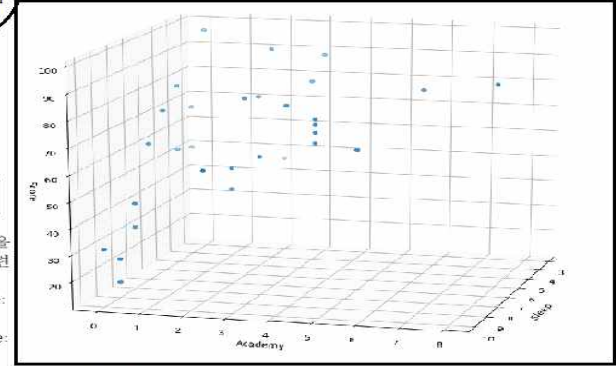
### < 선형 회귀 최종 결과 >

새로운 데이터를 넣었을 때의 성적 예측



Linear Regression

이 데이터들을 바탕으로 훈련 시작  
오차평균함수: RMSE  
Learning\_rate: 0.0001

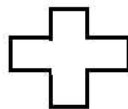


데이터의 전체적인 구조를 살펴보기 위해 성적에 대한 영향도가 높은 2가지(학원, 잠)를 주러내서 3d 형태 그래프로 분석

## ○ 2차 로지스틱 회귀 모델

Logistic Regression

만들어진 선형 모델로 생활 태도에 대한 예측점수 배열을 생성



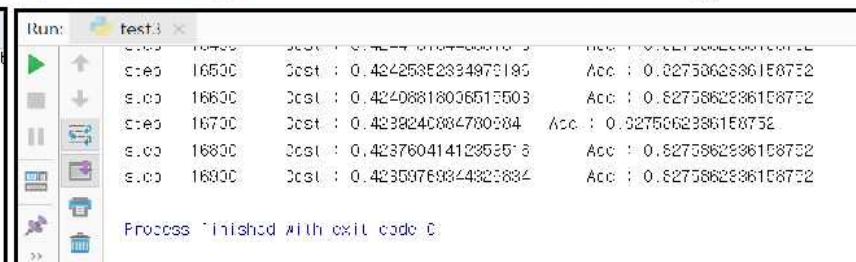
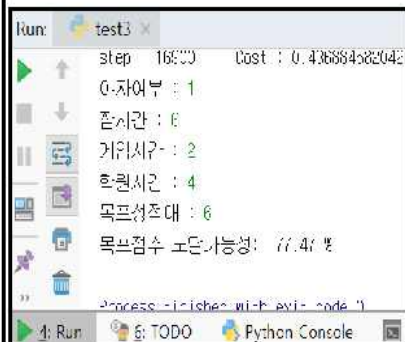
목표로 했던 성적대  
실문 데이터



예측 점수가 목표 점수보다 크면 성공(1), 아니라면 실패(0)을 담은 배열

Epochs = 170, Learning\_rate = 0.01, Acc : 0.8275, Loss : 0.4235

로지스틱 회귀



야자여부, 잠시간, 게임시간, 학원시간, 목표성적대 (X)0점대를 넣었을 때 목표 점수 도달 가능성을 예측해냄.

실제 시험 결과와 비교했을 때는 84퍼센트대의 정확도로 시뮬레이션 정확도와 비슷함.

