

| 딥러닝 실습 – Logistic Regression & Linear Regression



명지대학교
MYONGJI UNIVERSITY

과제 제출 유의사항

이번주 과제는 **.ipynb** 로 제출 바랍니다.

파일에 모든 과제가 다 들어갈 수 있도록 한 개 파일로 작성 (comment 로 구분 필요)

과제는 스스로 하기 바랍니다. (복사 적발시 보여준사람 및 본 사람 모두 -300%)

- 과제 제출에 대한 점수를 차등하여 부여할 수 있는 아이디어 받습니다.!

Linear Regression : 15점

1. 남, 여 데이터로 나누기: 5점
2. 1번 그림 그리기 및 분석 : 5점
3. 2번 그림 그리기 및 분석 : 5점

Logistic Regression :10점

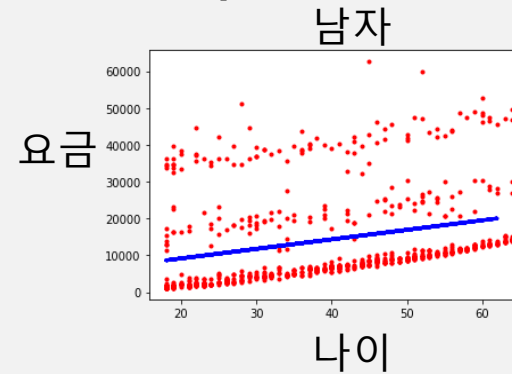
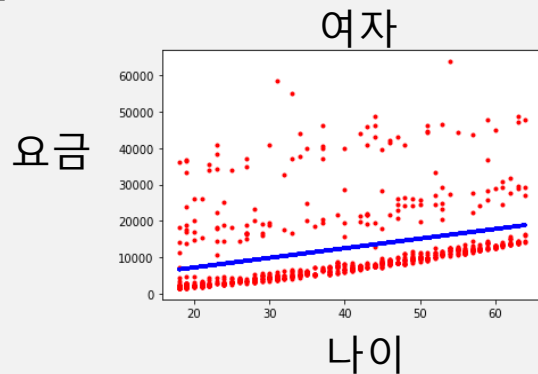
Lab – Linear Regression

- 첨부된 파일은 보험관련 데이터 이다.

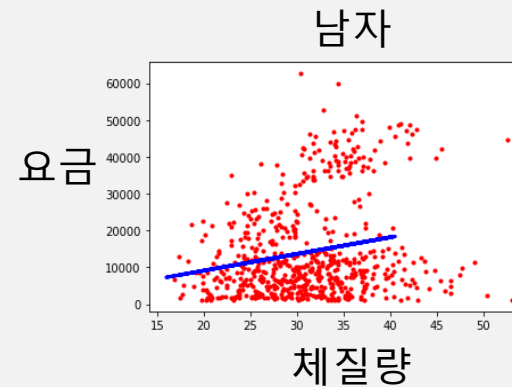
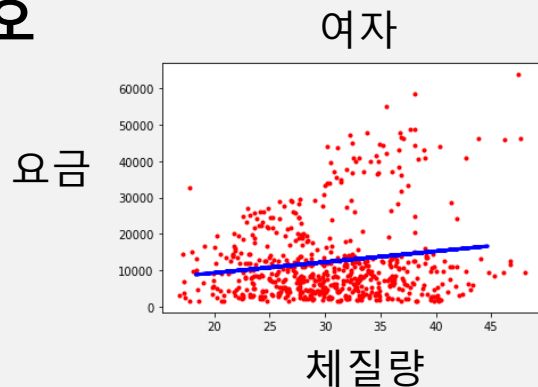
- 나이(age): 주 수혜자의 나이
- 성별(sex): 보험 계약자 성별, 여성, 남성
- 체질량 지수(bmi): 키에 비해 상대적으로 높거나 낮은 신체, 체중에 대한 이해를 제공하는 체질량 지수, 키와 몸무게의 비율을 사용하는 객관적인 체중 지수(kg/m^2), 이상적으로는 18.5~24.9입니다.
- 어린이(children): 건강 보험이 적용되는 자녀 수 / 부양 가족 수
- 흡연자(smoker): 흡연
- 지역(region): 미국, 북동부, 남동부, 남서부, 북서부 내 수혜자의 거주 지역.
- 요금(charges): 건강보험에서 청구하는 개별 의료비

Lab – Linear Regression (15점)

성별(남,여)에 따라 분리하여 다음 두 특징(나이vs 요금)으로 회귀 분석하고 해석 하시오.



성별(남,여)에 따라 분리하여 다음 두 특징(체질량vs 요금)으로 회귀 분석 하고 해석 하시오



Lab – Logistic Regression (10점)

주어진 데이터는 속성의 값에 따라 당뇨병 데이터로 당뇨병이 있는지(1) 없는지(0)에 대한 데이터이다. 아래 코드에서 X는 속성을 y는 당뇨병 여부의 라벨링을 의미한다. logistic regression으로 80:20 으로 cross validation 테스트하고 그결과를 confusion matrix로 구한 후 분석하여라. (TP, TN, FP, FN)

```
import pandas as pd
# Pandas Data 읽기맨위컬럼의값을 col_names 으로 정하기
col_names = ['pregnant', 'glucose', 'bp', 'skin', 'insulin', 'bmi', 'pedigree',
             'age', 'label']

#Pandas로 데이터 읽고 각 컬럼별 속성을 col_name으로 설정하기 정하기
pima = pd.read_csv("diabetes.csv", header=None, names=col_names)

feature_cols = ['pregnant', 'insulin', 'bmi', 'age', 'glucose', 'bp', 'pedigree']
X = pima[feature_cols] # Features
y = pima.label # Target variable
```

Lab – Logistic Regression (10점)

참고 : https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

Logistic regression – 5점
confusion matrix 구하기 3점 그 해석 2점