

로지스틱 회귀 연습문제

문제 1

피마 인디언 당뇨병 발병여부를 예측할 수 있는 분석 모델을 구현하기 위해 아래와 같은 항목들을 조사하였다. 분석하라.

변수	설명
Pregnancies	임신히트수
Glucose	포도당 부하 검사 수치
BloodPressure	혈압
SkinThickness	팔 삼두근 뒤쪽의 피하지방 측정값
Insulin	혈청 인슐린
BMI	체질량 지수
DiabetesPedigreeFunction	당뇨 내력 가중치 값
Age	나이
Outcome	당뇨여부(0 또는 1)

https://data.hossam.kr/E05/indian_diabetes.xlsx

단, 모든 독립변수는 명목형 변수를 포함하지 않으며 정규분포를 만족한다고 가정한다.

문제 2

다음의 데이터는 타이타닉 탑승객 명단 데이터이다. 적절한 데이터 전처리와 정제를 수행한 후 분석하라.

변수명	설명
PassengerId	탑승객의 ID(인덱스와 같은 개념)
Survived	생존유무(0은 사망 1은 생존)
Pclass	객실의 등급
Name	이름
Sex	성별
SibSp	동승한 형제 혹은 배우자의 수
Parch	동승한 자녀 혹은 부모의 수
Ticket	티켓번호
Fare	요금
Cabin	선실
Embarked	탑승지 (C = Cherbourg, Q = Queenstown, S = Southampton)

<https://data.hossam.kr/E05/titanic.xlsx>

단, 모든 독립변수는 정규분포를 만족한다고 가정한다.