

# 연습문제

---

## 문제 1

---

8-다중선형회귀 파일을 내용을 참고하여 선형회귀 모델을 다항회귀 모델로 구현하시오.

## 문제2

---

아래의 데이터셋은 미국 국립 당뇨병, 소화기 및 신장 질환 연구소에서 가져온 것으로 데이터셋의 목적은 데이터셋에 포함된 특정 진단 측정값을 기반으로 환자에게 당뇨병이 있는지 여부를 진단적으로 예측하는 것입니다.

<https://data.hossam.kr/F03/pima-indians.xlsx>

전체 모집단으로부터 몇 가지 제약이 적용되어 표본이 추출되었습니다.

피마 인디언 혈통의 21세 이상의 여성

데이터 세트는 여러 개의 의료 독립 변수와 하나의 종속 변수로 구성됩니다

종속 변수는 Outcome이며 독립 변수에는 환자의 임신 횟수, BMI, 인슐린 수준, 연령 등이 포함됩니다.

이 데이터셋을 활용하여 당뇨병 발병 여부를 판단할 수 있는 분류 모델을 구현하시오.

적용해야 하는 분석모델의 종류는 데이터의 형태에 따라 스스로 결정해야 합니다.

## 문제 3

---

아래의 데이터셋은 이탈리아의 한 지방에서 생산되는 와인 3종류의 성분을 화학 분석한 데이터셋이다.

[https://data.hossam.kr/F03/wine\\_3type.xlsx](https://data.hossam.kr/F03/wine_3type.xlsx)

종속변수인 와인의 종류는 `class` 라는 이름의 변수에 저장되어 있으며 그 외의 변수들은 화학적 성분을 의미한다.

이 데이터를 활용하여 와인의 종류를 분류할 수 있는 모델을 구현하라.

적용해야 하는 분석모델의 종류는 데이터의 형태에 따라 스스로 결정해야 합니다.