

# 연습문제

---

## 문제 1

---

8-다중선형회귀 파일의 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.

## 문제 4

---

아래의 이미지 셋은 비행기와 자동차를 구분해 놓은 이미지셋이다.

이 이미지를 통해 비행기와 자동차를 구분할 수 있는 학습모델을 구현하라.

<https://data.hossam.kr/F03/AirplaneAndCar.zip>

## 문제 5

---

아래의 이미지 셋은 가위,바위,보에 대한 이미지 셋이다.

이 이미지를 통해 가위,바위,보를 구분할 수 있는 학습 모델을 구현하라

<https://data.hossam.kr/F03/Rock-Paper-Scissors.zip>

## 문제 5

---

CIFAR-10 데이터 세트는 10개 클래스의 60000개의 32x32 컬러 이미지로 구성되며 클래스당 이미지는 6000개입니다. 50000개의 훈련 이미지와 10000개의 테스트 이미지가 있다.

데이터 세트는 5개의 훈련 배치와 1개의 테스트 배치로 나누어지며 각각 10,000개의 이미지가 있습니다. 테스트 배치에는 각 클래스에서 무작위로 선택된 이미지 1000개가 포함되어 있다. 훈련 배치에는 나머지 이미지가 무작위 순서로

포함되어 있지만 일부 훈련 배치에는 한 클래스의 이미지가 다른 클래스보다 더 많이 포함될 수 있습니다. 그 사이에 훈련 배치에는 각 클래스의 정확히 5000개의 이미지가 포함된다.

- airplane
- automobile
- bird
- cat
- deer
- dog
- frog
- horse
- ship
- truck

이 이미지를 각각의 카테고리를 구분할 수 있는 학습 모델을 구현하라

<https://data.hossam.kr/F03/CIFAR-10.zip>