

# 으뜸 머신러닝

강의자를 위한 안내서

# 책의 특징과 구성

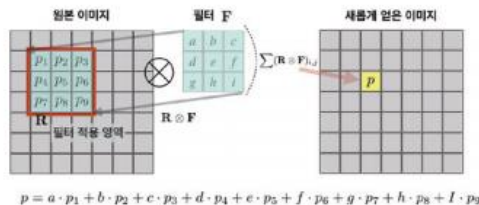
## 한눈에 읽히는 하나의 주제

이 책의 모든 절은 책을 펼쳤을 때에 좌우의 두 페이지에 담을 수 있는 양으로 나뉘어 언제나 한눈에 하나의 주제를 전체적으로 파악할 수 있도록 하였습니다.

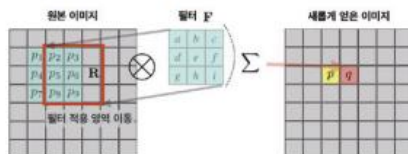
### 9.2 합성곱의 기본 개념 이해

이미지를 조작할 때 사용하는 방법 중에 **커널(kernel)** 혹은 **필터(filter)**라고 불리는 작은 행렬을 사용하여 새로운 값을 얻는 방법이 있다. SVM을 다룰 때, 커널이라는 개념을 사용했기 때문에, 혼돈을 피하기 위해 **이미지를 처리할 때 사용하는 이 행렬은 필터라고 부를 것이다**.

필터를 통해 이미지를 변형하는 것은 다음 그림과 같은 방식을 사용한다. 필터는 **F**로 표시된 작은 크기의 **파란색 행렬**이며, 이 영역을 원본 이미지에 덮었을 때, 필터가 덮이는 영역을 표시하는 부분 행렬이 **R**이다. 현재  $p_1$  위치의 픽셀 값을 필터가 적용된 이미지의  $p$ 로 바꾸는 작업이 진행되고 있다. 즉, 아래의 수식과 같이 필터와 원본 이미지의 대응되는 값들을 서로 곱한 뒤에 모두 더하여 새로운 이미지의 픽셀값  $p$ 를 구하는 것이다. 이러한 작업을 모든 픽셀에 대해 적용하면 이미지 전체의 변경이 이루어진다. 이러한 연산을 합성곱 연산, 혹은 컨볼루션이라고 한다.



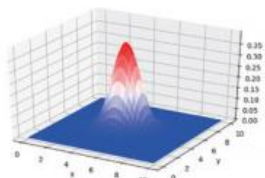
그 다음 픽셀  $q$ 를 구하는 일은 필터를 옆으로 한 칸 옮기면 된다. 이와 같이, 필터를 원본 이미지의 왼쪽 상단에서 시작하여 차례로 픽셀을 옮겨가며 전체 영역에 적용해 보자. 이런 방식으로 새롭게 얻은 이미지는 왼쪽과 같은 것이다.



간단한 필터링은 하나의 픽셀 값을 변경할 때 주위의 값을 고려하여 평균을 취하는 방법이다. 이렇게 평균을 취할 경우 튀는 값이 사라지고 좀 더 부드러운 이미지가 될 것이다. 만일 3x3 크기의 필터를 사용한다면 다음과 같은 행렬 **M**을 생각해 볼 수 있을 것이다. 이것은 인접한 픽셀과 자기 자신의 평균을 내는 필터가 될 것이다. 필터 중에서 이렇게 특정 픽셀 주변의 픽셀들이 가진 값의 평균을 취하는 필터를 **평균 필터(average filter)** 혹은 **상자 필터(box filter)**라고 한다.

$$M = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

가우스 필터는 필터 영역의 모든 픽셀에 동일한 중요성을 부여하는 것이 아니라 중심 픽셀에는 더 높은 중요도를 부여하고, 중심에서 멀어질수록 중요성을 낮게한다. 이 중요도를 **가중치(weight)**라고 부르고, 가우스 함수를 통해 그 값을 결정한다. 가우스 함수는 왼쪽 그림과 같은 **종 모양(bell shaped)** 함수 혹은 **정규 분포(normal distribution)** 함수로 잘 알려져 있다. 가운데 볼록하게 솟은 부분은 현재 색상을



변경할 대상 픽셀의 위치이다. 이 솟아 오른 정도가 중요도, 혹은 가중치로 사용된다. 상자 필터와 가우스 필터로 이미지를 처리하면 아래 그림의 왼쪽 첫 그림과 같은 원본 이미지가 상자 필터에 의해서는 가운데 그림, 가우스 필터에 의해서는 오른쪽 끝 그림과 같이 바뀐다.



원본 이미지      상자 필터를 이용한 합성곱      가우스 필터를 이용한 합성곱

# 설명과 구현, 생각의 폭 넓히기

이 책은 교재와 자습서 모두로 사용될 수 있도록 이론 설명과 함께 따라할 수 있는 실습 코드가 늘 다음과 같이 코랩 환경에서 돌아가는 형태로 제공됩니다.



```
svm_simple.predict([[0.12, 0.56], [-4, 40], [0, 40], [5, 20]])
```

```
array([0, 1, 0, 0])
```

## 생각의 폭을 넓히기

보충 설명이 필요하거나, 알아두면 좋은 내용들은 본문의 흐름에 방해되지 않도록 잠깐이라는 이름의 상자에 담아 정리했습니다.



### 잠깐 - 복잡한 함수로 데이터를 설명하는 일은 쉽다

현대적 컴퓨터의 구조를 확립한 **존 폰 노이만** John von Neumann은 이런 말을 남겼다. "내게 네 개의 파라미터만 주면 코끼리를 만들 수 있고, 다섯 개라면 코를 흔들게 할 수 있다." 이것은 복잡한 함수의 모양이 데이터에 담긴 특징이나 패턴과는 무관할 수 있음을 나타낸다. "웃자고 한 이야기에 죽자고 덤벼드는" 과학자들은 실제로 정확히 네 개의 복소수 파라미터와 **푸리에** Fourier 급수를 이용하여 이런 코끼리를 그렸다<sup>1)</sup>.



# 주제별 LAB

필요할 때마다, 학습한 내용을 정리할 수 있는 집중적인 실습을 LAB으로 제공하고 있습니다. LAB은 아래와 같은 제목을 시작으로 별도로 정리된 절을 구성하는데, 이때는 페이지의 제한 없이 하나의 주제를 충분히 다룰 수 있는 완결된 실습이 되도록 했습니다.

## LAB<sup>6-1</sup>

### 다항 회귀의 회귀 함수를 그려 보자

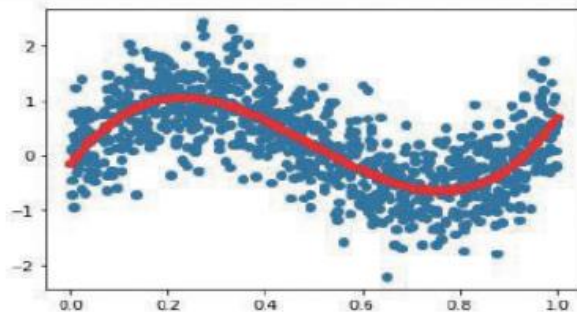
#### 실습 목표

아래의 위치에는 직선으로 표현하기 어려운 데이터가 있다.

<https://github.com/dknife/ML/raw/main/data/nonlinear.csv>

이 데이터를 화면에 그려보고 이 데이터를 설명하는 회귀 함수를 다항 회귀를 이용하여 찾아 보자. 그리고 회귀 함수를 데이터의 독립 변수 범위 내에서 가시화해 보라.

#### 원하는 결과





# 미니 프로젝트

- 난이도가 상당히 높은 프로젝트로 2-4인의 학습자 팀으로 구현해 보는 것이 좋습니다.

책이 다루는 주제에 따라 적절한 시점에 미니 프로젝트를 제시하고 있습니다. 미니 프로젝트는 다소 도전적인 문제를 해결할 수 있도록 돕기 위해 만들어졌으며, 문제를 따라하면서 책에서 다룬 설명을 확인할 수 있을 뿐만 아니라, 다양한 문제로 이론을 응용하여 적용할 수 있는 능력을 키울 수 있을 것입니다. 때로는 **까다롭고도 도전적인 주제**를 다루고 있어 따라하고 개선하는 과제로 활용하면 좋을 것입니다.

## 미니 프로젝트 B1

### 얼굴 찾기: SVM으로 분류하기

#### 목표

SVM을 이용하여 주어진 이미지가 사람의 얼굴인지 아닌지를 구분하는 프로그램을 만들어 보자.

#### 1. 데이터 확인

데이터의 위치는 아래와 같다.

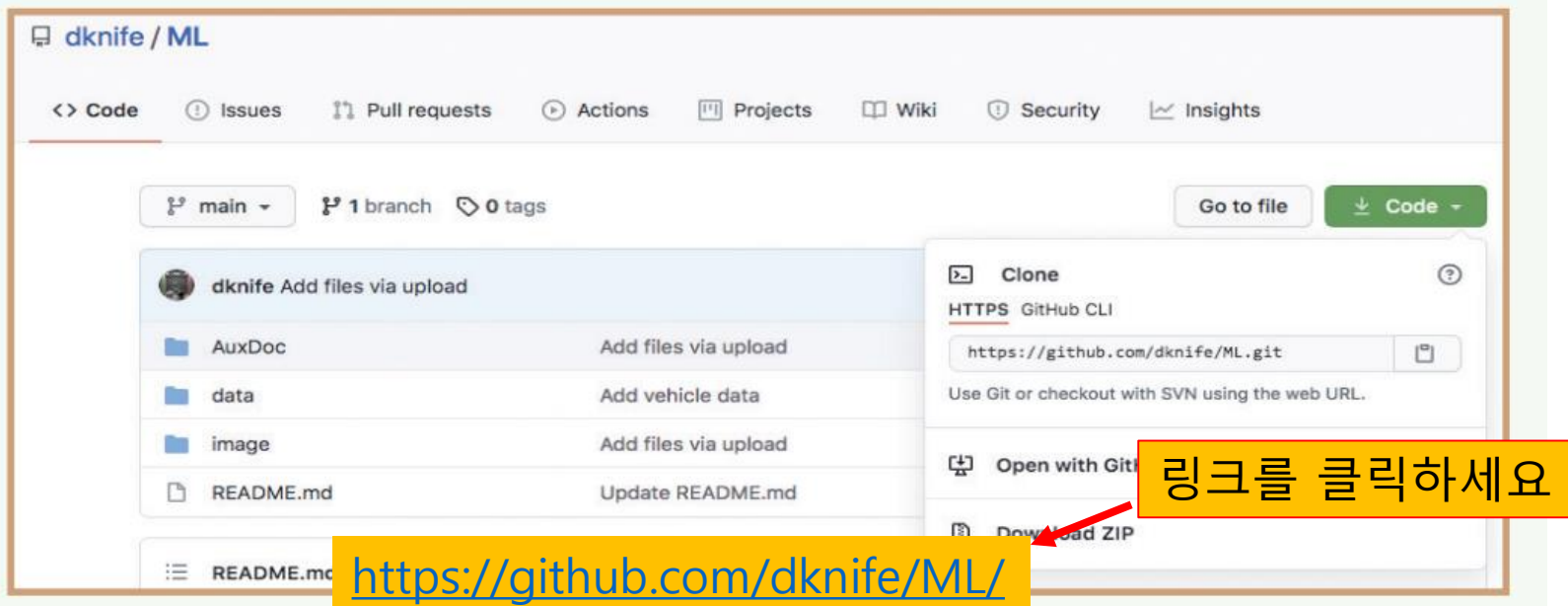
<https://github.com/dknife/ML/raw/main/data/Proj2>

# 소스코드와 데이터

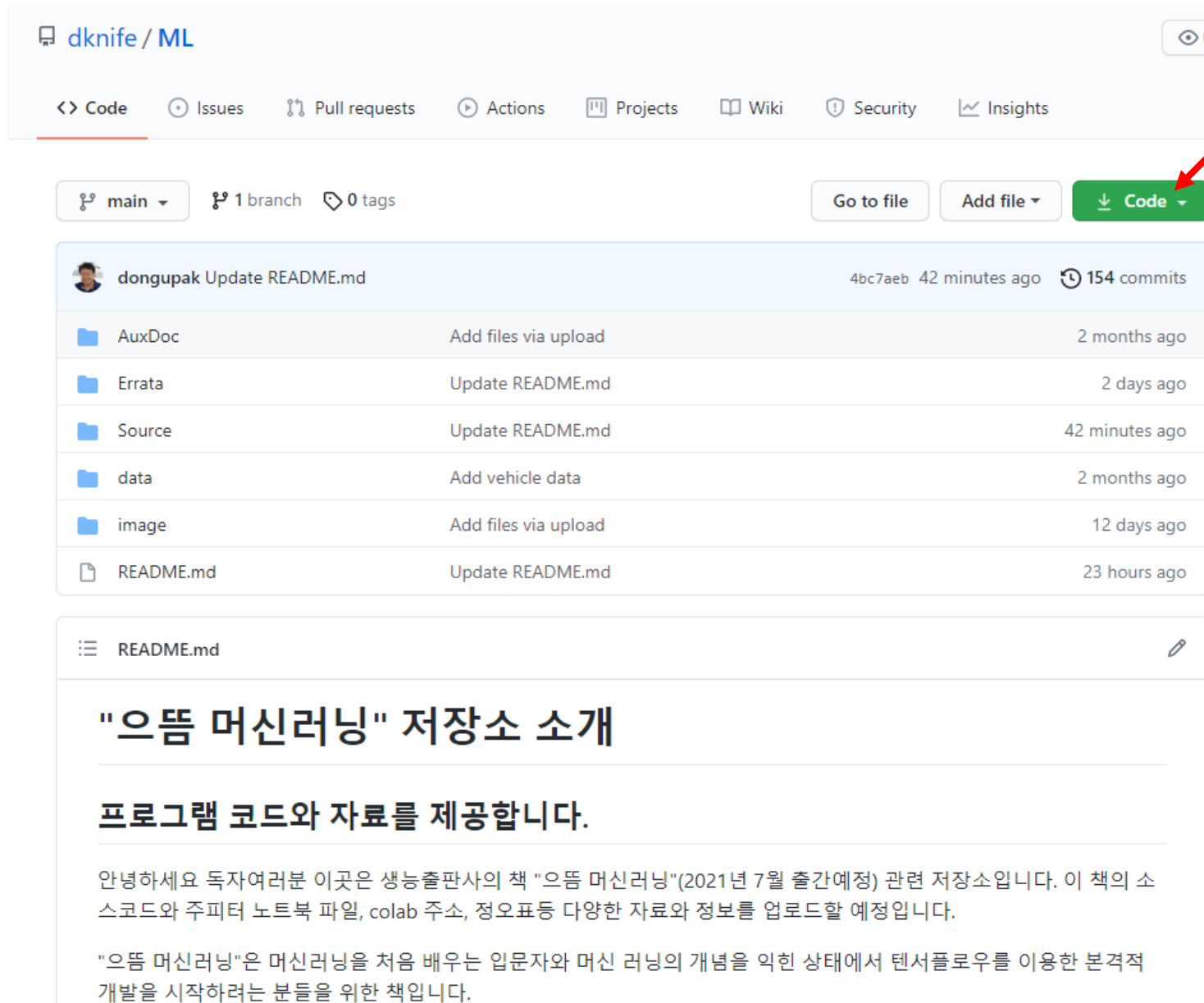
이 책에 나오는 모든 소스 코드와 각종 데이터 파일은 저자의 [github.com](https://github.com/dknife/ML) 사이트에서 다운받을 수 있습니다. [github](https://github.com/dknife/ML)은 프로그램 개발자들이 소프트웨어의 버전을 관리하고 소스코드를 공유하는 협업서비스로, 이 책과 관련된 저자의 [github](https://github.com/dknife/ML) 사이트는 다음과 같습니다.

<https://github.com/dknife/ML>

[github](https://github.com/dknife/ML)의 사용법에 익숙한 독자라면 포크 기능을 사용해서 코드를 자신의 [github](https://github.com/dknife/ML)에 복사하는 방법이 있으나, [github](https://github.com/dknife/ML)의 사용법에 익숙하지 않다면 아래 그림과 같이 초록색 Code 메뉴를 누르면 펼쳐지는 팝업 메뉴에서 Download ZIP 메뉴를 선택하여 코드를 다운받을 수 있습니다.



# github 접속하여 자료와 파일 다운받기



dknife / ML

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main 1 branch 0 tags

Go to file Add file

Code

dongupak Update README.md 4bc7aeb 42 minutes ago 154 commits

AuxDoc	Add files via upload	2 months ago
Errata	Update README.md	2 days ago
Source	Update README.md	42 minutes ago
data	Add vehicle data	2 months ago
image	Add files via upload	12 days ago
README.md	Update README.md	23 hours ago

README.md

## "으뜸 머신러닝" 저장소 소개

### 프로그램 코드와 자료를 제공합니다.

안녕하세요 독자여러분 이곳은 생능출판사의 책 "으뜸 머신러닝"(2021년 7월 출간예정) 관련 저장소입니다. 이 책의 소스코드와 주피터 노트북 파일, colab 주소, 정오표등 다양한 자료와 정보를 업로드할 예정입니다.

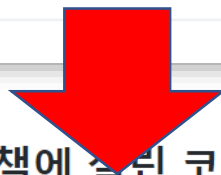
"으뜸 머신러닝"은 머신러닝을 처음 배우는 입문자와 머신 러닝의 개념을 익힌 상태에서 텐서플로우를 이용한 본격적 개발을 시작하려는 분들을 위한 책입니다.

파일 전체 다운로드가 가능합니다.

## 책의 소스코드(colab 주소, .py 파일, ipynb 파일)

- [소스코드](#) : "으뜸 머신러닝" 책의 장별 소스코드가 .ipynb, colab 페이지를 통해서 제공됩니다

"소스코드" 링크 클릭 보면 좋은 책



### 으뜸 머신러닝 책에 실린 코드들입니다.

위의 코드가 노트북 환경에서 열리지 않는다면, 아래 링크를 클릭하시기 바랍니다.

[3장 코드](#) 구현을 위한 도구 (파이썬, 코랩, 넘파이)

- 코랩 주소 : [https://colab.research.google.com/drive/11J--272C0cPrHqZkWUDJMi2\\_vsL4nmxmK](https://colab.research.google.com/drive/11J--272C0cPrHqZkWUDJMi2_vsL4nmxmK)

[4장 코드](#) 선형 회귀로 이해하는 지도학습 (선형회귀 구현, Scikit-learn 활용)

- 코랩 주소 : <https://colab.research.google.com/drive/1jB-Cnzgd30hzNPhlggXN-s8QG8dgsUUc>

[5장 코드](#) 분류와 군집화로 이해하는 지도 학습과 비지도 학습 (k-NN, k-means)

- 코랩 주소 : [https://colab.research.google.com/drive/1XQe7\\_FCAHlbnig44Lplew2Bildufn7oF](https://colab.research.google.com/drive/1XQe7_FCAHlbnig44Lplew2Bildufn7oF)

[5장 미니프로젝트 A1](#) 미니 프로젝트 A1 잡음제거: k-NN의 활용

- 코랩 주소 : <https://colab.research.google.com/drive/1X45XKyEHgOpR0Mq56eZvEGZPpoA4v8Vq>

[6장 코드](#) 다양한 머신러닝 기법들: 다항 회귀, 결정 트리, SVM

- 코랩 주소 : [https://colab.research.google.com/drive/1NKyKM9P85\\_yybabF-3zSo1jH2hBWc13C](https://colab.research.google.com/drive/1NKyKM9P85_yybabF-3zSo1jH2hBWc13C)

[6장 미니프로젝트 B1](#) 얼굴 찾기: SVM으로 분류하기

- 코랩 주소 : [https://colab.research.google.com/drive/1UrZjeFCtwcLHqefXhZBa2mzeltx9\\_SSi](https://colab.research.google.com/drive/1UrZjeFCtwcLHqefXhZBa2mzeltx9_SSi)

[7장 코드](#) 인공 신경망 기초 - 문제와 돌파구 (퍼셉트론, 다항입력, 다층 퍼셉트론)

- 코랩 주소 : [https://colab.research.google.com/drive/1USlwBX\\_eo9h-3aMyqFe8pDxFbSbkMQZN](https://colab.research.google.com/drive/1USlwBX_eo9h-3aMyqFe8pDxFbSbkMQZN)



"6장 코드" 링크 클릭시  
책의 코드를 보실수 있습니다.  
- 복사하여 사용가능

생능 출판사 "으뜸 머신러닝 (1권) 프로젝트 코드 코드

## 6장 다양한 머신러닝 기법들

- 출판사 : 생능 출판사( <http://www.booksr.co.kr/> )
- 으뜸 파이썬 저자 : 강영민, 박동규, 김성수
- 소스코드 저장소 : <https://github.com/dknife/ML>
- 저작권 : 본 주피터 노트북 코드는 자유롭게 배포가능하지만 위의 출판사, 저서, 저자표기와 함께 배포해 주십시오.

### 6.1 다항 회귀

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

data_loc = 'https://github.com/dknife/ML/raw/main/data/'
life = pd.read_csv(data_loc + 'life_expectancy.csv')
life.head()

life = life[['Life expectancy', 'Alcohol', 'Percentage expenditure',
            'Polio', 'BMI', 'GDP', 'Thinness 1-19 years']]
life.dropna(inplace = True)

X = life[['Alcohol', 'Percentage expenditure', 'Polio',
            'BMI', 'GDP', 'Thinness 1-19 years']]
y = life['Life expectancy']

from sklearn.preprocessing import PolynomialFeatures

poly_feature = PolynomialFeatures(degree = 3)
X = poly_feature.fit_transform(X)
```

“코랩 코드” 링크 클릭시  
책의 코드를 코랩 환경에서 보고 바로 실행할 수  
있습니다.

목차

생능 출판사 "으뜸 머신러닝"(1판) 교재의  
소스 코드

6.1 다항 회귀

6.2 다항 회귀의 문제점 - 과적합, 그리고  
폭발적인 복잡도 증가

6.3 과적합과 과소적합 - 공짜 점심은  
없다

LAB 6-1: 다항 회귀의 회귀 함수를 그  
려 보자

6.4 결정 트리를 이용한 분류

6.5 어떤 속성이 가장 중요한가?

6.6 결정 트리를 손으로 만들어 보자 -  
ID3 알고리즘

6.7 지니 불순도를 이용한 효율적인 평  
가 - CART 알고리즘

6.8 사이킷런의 결정 트리로 붓꽃 분류  
하기

LAB 6-2: 꽃받침의 너비와 길이로 결  
정트리를 만들자

LAB 6-3: 엔트로피를 이용하여 결정

+ 코드 + 텍스트

연결

생능 출판사 "으뜸 머신러닝"(1판) 교재의 소스 코드

6장 다양한 머신러닝 기법들

- 출판사 : 생능 출판사( <http://www.booksr.co.kr/> )
- 으뜸 파이썬 저자 : 강영민, 박동규, 김성수
- 소스코드 저장소 : <https://github.com/dknife/ML>
- 저작권 : 본 주피터 노트북 코드는 자유롭게 배포가능하지만 위의 출판사, 저서, 저자표기와 함께 배포해 주

6.1 다항 회귀

```
[ ] 1 import pandas as pd
    2 import matplotlib.pyplot as plt
    3 from sklearn.model_selection import train_test_split
    4 from sklearn.linear_model import LinearRegression
    5 from sklearn.metrics import mean_squared_error
```

경고: 이 노트는 Google에서 작성하지 않았습니다.

이 노트는 **dgpark@gs.cwnu.ac.kr**님이 작성했습니다. 노트가 Google에 저장된 데이터에 액세스하거나 다른 세션의 데이터 및 사용자 인증 정보를 읽을 권한을 요청할 수 있습니다. 노트를 실행하기 전에 소스 코드를 검토하세요. 궁금한 점이 더 있다면 노트 작성자 (dgpark@gs.cwnu.ac.kr)에게 문의하시기 바랍니다.

취소

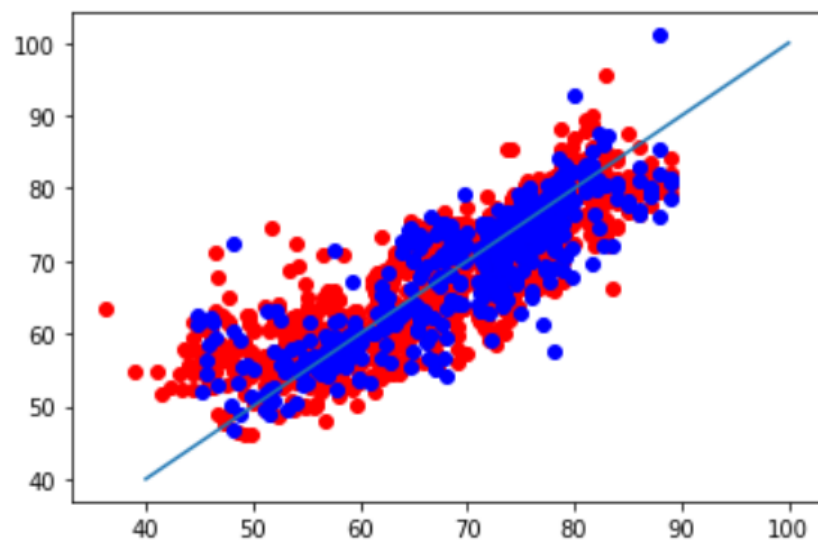
무시하고 계속하기

"무시하고 계속하기 " 선택하세요

아래와 같이 코드의 수행결과를 코랩에서 바로 확인하실 수 있습니다.

```
1 X_train,X_test,y_train,y_test = train_test_split(X, y, test_size = 0.2)
2 lin_model = LinearRegression()
3 lin_model.fit(X_train, y_train)
4
5 y_hat_train = lin_model.predict(X_train)
6 y_hat_test = lin_model.predict(X_test)
7 plt.scatter(y_train, y_hat_train, color='r')
8 plt.scatter(y_test, y_hat_test, color='b')
9 plt.plot([40, 100], [40, 100])
10 print('Mean squared error:', mean_squared_error(y_test, y_hat_test))
```

Mean squared error: 28.814491614427215

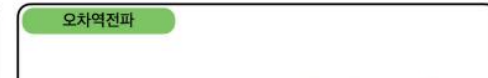
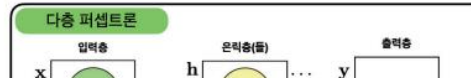
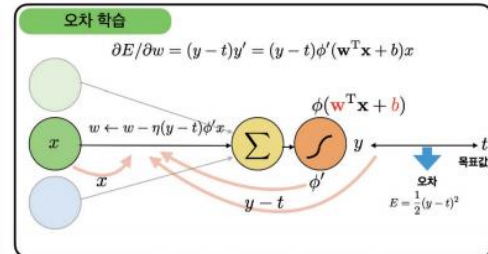
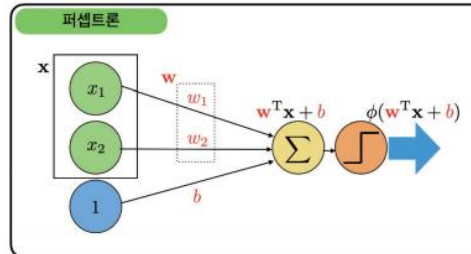
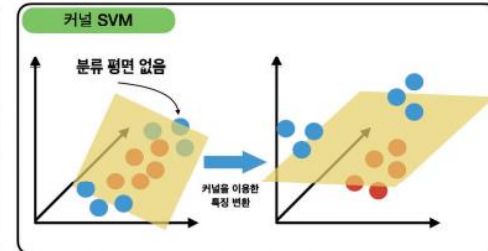
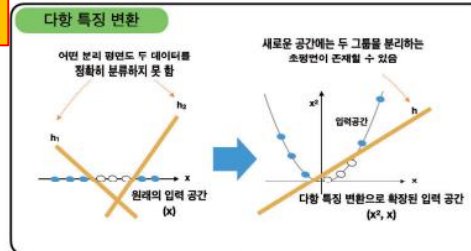
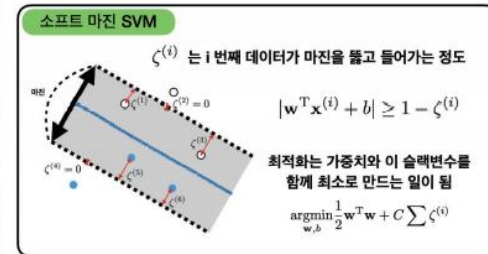
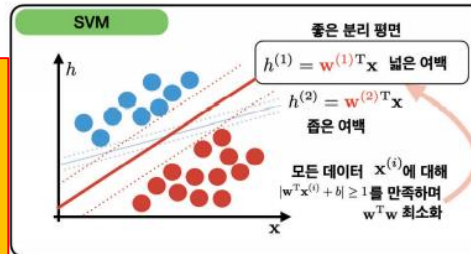


✓ 1초 오후 4:33에 완료됨

# 치트 시트이용 : 책 내용의 비주얼 요약

학습자들의 평가시에 치트시트를 이용하도록 하여 평가하는 것도 가능합니다.  
책 내용의 핵심을 담고 있으며,  
절취선을 따라서 잘라주시면 됩니다.

## 으뜸 머신러닝 - Cheat Sheet



절  
취  
선



# 문의사항

- 아래의 저자 이메일을 이용해 주십시오
  - 대표저자 : 강영민교수(동명대학교)
  - 이메일 : [ymkang@tu.ac.kr](mailto:ymkang@tu.ac.kr)
  - github : <https://github.com/dknife/ML/>