



python for AI

Python is Most Popular programming language in AI

- 1) Python the Best programming language in AI**
- 2) Community support**
- 3) Good visualization options**
- 4) Easy Readability**
- 5) Platform independence**
- 6) Flexibility**
- 7) Easy Coding**
- 8) A rich Python Libraries for AI Projects**

Python

프로그래밍 언어 파이썬은 인공지능, 데이터 분석에서 막강하며, 웹, 데이터베이스, 수학/과학계산, 인공지능 솔루션 개발 등에 쓸 수 있는 수많은 라이브러리가 무료인 오픈소스로 제공하고 있다.

<https://github.com/kgpark88/python/>

프로그래밍을
배워야 하는 이유



컴퓨팅 사고력

통합개발환경에서 코딩하기



프로그램 제어문



왜 파이썬 인가?

IDE보다 간편한 코드에디터

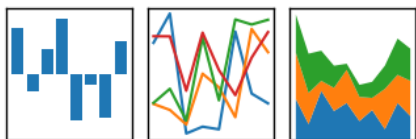
Python 주요 라이브러리



행렬과 대규모 다차원 배열을 쉽게 처리 할 수 있게 해주는 라이브러리

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



데이터 분석 패키지로 데이터를 처리하고 분석하는 데 효과적인 도구



머신러닝 교육 및 실무를 위한 머신러닝 라이브러리



기계 학습과 딥러닝을 위해 구글에서 만든 오픈소스 라이브러리



하이레벨 딥러닝 API로 대부분의 딥러닝 라이브러리 지원
(TensorFlow, Theano, CNTK)

NumPy

NumPy는 파이썬이 계산과학분야에 이용될 때 핵심 역할을 하는 라이브러리로
고성능의 다차원 배열 객체와 이를 다룰 도구를 제공한다.

```
data = np.array([1,2,3])
```

data

1
2
3

data

1
2
3

.max() =

3

data

1
2
3

.min() =

1

data

1
2
3

.sum() =

6

```
np.array([ [[1,2],[3,4]],  
          [[5,6],[7,8]] ])
```



		5	6
1	2		
3	4		8

```
np.random.random((4,3,2))
```

		0.3	0.6	0.8
	0.2	0.5	0.3	0.8
	0.7	0.6	0.1	0.5
	0.4	0.5	0.5	0.3
	0.1	0.1	0.4	

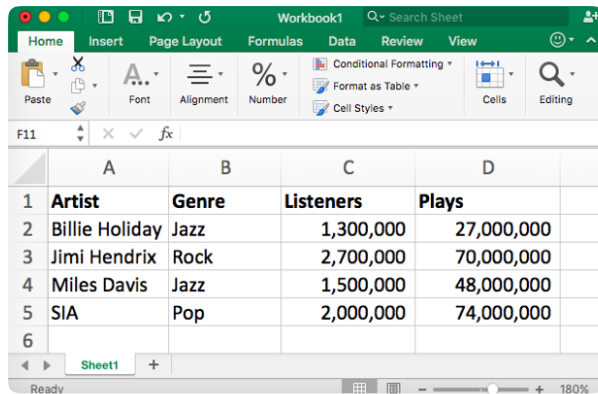
$$\text{MeanSquareError} = \frac{1}{n} \sum_{i=1}^n (Y_{\text{prediction}_i} - Y_i)^2 \quad \rightarrow \quad \text{error} = (1/n) * \text{np.sum}(\text{np.square}(\text{predictions} - \text{labels}))$$

출처 : <http://jalammar.github.io/visual-numpy/>

Pandas

데이터분석 라이브러리로 행과 열로 이루어진 데이터 객체를 만들어 다룰 수 있으며,
보다 안정적으로 대용량의 데이터들을 처리하는데 매우 편리한 도구이다.

music.csv



	Artist	Genre	Listeners	Plays
1	Billie Holiday	Jazz	1,300,000	27,000,000
2	Jimi Hendrix	Rock	2,700,000	70,000,000
3	Miles Davis	Jazz	1,500,000	48,000,000
4	SIA	Pop	2,000,000	74,000,000

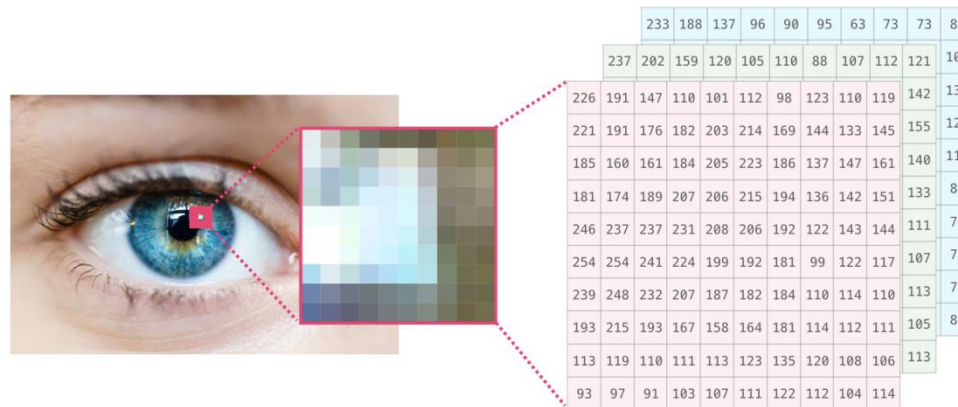
`pandas.read_csv('music.csv')`



	Artist	Genre	Listeners	Plays
0	Billie Holiday	Jazz	1,300,000	27,000,000
1	Jimi Hendrix	Rock	2,700,000	70,000,000
2	Miles Davis	Jazz	1,500,000	48,000,000
3	SIA	Pop	2,000,000	74,000,000

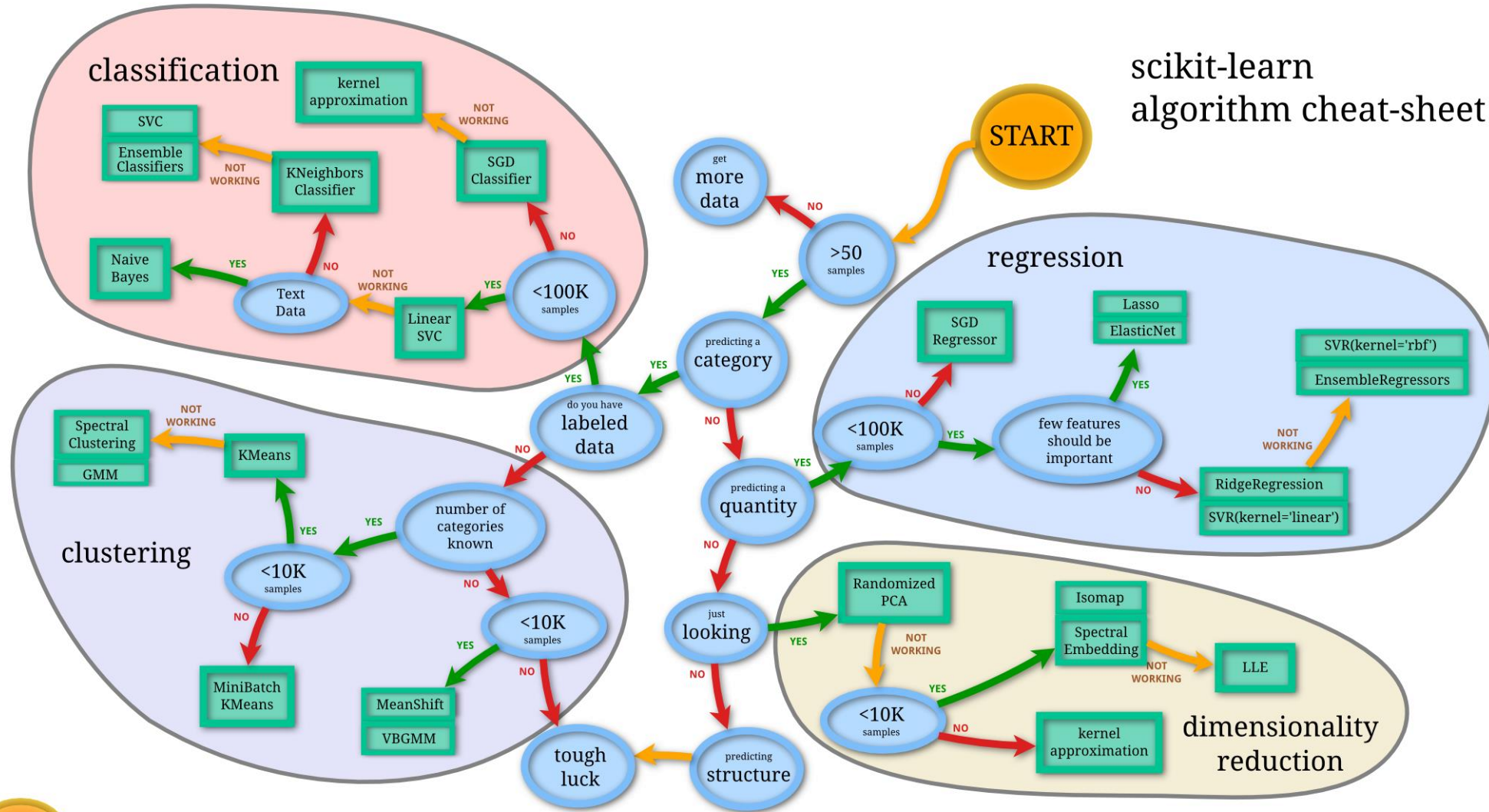
`df['Avg Plays'] = df['Plays']/df['Listeners']`

	Artist	Genre	Listeners	Plays	Avg Plays
0	Billie Holiday	Jazz	1,300,000	27,000,000	20
1	Jimi Hendrix	Rock	2,700,000	70,000,000	25
2	Miles Davis	Jazz	1,500,000	48,000,000	32
3	SIA	Pop	2,000,000	74,000,000	37



Scikit-learn

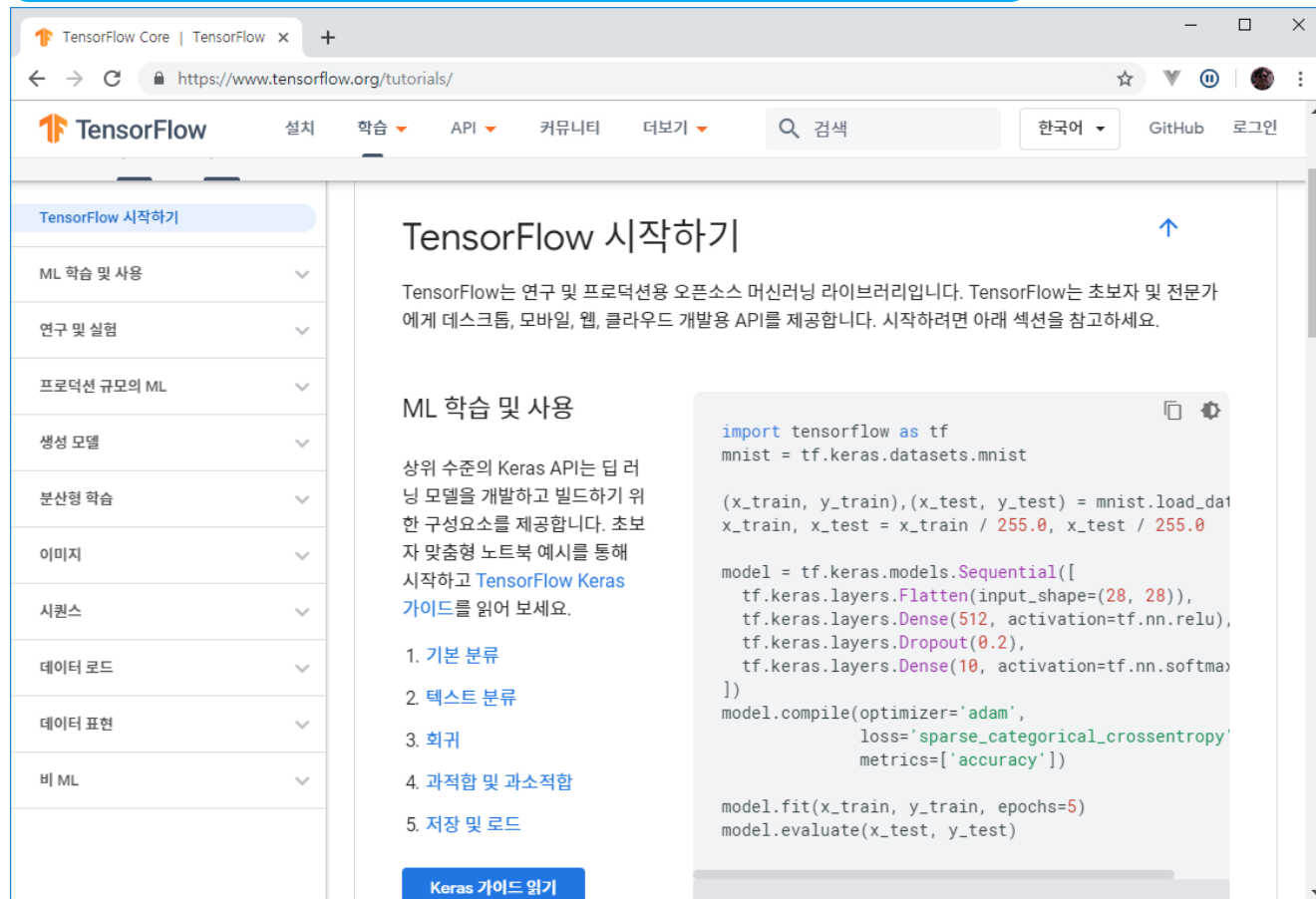
scikit-learn algorithm cheat-sheet



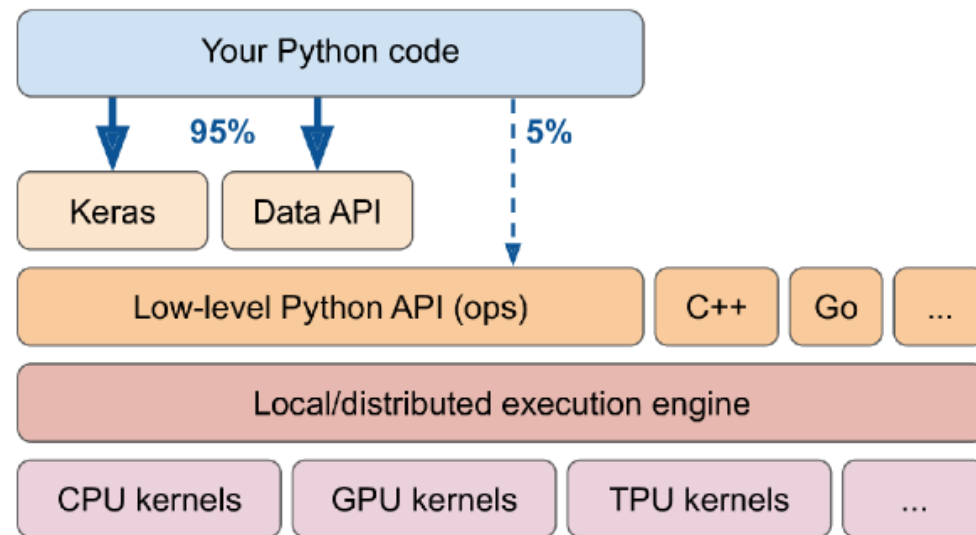
TensorFlow, Keras

텐서플로는 강력한 수치계산용 라이브러리로 현재 가장 인기 있는 딥러닝 라이브러리입니다. 케라스(Keras)는 모든 종류의 신경망을 쉽게 만들고 훈련, 평가, 실행할 수 있는 고수준 딥러닝 API입니다.

<https://www.tensorflow.org/tutorials/>

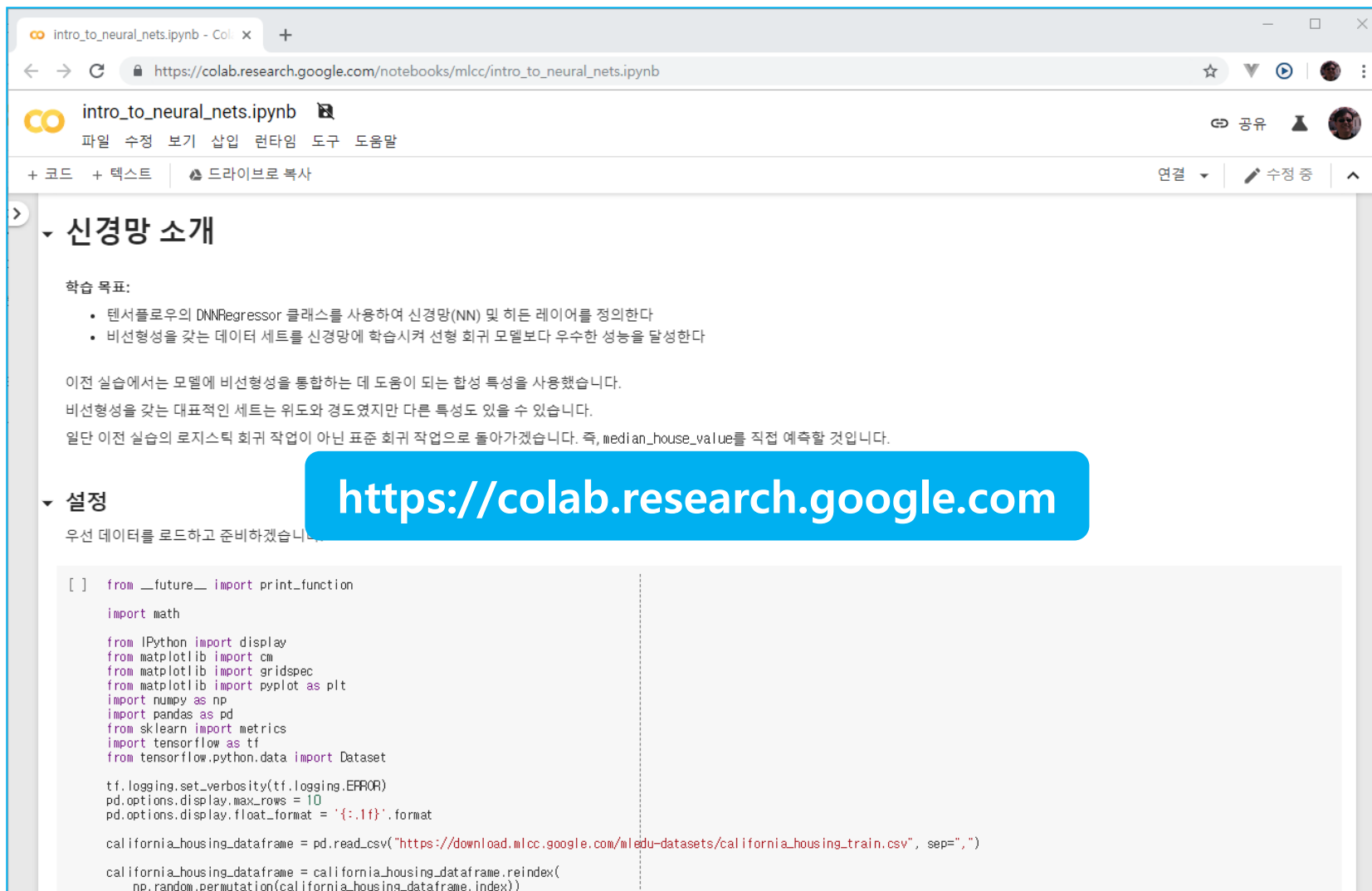


텐서플로 구조



Google Colaboratory

프로그램 설치 없이 클라우드에서 실행되는 개발 환경으로 웹브라우저에서 코드를 작성 및 실행한다.



The screenshot displays a Google Colaboratory notebook titled "intro_to_neural_nets.ipynb". The interface includes a browser address bar with the URL https://colab.research.google.com/notebooks/mlcc/intro_to_neural_nets.ipynb. The notebook content is in Korean and covers the following sections:

- 신경망 소개** (Introduction to Neural Networks):
 - 학습 목표:** (Learning Objectives)
 - 텐서플로우의 DNNRegressor 클래스를 사용하여 신경망(NN) 및 히든 레이어를 정의한다
 - 비선형성을 갖는 데이터 세트를 신경망에 학습시켜 선형 회귀 모델보다 우수한 성능을 달성한다
 - 이전 실습에서는 모델에 비선형성을 통합하는 데 도움이 되는 합성 특성을 사용했습니다.
 - 비선형성을 갖는 대표적인 세트는 위도와 경도였지만 다른 특성도 있을 수 있습니다.
 - 일단 이전 실습의 로지스틱 회귀 작업이 아닌 표준 회귀 작업으로 돌아가겠습니다. 즉, median_house_value를 직접 예측할 것입니다.
- 설정** (Setup):
 - 우선 데이터를 로드하고 준비하겠습니다.

The code cell under the "설정" section contains the following Python code:

```
[ ] from __future__ import print_function
import math

from IPython import display
from matplotlib import cm
from matplotlib import gridspec
from matplotlib import pyplot as plt
import numpy as np
import pandas as pd
from sklearn import metrics
import tensorflow as tf
from tensorflow.python.data import Dataset

tf.logging.set_verbosity(tf.logging.ERROR)
pd.options.display.max_rows = 10
pd.options.display.float_format = '{:.1f}'.format

california_housing_dataframe = pd.read_csv("https://download.mlcc.google.com/mledu-datasets/california_housing_train.csv", sep=",")

california_housing_dataframe = california_housing_dataframe.reindex(
    np.random.permutation(california_housing_dataframe.index))
```

프로그램 설치

■ Python 설치

<https://www.python.org/>



■ Python 가상환경 설치 : 윈도우즈

소스코드를 저장한 디렉토리로 이동

```
python -m venv venv
```

```
cd venv
```

```
cd scripts
```

```
activate
```

■ Python 가상환경 설치 : 리눅스

소스코드를 저장한 디렉토리로 이동

```
python3 -m venv venv
```

```
cd venv
```

```
cd bin
```

```
source activate
```

■ Python 패키지 설치

```
pip install pandas
```

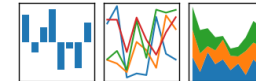
```
pip install matplotlib
```

```
pip install jupyter
```

```
pip install tensorflow
```

```
pytorch : pytorch 홈페이지 Get Started 참조
```

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



matplotlib

jupyter



PYTORCH

* PIP connection Error : SSL CERTIFICATE VERIFY FAILED 발생시

```
pip install --trusted-host pypi.org
```

```
--trusted-host pypi.python.org
```

```
--trusted-host files.pythonhosted.org 패키지명
```

■ 코드에디터 설치

Visual Studio Code , Sublime Text, Atom 등 선호하는 에디터

Thank you