ML Exercise Guide for Samsung DS^2

기계학습 실습환경 설정 가이드

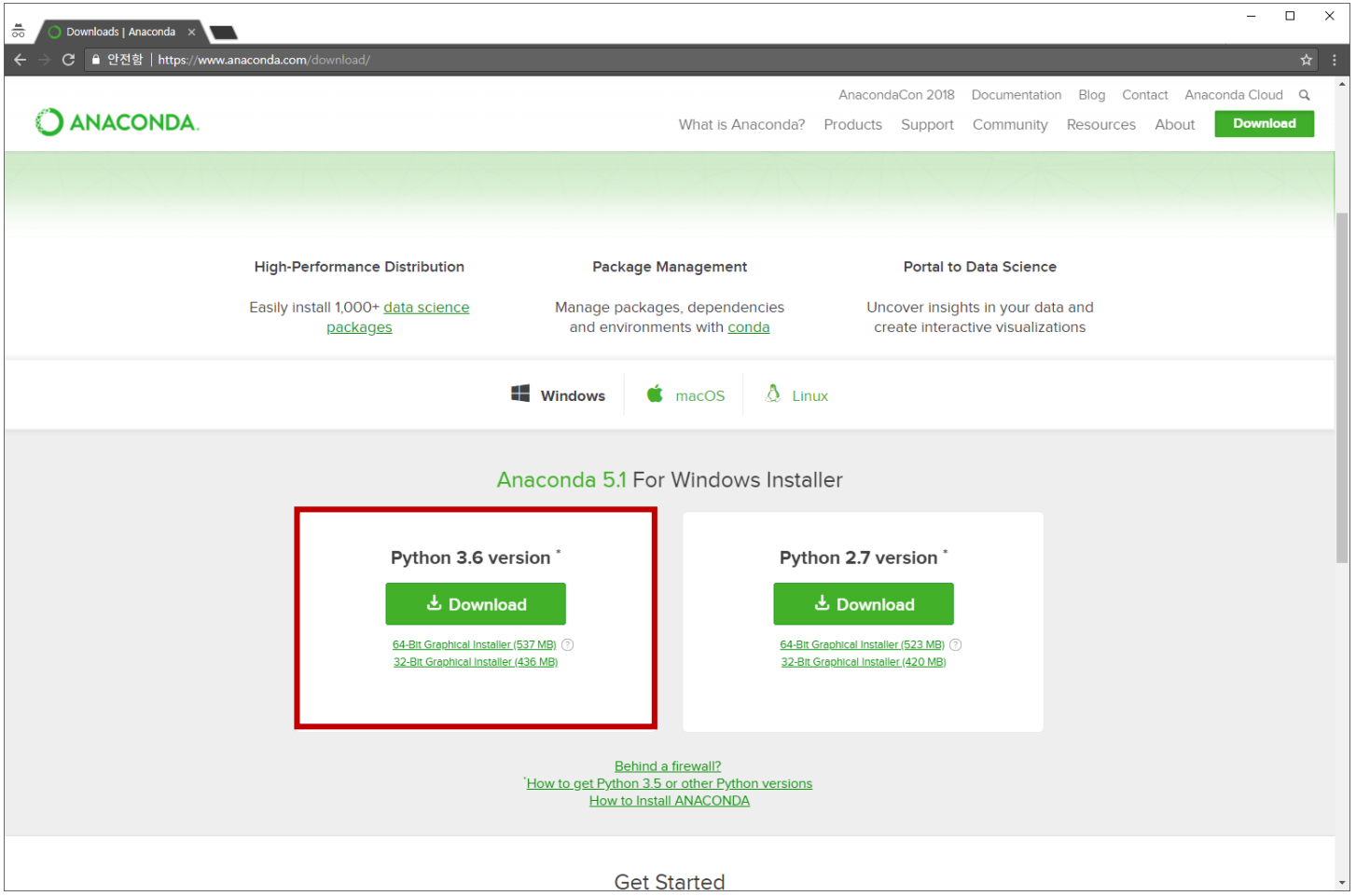
2018.10.12

Wonjong Rhee, Hyunghun Cho, Daeyoung Choi
Seoul National University
Graduate School of Convergence Science and Technology
Applied Data Science Lab.

실습 환경 구축

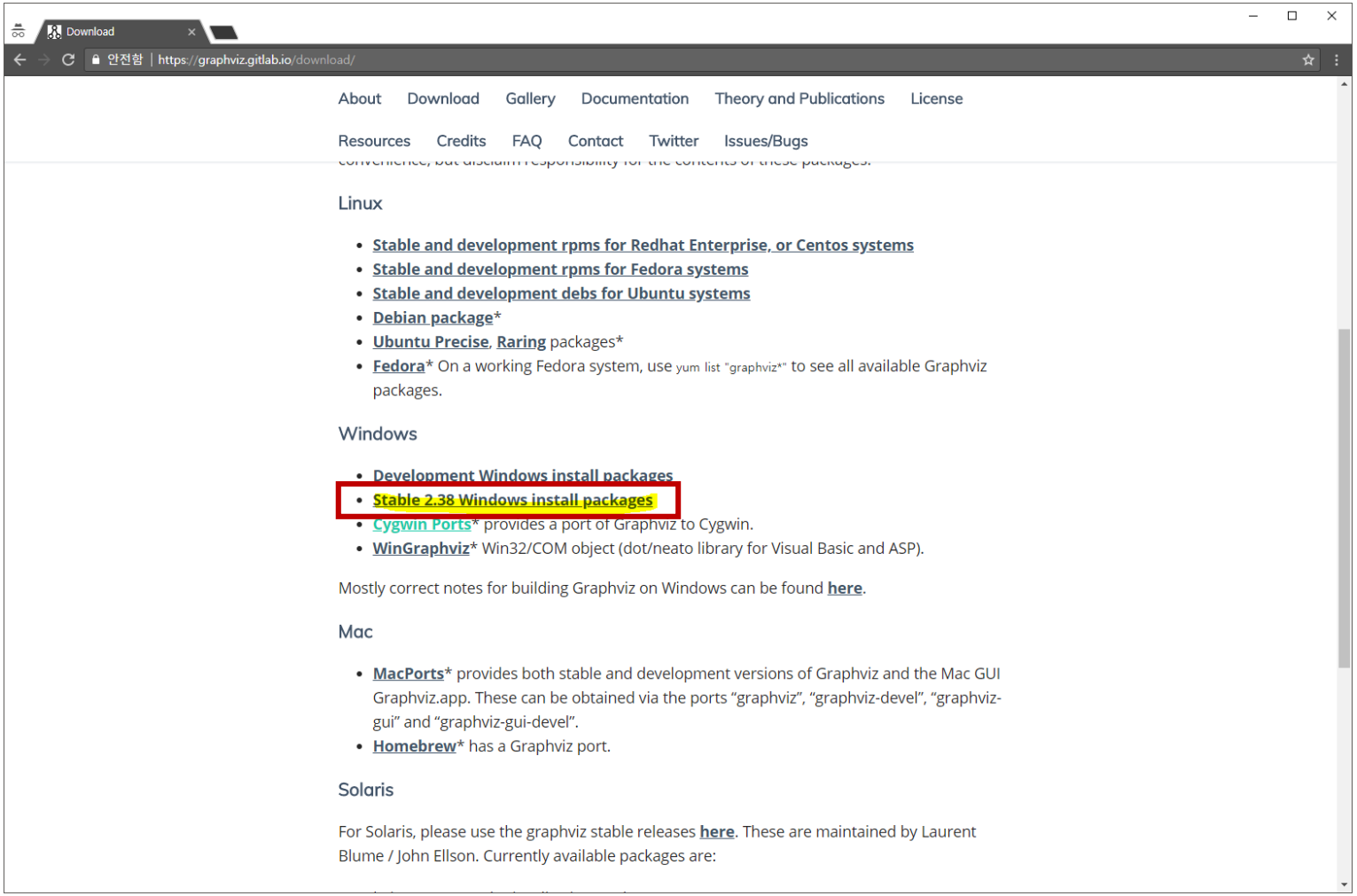
Prerequisites (1)

- Anaconda 설치 (www.anaconda.com/download/)



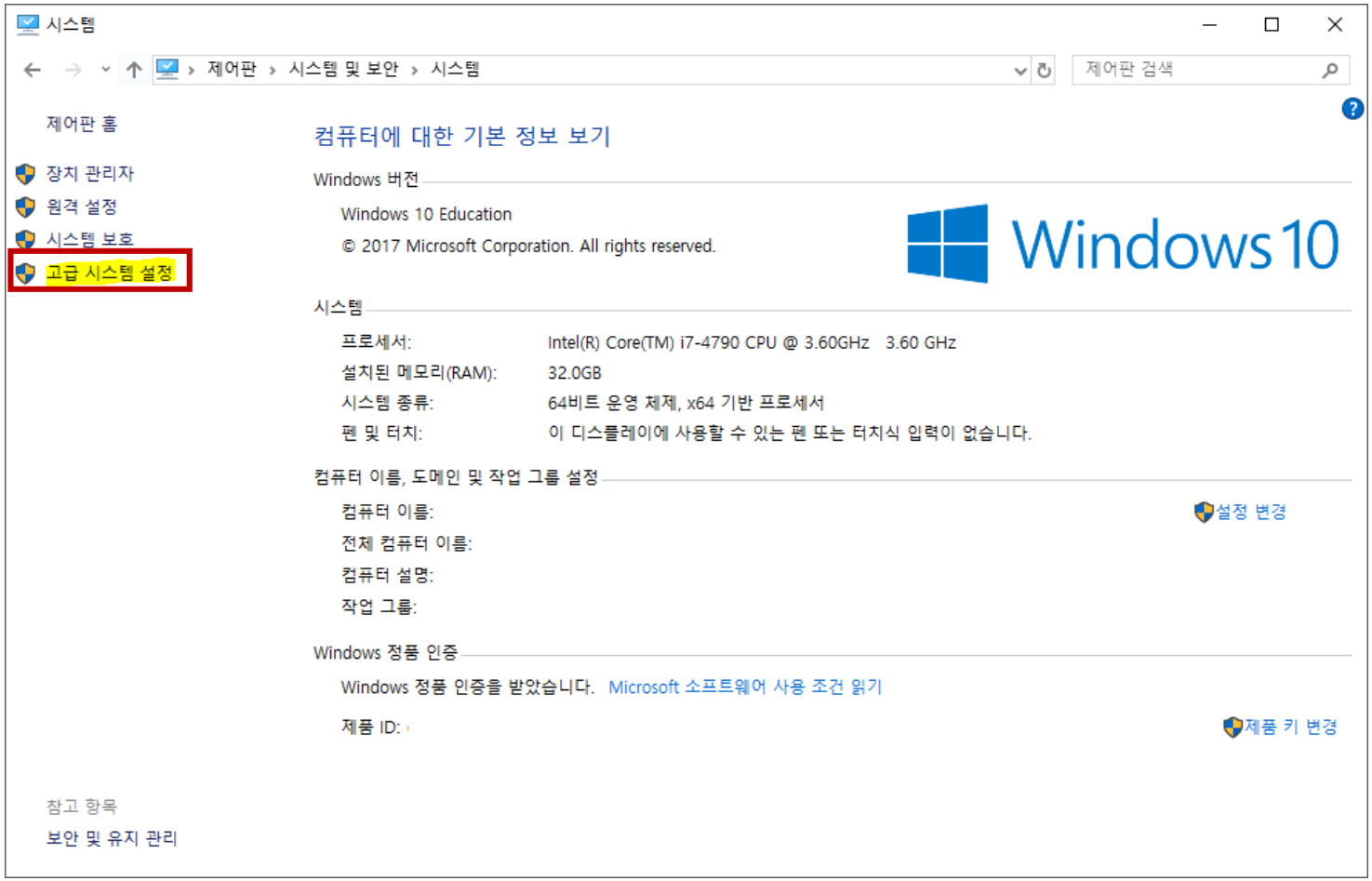
Prerequisites (2)

- Graphviz 설치 (https://graphviz.gitlab.io/download/)



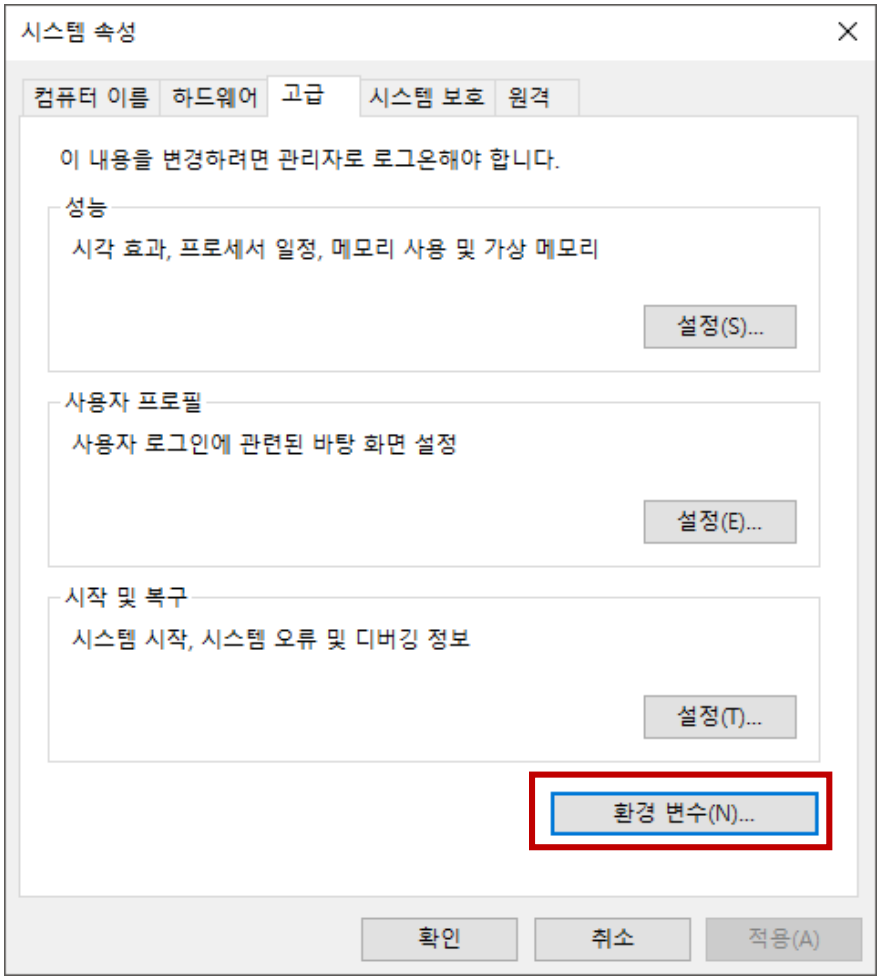
Prerequisites (2)

- Graphviz 설치 완료 후 bin 경로 PATH 추가
  - 제어판 > 시스템 및 보안 > 시스템 > 고급 시스템 설정



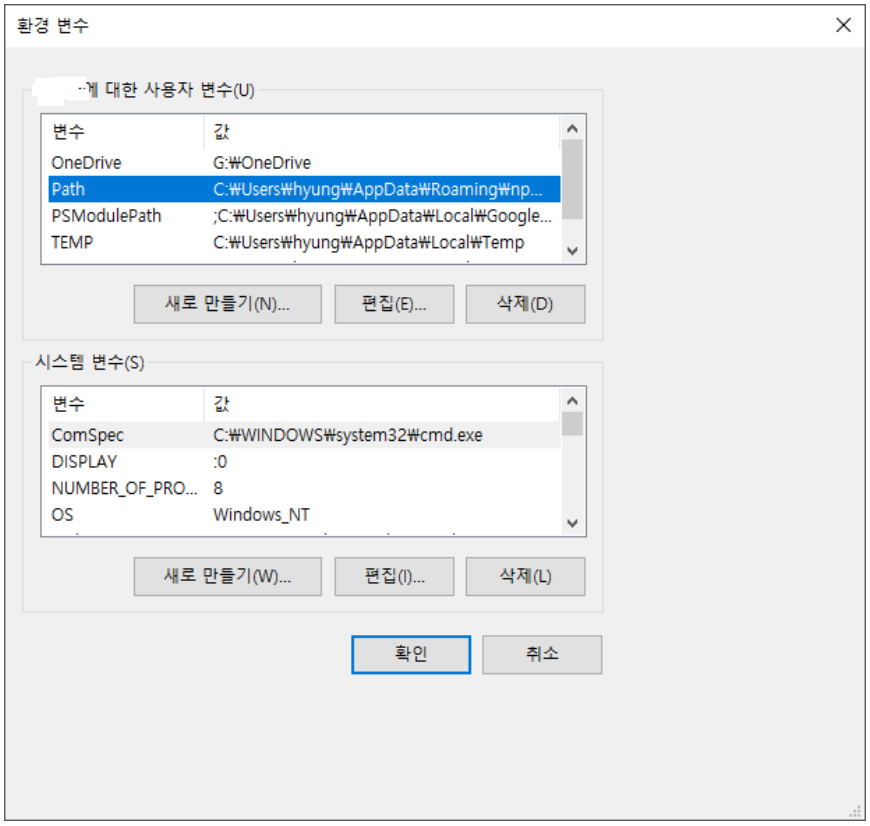
Prerequisites (2)

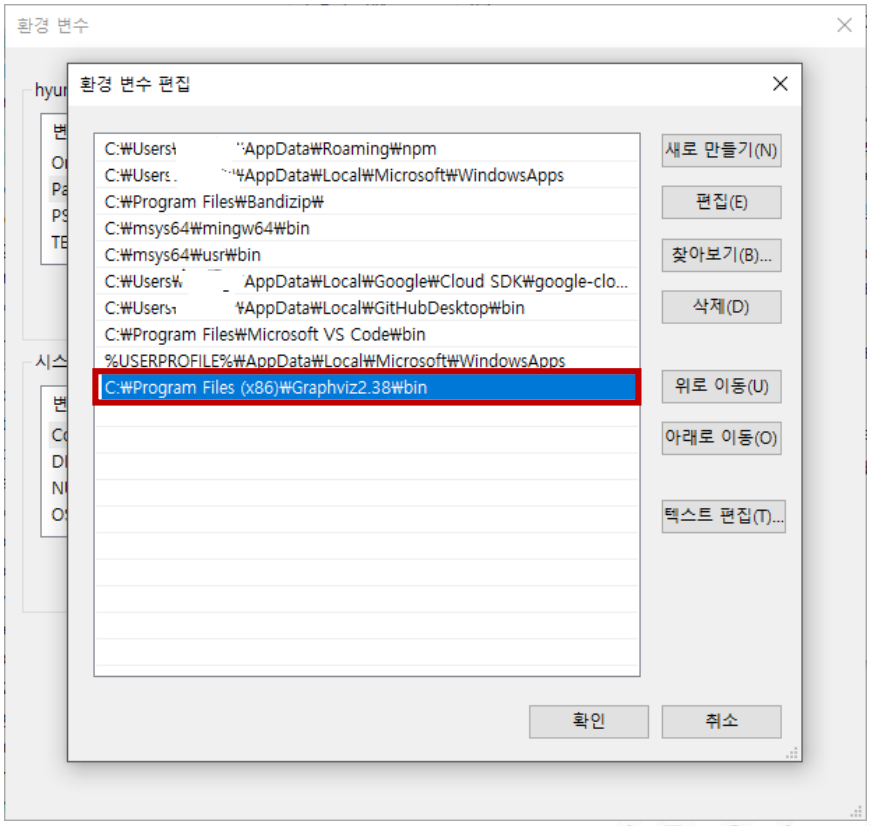
- Graphviz 설치 완료 후 bin 경로 PATH 추가
  - 고급 탭 > 환경 변수 버튼 선택



Prerequisites (2)

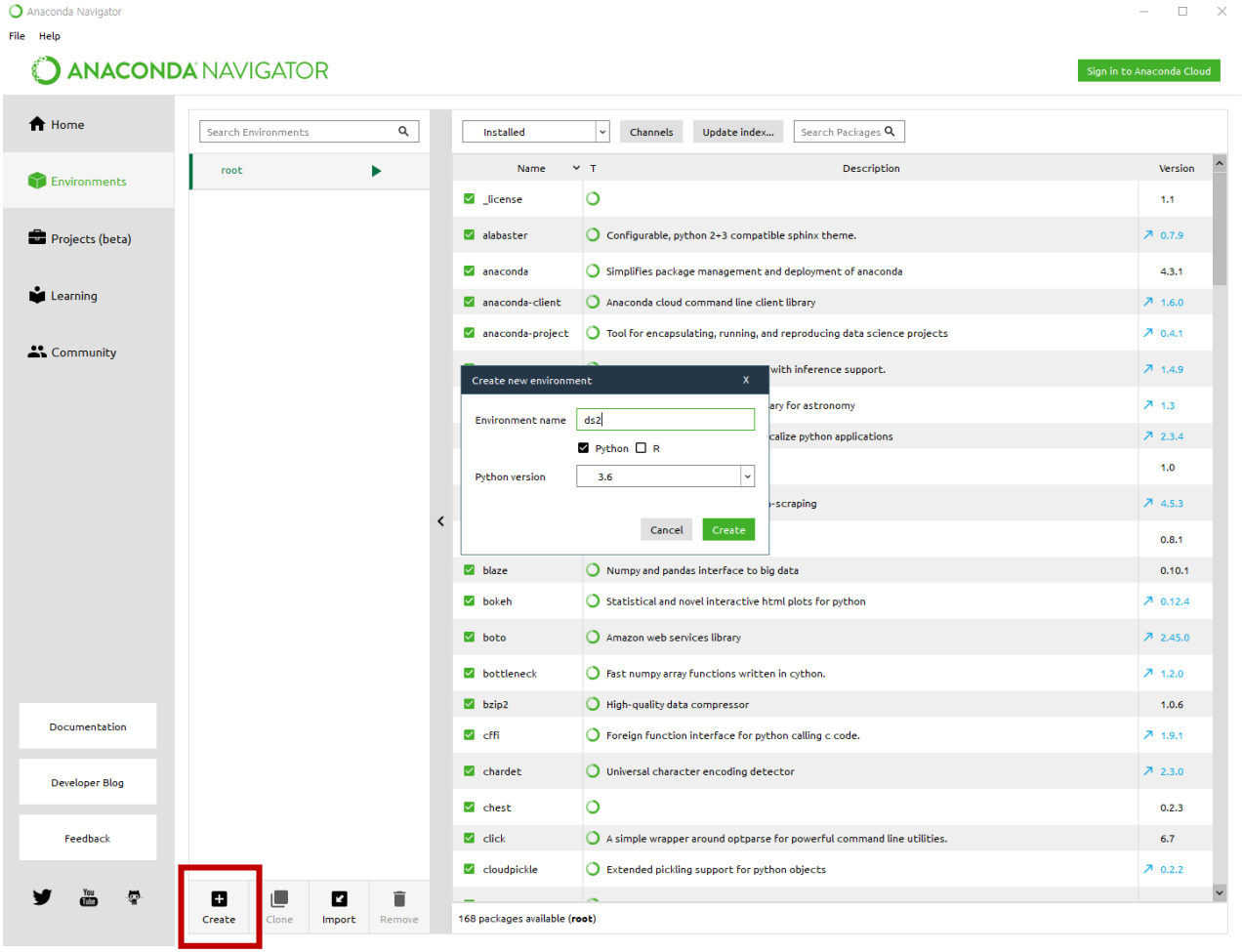
- Graphviz 설치 완료 후 bin 경로 PATH 추가
  - 사용자 변수의 Path 더블 클릭
  - 환경 변수 편집 창에 새로 만들기를 클릭하고 Graphiz 의 bin 경로 입력
    - C:\Program Files (x86)\Graphviz2.38\bin





Installation

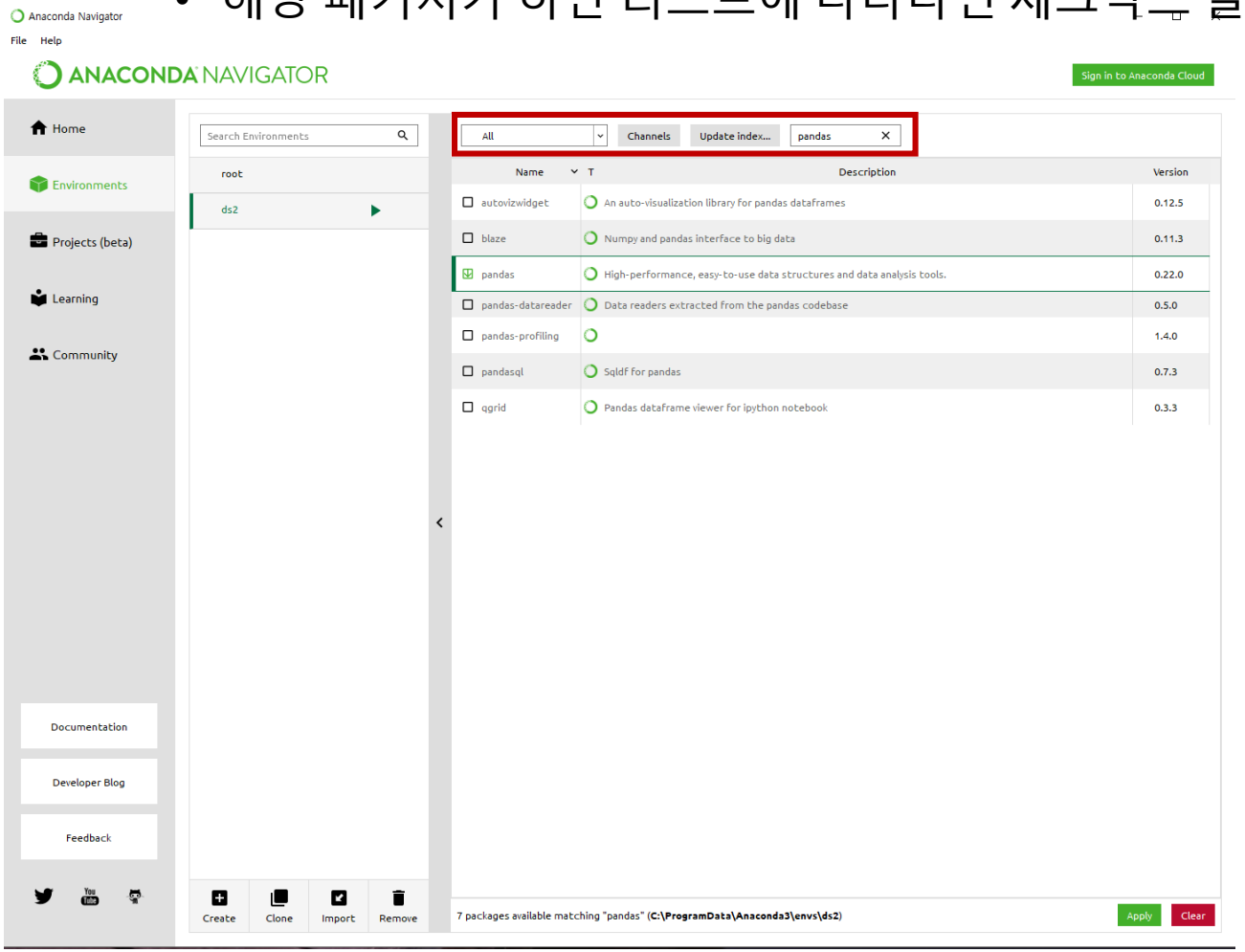
- 시작 메뉴에서 anaconda navigator를 관리자 권한으로 실행
  - Environments 탭 > Create 버튼을 누른 후 Environment name 설정
  - Environment name은 편의상 **ds2** 로 설정



ADS Lab.

Installation

- 생성한 Environment 이름(ds2)을 선택하고 필요한 패키지를 설치
  - 상단 툴바에서 All를 선택하고 검색창에 필요한 패키지 이름 입력
  - 해당 패키지가 하단 리스트에 나타나면 체크박스 클릭

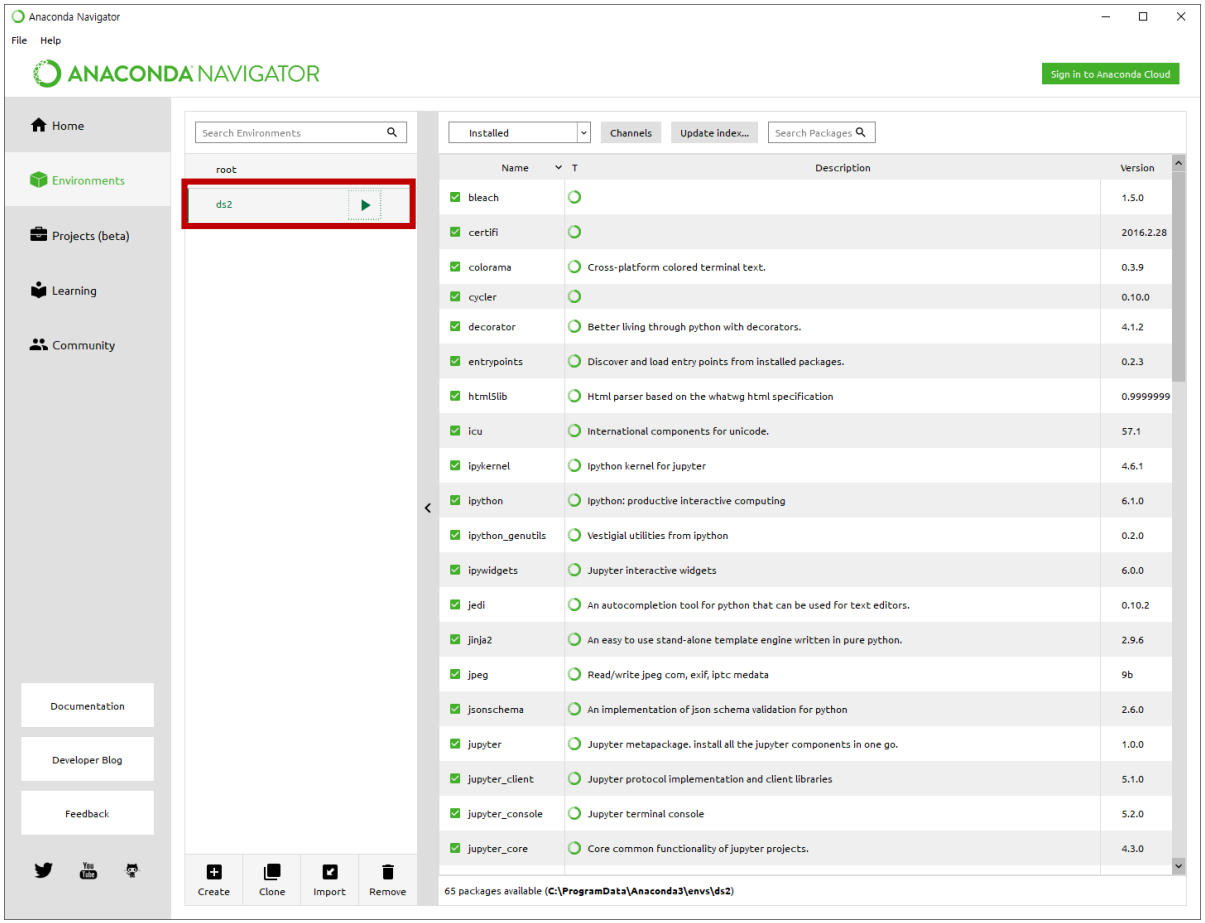


- 패키지 리스트
  - pandas
  - numpy
  - scipy
  - scikit-learn
  - jupyter
  - statsmodels
  - patsy
  - matplotlib
  - seaborn
  - xlrd
- 모두 선택 후 Apply 클릭

ADS Lab.

Installation

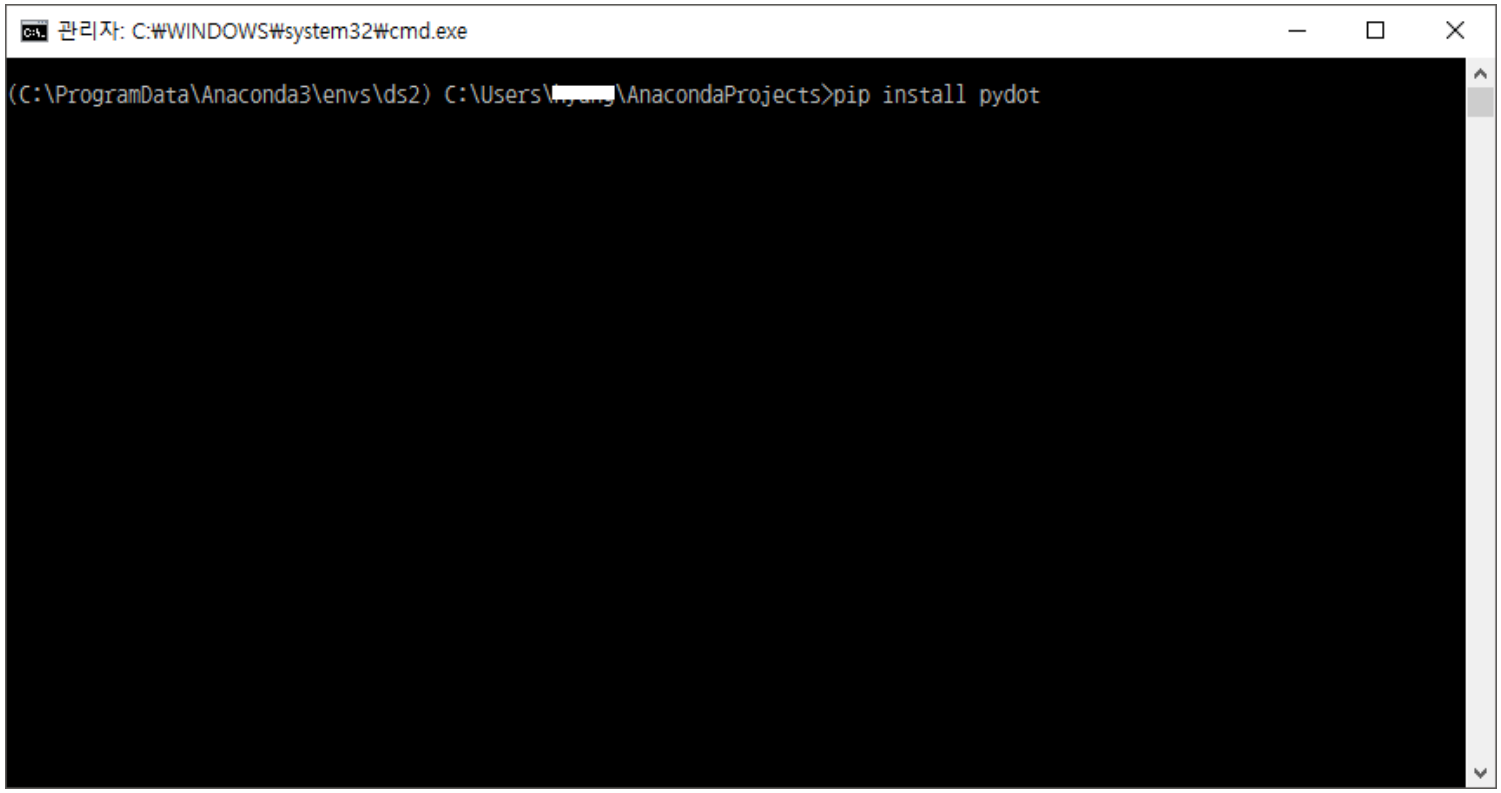
- 설치가 완료된 이후 필요한 패키지가 모두 설치되었는지 확인
  - 상단 툴바에서 Installed를 선택하고 패키지가 하단 리스트에 있는지 확인
- 확인 완료 후 environment 이름 옆 ► 클릭해서 'Open Terminal' 선택



- 패키지 리스트
  - pandas
  - numpy
  - scipy
  - scikit-learn
  - jupyter
  - statsmodels
  - patsy
  - matplotlib
  - seaborn
  - xlrd

Installation

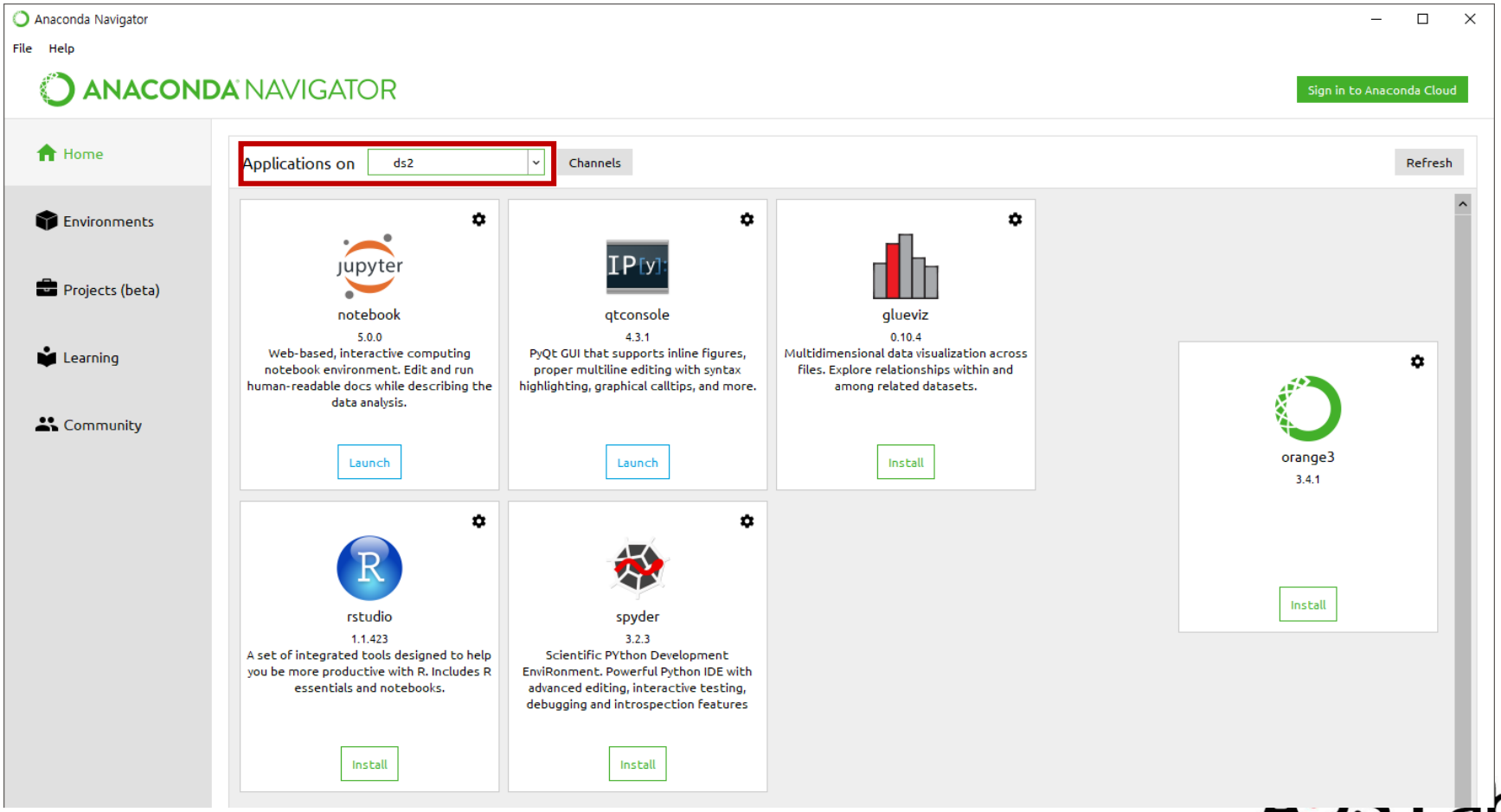
- Console 창에 다음 명령을 수행하여 추가 패키지 설치
  - `pip install pydot`
- 설치가 정상적으로 완료되면 창을 닫음



실습 수행

Run Jupyter

- 시작 메뉴에서 Anaconda Navigator를 선택해서 실행
- Home 탭으로 이동해 지정한 Environment 이름을 선택하고 jupyter notebook 메뉴의 Launch 버튼 클릭



Coding on Jupyter notebook

- 브라우저가 실행되면 Notebooks 폴더로 이동해서 .ipynb 파일을 선택
- notebook editor 창에서 python 코딩과 실습 수행 가능
  - 노트북 사용법

Chapter 3 - Linear Regression

- Load Datasets
- 3.1 Simple Linear Regression
- 3.2 Multiple Linear Regression
- 3.3 Other Considerations in the Regression Model

In [ ]:

```
# %load ../standard_import.txt
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import axes3d
import seaborn as sns

from sklearn.preprocessing import scale
import sklearn.linear_model as skl_lm
from sklearn.metrics import mean_squared_error, r2_score
import statsmodels.api as sm
import statsmodels.formula.api as smf

%matplotlib inline
plt.style.use('seaborn-white')
```