

1장

머신러닝 소개

이홍석 (hsyi@kisti.re.kr)

한국과학기술정보연구원 슈퍼컴퓨팅응용센터





Contents

목차

 Contents 1	K-Means 개요
 Contents 2	K-Means의 이해
 Contents 3	K-Means 상가고객 군집
 Contents 4	K-Means 적용 붓꽃
 Contents 5	GMM 군집화
 Contents 6	DBSCAN 군집화

01. 강의소개

이홍석 (hsyi@kisti.re.kr)





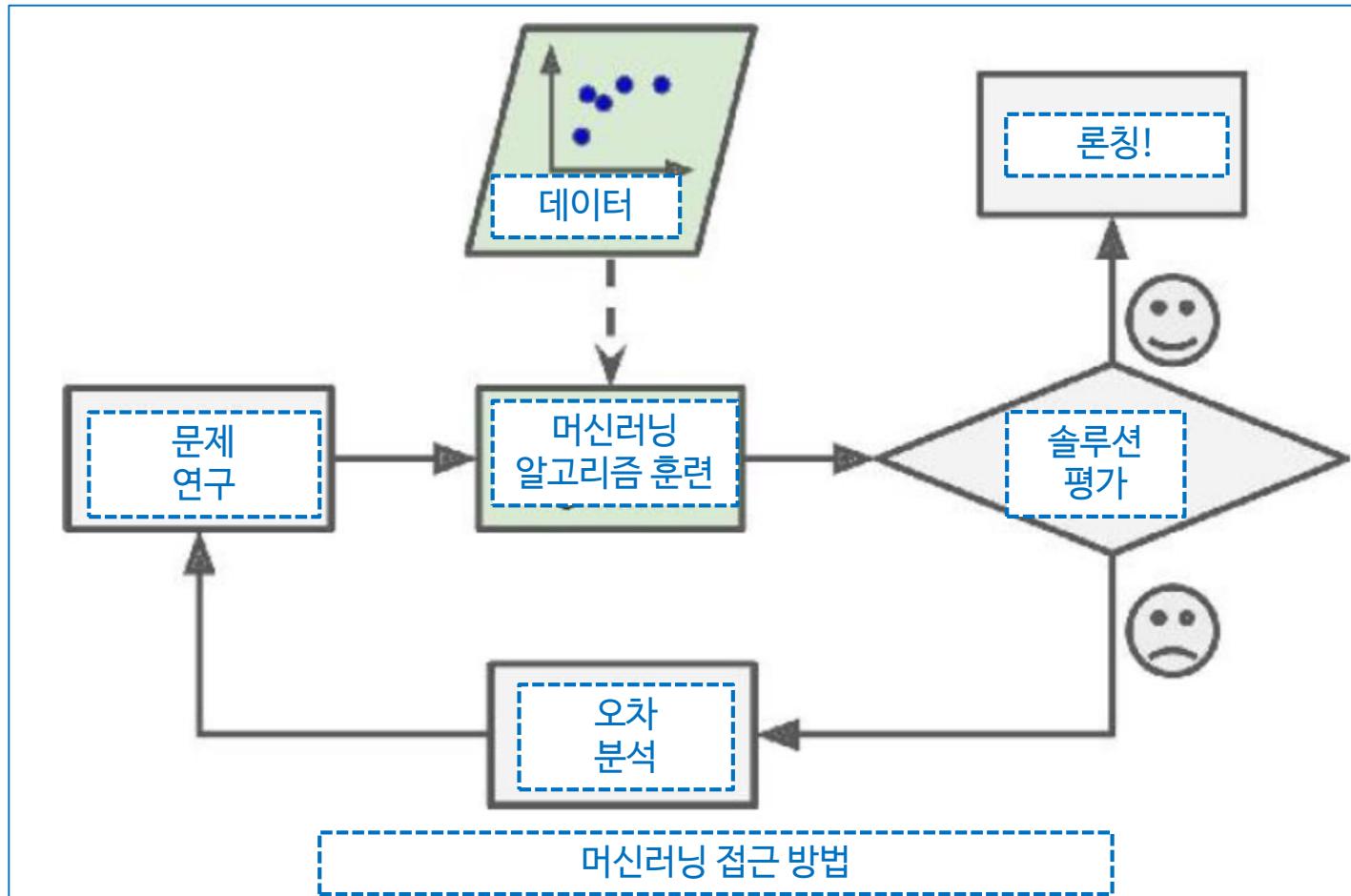
머신러닝이란?

- **머신러닝**

- 데이터로 부터 학습하도록 컴퓨터를 프로그래밍하는 과학
- 어떤 작업 T에 대한 컴퓨터 프로그램의 성능을 P로 측정했을 때 경험 E로 인해 성능이 향상됐다면, 이 컴퓨터 프로그램은 작업 T와 성능 측정 P에 대한 경험 E로 학습한 것

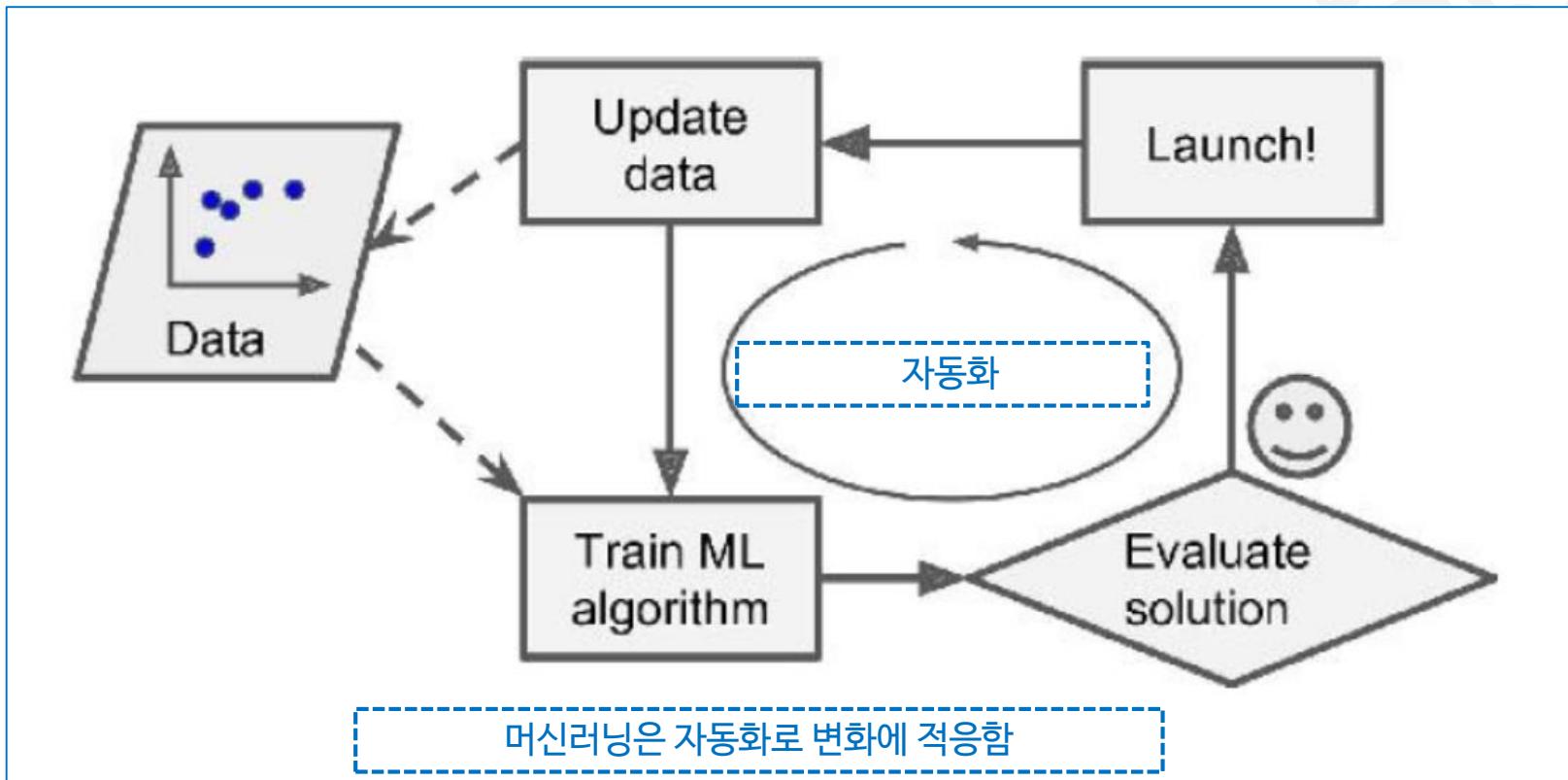


왜 머신러닝을 사용하는가? (1)



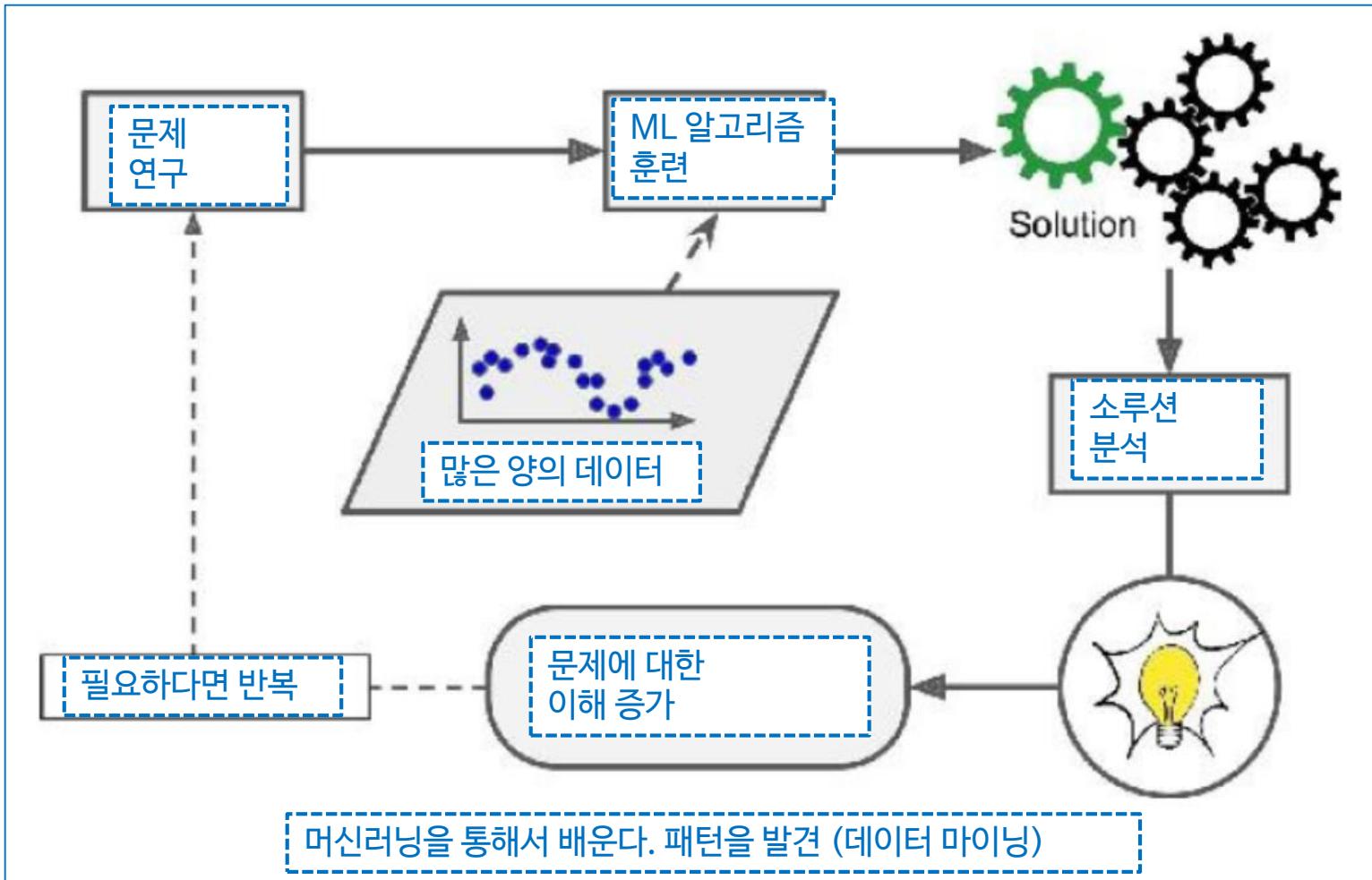


왜 머신러닝을 사용하는가? (2)





왜 머신러닝을 사용하는가? (3)

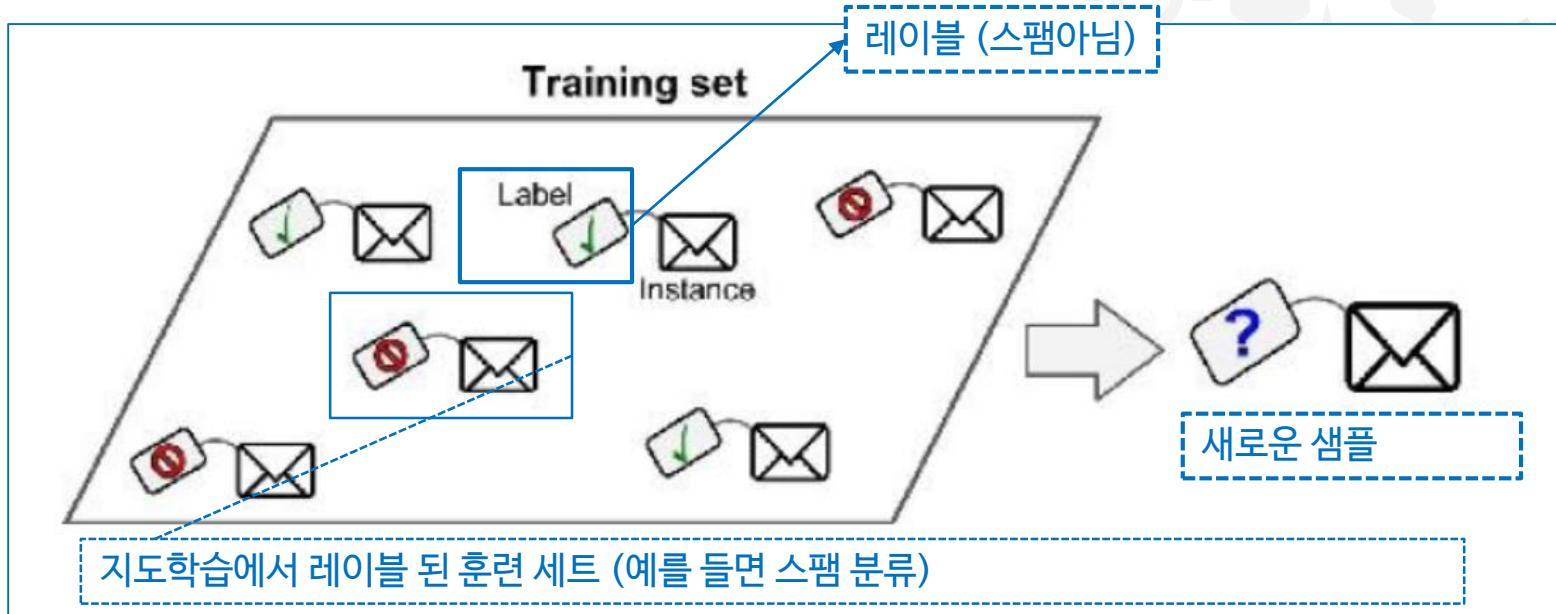




머신러닝 시스템의 종류(1)

- 지도학습

- 알고리즘에 주입하는 훈련 데이터에 레이블이라는 원하는 답이 포함됨





지도학습 알고리즘

- K-NN (k-최근접 이웃)
- 선형회귀
- 로지스틱 회귀
- 서포트 벡터 머신 (SVM)
- 결정 트리
- 램덤 포레스트
- 신경망

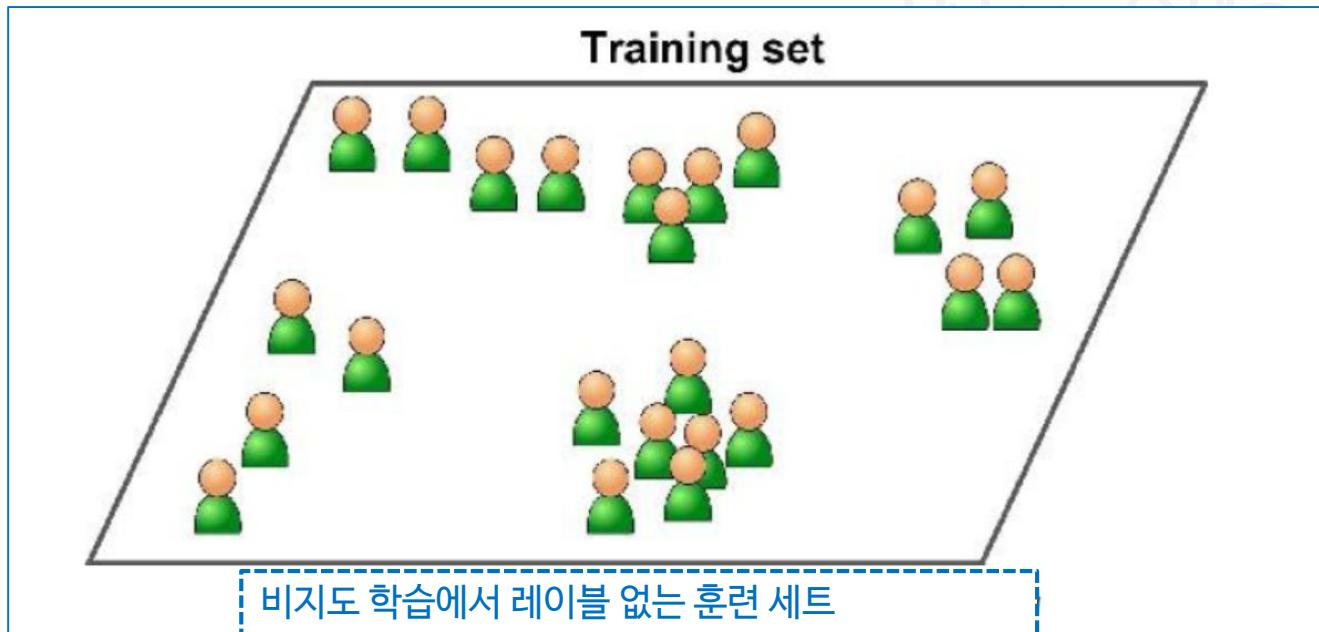




비지도학습

- **비지도학습(Unsupervised Learning)**

- 훈련데이터에 레이블이 없다.
- 시스템이 아무런 도움없이 학습을 해야 한다.
- (예) 중간 시험은 봤는데 채점은 하지않고 시험만 계속 봄



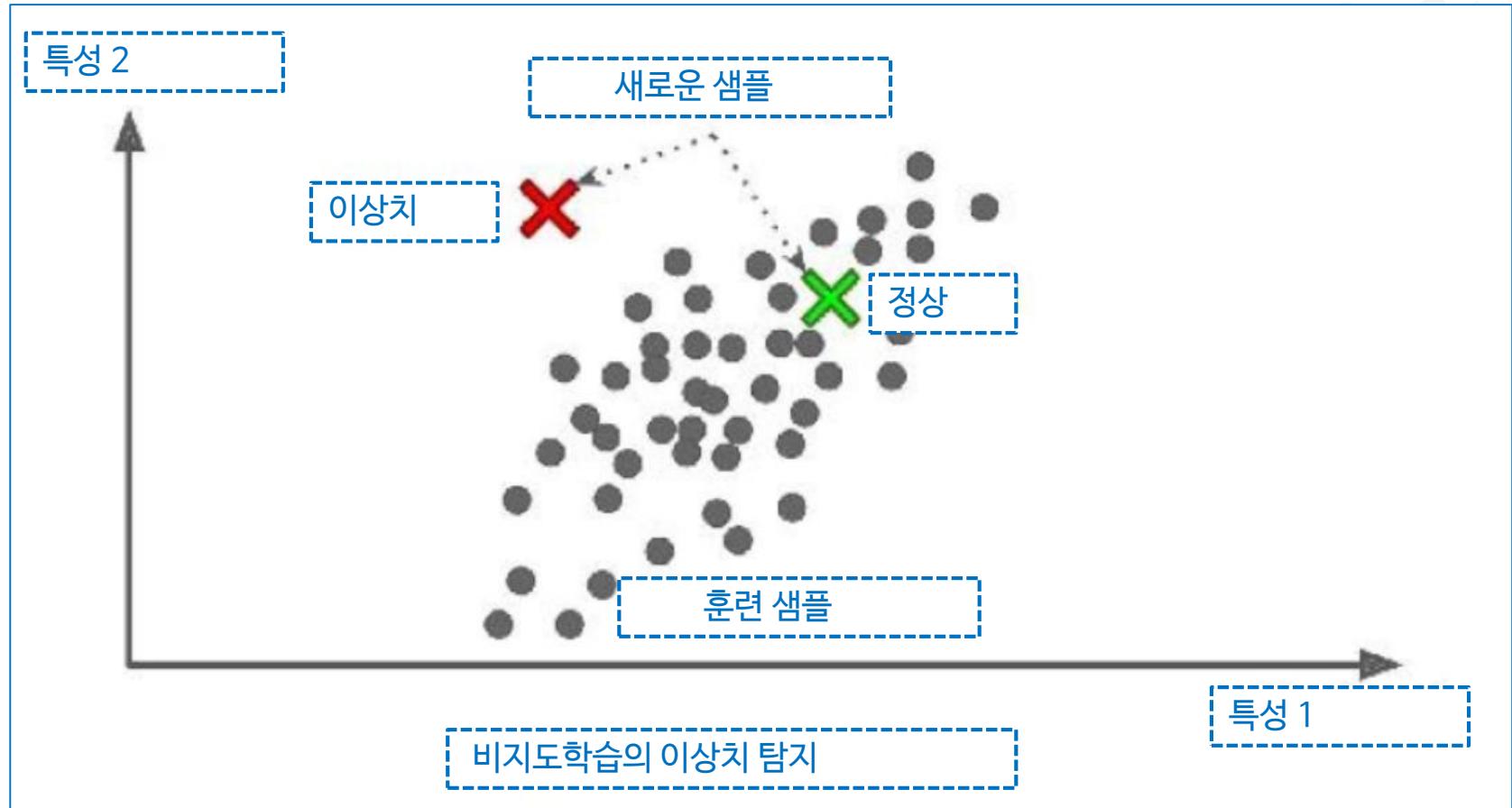


비지도 학습의 알고리즘

- 군집 (Clustering)으로 K-평균
 - K-Means
- 차원축소로 주성분분석 (PCA)
 - PCA (Principal Component Analysis)
- 이상치 탐지 (Anomaly detection)
 - 부정 거래를 막기 위해 이상한 신용카드 거래를 탐지
 - 제조 결함을 잡아내고, 학습 이전에 데이터 셋에서 이상한 값을 자동으로 제거하는 것
 - 도로에서 주행하는 차량이 차선 위반, 속도위반을 자동으로 검출하는 기술



비지도학습의 이상치 탐지

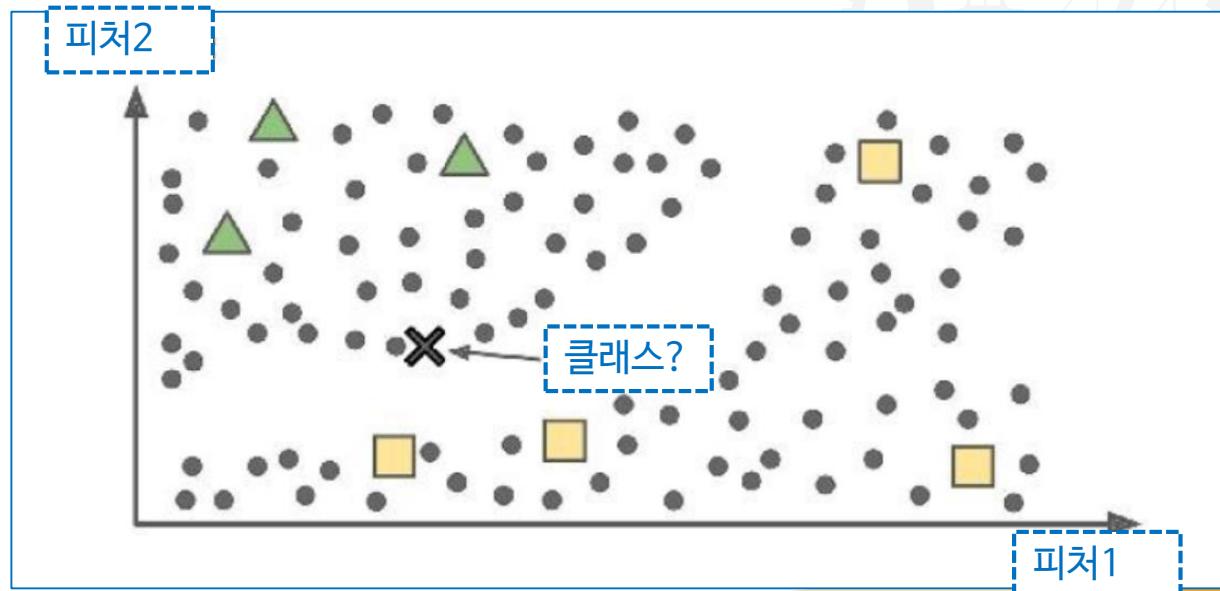




준지도학습

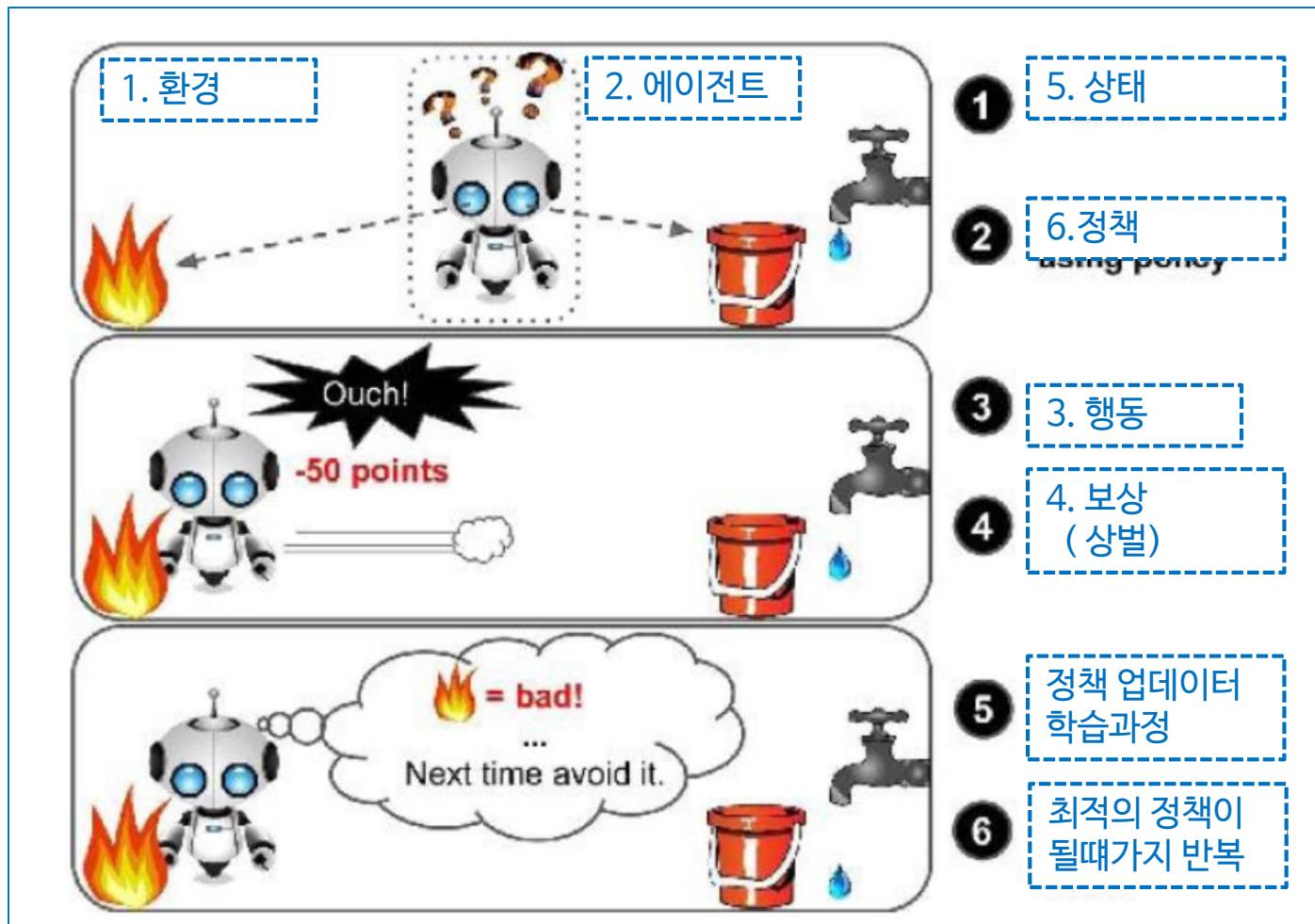
- **준지도학습**

- 보통은 레이블이 없는 데이터가 많고, 레이블이 있는 데이터는 아주 조금이다.
- 지도학습과 비지도학습의 조합으로 이루어짐
- (예) 신경망에서 DBN (Deep Belief Network)는 RBM (Restricted Boltzmann Machine) 비지도학습에 기초함
- RBM으로 순차적으로 훈련된 다음, 전체 시스템이 지도학습 방식으로 조정





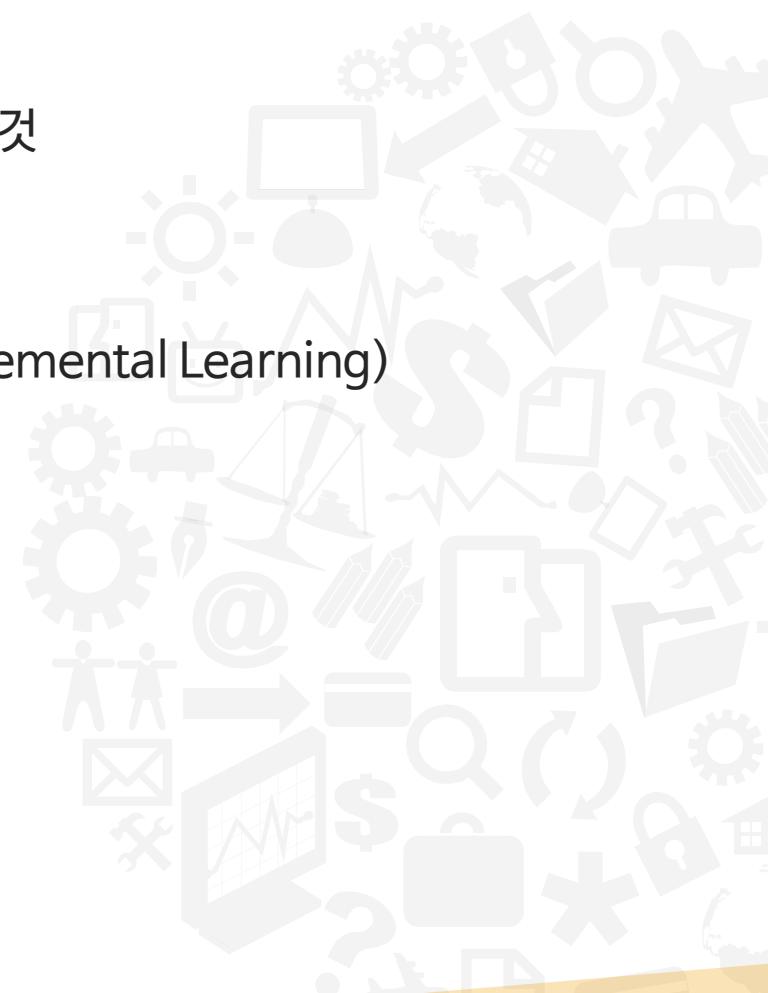
강화학습





머신러닝에서 데이터 학습 방법 (1)

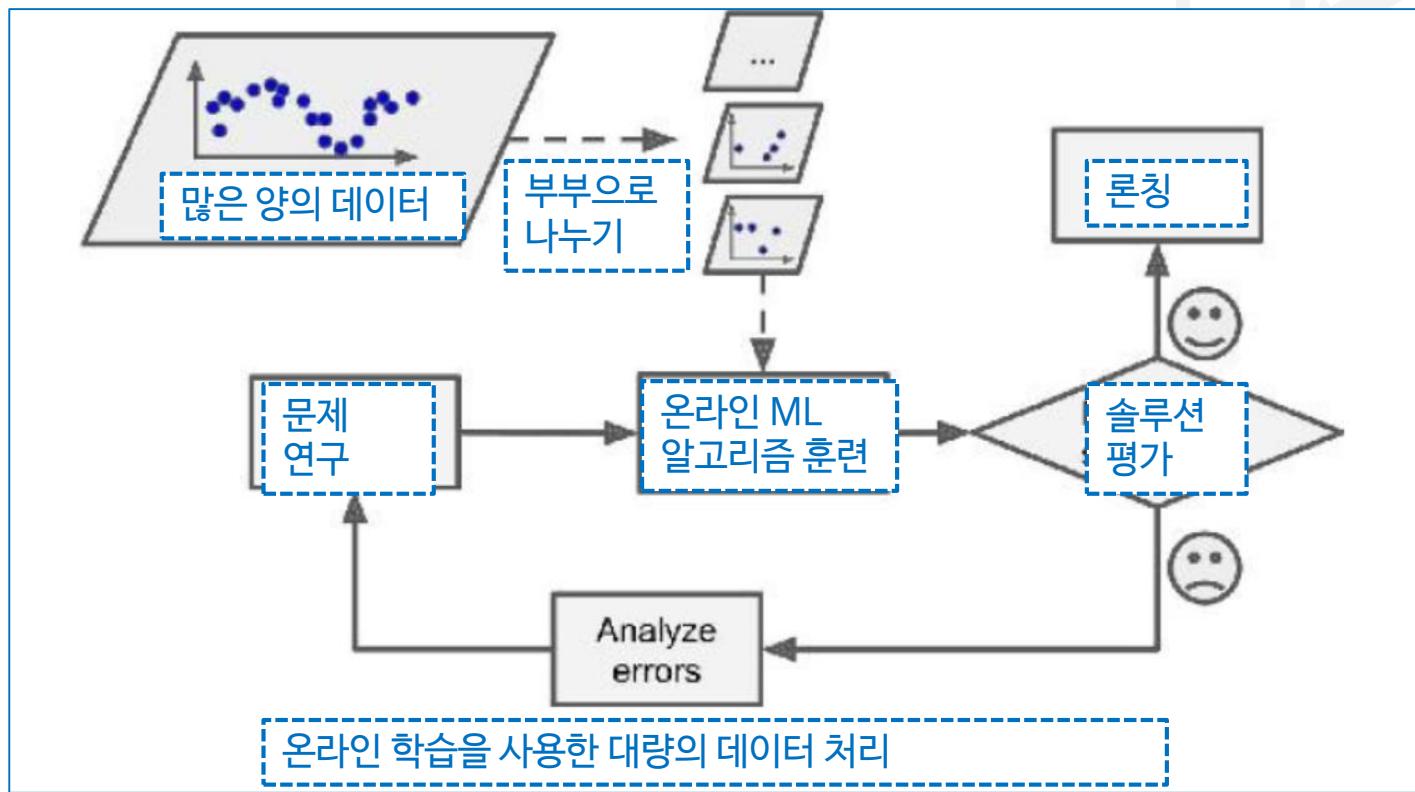
- **오프라인 학습은 배치 학습**
 - Batch 학습은 모든 데이터를 한번에 학습하는 것
 - 큰 메모리 필요, 긴 계산시간 등
- **온라인 학습**
 - 미니 배치 학습으로, 실시간 학습에 적합 (Incremental Learning)
 - 단점에 나쁜 데이터를 학습하면 곧 망가짐





머신러닝에서 데이터 학습 방법 (2)

- 온라인 학습을 사용한 대량의 데이터 처리





머신러닝에 사용하는 패키지

- Scikit-learn
- Pandas
- Numpy
- Seaborn
- Matplotlib





Scikit-learn 사이킷런

<https://scikit-learn.org/stable/>

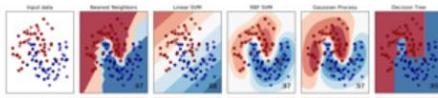
The screenshot shows the official scikit-learn documentation page. At the top, there's a navigation bar with links for 'Install', 'User Guide', 'API', 'Examples', and 'More'. Below the header, the title 'scikit-learn' is displayed in large white letters on a blue background, followed by 'Machine Learning in Python'. There are three orange buttons: 'Getting Started', 'What's New in 0.22.1', and 'GitHub'. To the right, a yellow sidebar lists the project's benefits: 'Simple and efficient tools for predictive data analysis', 'Accessible to everybody, and reusable in various contexts', 'Built on NumPy, SciPy, and matplotlib', and 'Open source, commercially usable - BSD license'. The main content area is divided into three sections: 'Classification', 'Regression', and 'Clustering', each with a brief description, applications, algorithms, and associated visualizations.

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.

Algorithms: SVM, nearest neighbors, random forest, and more...



Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.

Algorithms: SVR, nearest neighbors, random forest, and more...



Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, spectral clustering, mean-shift, and more...



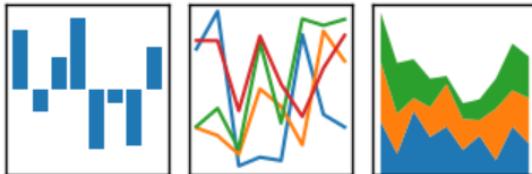


Pandas

<https://pandas.pydata.org/>

pandas

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



[home](#) // [about](#) // [get pandas](#) // [documentation](#) // [community](#) // [talks](#) // [donate](#)

Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

pandas is a [NumFOCUS](#) sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to [donate](#) to the project.

A Fiscally Sponsored Project of
NUMFOCUS
OPEN CODE = BETTER SCIENCE

VERSIONS

Release Candidate

1.0.0 - January 2020

[download](#) // [docs](#) // [pdf](#)

Release

0.25.3 - November 2019

[download](#) // [docs](#) // [pdf](#)

Development

1.0.0 - September 2019

[github](#) // [docs](#)

Previous Releases

0.25.2 - [download](#) // [docs](#) // [pdf](#)



NumPy

NumPy

NumPy.org

NumPy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the [BSD license](#), enabling reuse with few restrictions.

Getting Started



Matplotlib

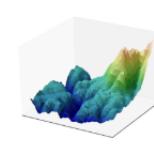
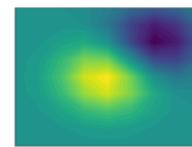
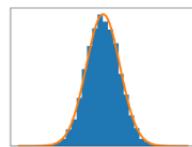
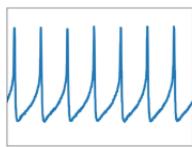
<https://matplotlib.org/>



[Installation](#) [Documentation](#) [Examples](#) [Tutorials](#) [Contributing](#)

[home](#) | [contents](#) »

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.



Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code. For examples, see the [sample plots](#) and [thumbnail gallery](#).

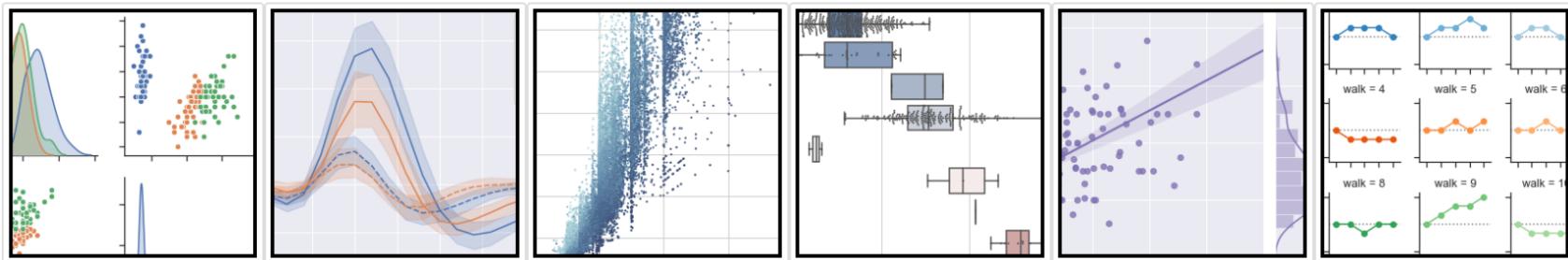
For simple plotting the `pyplot` module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.



Seaborn

<https://seaborn.pydata.org/>

seaborn: statistical data visualization



Seaborn is a Python data visualization library based on [matplotlib](#). It provides a high-level interface for drawing attractive and informative statistical graphics.

For a brief introduction to the ideas behind the library, you can read the [introductory notes](#). Visit the [installation page](#) to see how you can download the package. You can browse the [example gallery](#) to see what you can do with seaborn, and then check out the [tutorial](#) and [API reference](#) to find out how.

To see the code or report a bug, please visit the [github repository](#). General support issues are most at home on [stackoverflow](#), where there is a seaborn tag.

Contents

- [Introduction](#)
- [Release notes](#)
- [Installing](#)
- [Example gallery](#)
- [Tutorial](#)
- [API reference](#)

Features

- Relational: [API](#) | [Tutorial](#)
- Categorical: [API](#) | [Tutorial](#)
- Distributions: [API](#) | [Tutorial](#)
- Regressions: [API](#) | [Tutorial](#)
- Multiples: [API](#) | [Tutorial](#)
- Style: [API](#) | [Tutorial](#)
- Color: [API](#) | [Tutorial](#)

02. 실습용 데이터 셋트

이홍석 (hsyi@kisti.re.kr)





데이터 다운로드 사이트



Machine Learning Repository

[Center for Machine Learning and Intelligent Systems](#)

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 488 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable index or visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please feel free to [contact the Repository librarians](#).

Supported By:



In Collaboration With:





가장 있기 있는 데이터 세트

Newest Data Sets:

10-06-2019:		WISDM Smartphone and Smartwatch Activity and Biometrics Dataset
09-30-2019:		Hepatitis C Virus (HCV) for Egyptian patients
09-23-2019:		QSAR fish toxicity
09-23-2019:		QSAR aquatic toxicity
09-21-2019:		Online Retail II
09-20-2019:		Human Activity Recognition from Continuous Ambient Sensor Data
09-20-2019:		Beijing Multi-Site Air-Quality Data
09-20-2019:		MEx
07-30-2019:		PPG-DaLiA

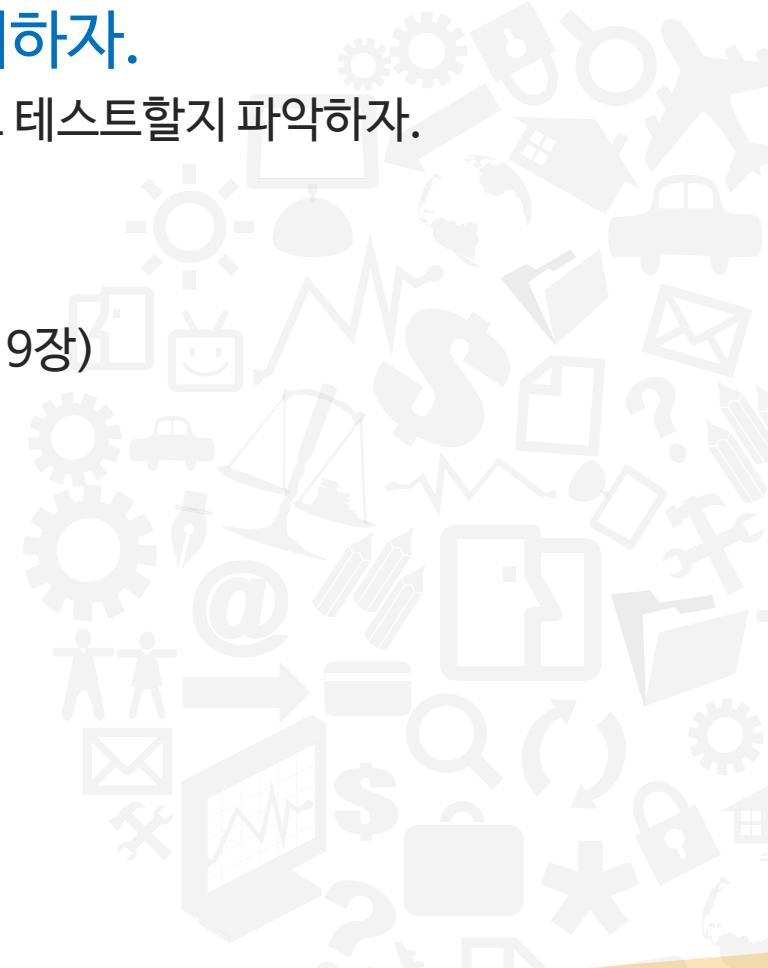
Most Popular Data Sets (hits since 2007):

3091299:		Iris
1707888:		Adult
1324210:		Wine
1123015:		Wine Quality
1122923:		Heart Disease
1114183:		Car Evaluation
1107500:		Breast Cancer Wisconsin (Diagnostic)
1097941:		Bank Marketing
933570:		Human Activity Recognition Using Smartphones



데이터 셋트 이해하기

- **공통으로 사용하는 데이터 셋트를 이해하자.**
 - 데이터의 특성을 잘 이해하고, 무엇을 훈련하고 테스트할지 파악하자.
- **주로 사용하는 데이터 셋트**
 - 타이타닉 데이터 생존 데이터 (2장, 3장, 4장)
 - 붓꽃(IRIS) 데이터 (2장, 5장, 8장, 9장)
 - 피마 인디언 당뇨병 데이터 (3장)
 - 위스콘신 유방암 데이터 (6장, 8장)
 - 산탄데르 고객 만족 데이터 (10장)
 - 캐글 신용카드 사기 검출 데이터 (9장)
 - 보스턴 주택 가격 데이터 (7장, 8장)
 - 자전거 대여 수요 데이터 (8장,
 - 사용자 행동 인식 데이터 (6장)





데이터를 잘 이해하자

날짜	데이터	담당자
	타이타닉	
	붓꽃(IRIS)	
	위스콘신 유방암	
	보스턴 주택 가격	
	사용자 행동 인식	
	자전거 대여 수요	
	신용카드 사기 검출	
	산탄데르 고객 만족	
	피마 인디언 당뇨병	



1. 타이타닉 생존 데이터

Getting Started Prediction Competition

Titanic: Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

 Kaggle · 15,671 teams · Ongoing

[Overview](#) [Data](#) [Notebooks](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#)

Overview

Description  Ahoy, welcome to Kaggle! You're in the right place.

Evaluation This is the legendary Titanic ML competition – the best, first challenge for you to di



2. 붓꽃 (IRIS)

<https://archive.ics.uci.edu/ml/datasets/iris>

Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Iris Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Famous database; from Fisher, 1936



Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	3091300



3. 신용카드 고객 데이터 세트

<https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>

This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods

The screenshot shows the UCI Machine Learning Repository homepage. At the top left is the UCI logo with a stylized antelope illustration. Below it is the text "Machine Learning Repository" and "Center for Machine Learning and Intelligent Systems". On the right side, there is a navigation bar with links like "About", "Contact", and "Logout".

default of credit card clients Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: This research aimed at the case of customers' default payments in Taiwan and compares the predictive accuracy of probability of default among six data mining methods.

Data Set Characteristics:	Multivariate	Number of Instances:	30000	Area:	Business
Attribute Characteristics:	Integer, Real	Number of Attributes:	24	Date Donated	2016-01-26
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	431729



4. 자전거 대여 수요 데이터

<https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset>

This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information.



Bike Sharing Dataset Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: This dataset contains the hourly and daily count of rental bikes between years 2011 and 2012 in Capital bikeshare s

Data Set Characteristics:	Univariate	Number of Instances:	17389	Area:	Social
Attribute Characteristics:	Integer, Real	Number of Attributes:	16	Date Donated	2013-12-20
Associated Tasks:	Regression	Missing Values?	N/A	Number of Web Hits:	441354



5. 위스콘신 유방암 데이터 세트

[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))

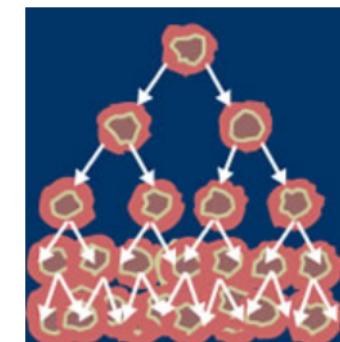


The logo for the UCI Machine Learning Repository. It features the letters "UCI" in yellow on a white background, followed by a blue silhouette of a hand pointing to the right. Below this is the text "Machine Learning Repository" in yellow, and "Center for Machine Learning and Intelligent Systems" in smaller blue text.

Breast Cancer Wisconsin (Diagnostic) Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Diagnostic Wisconsin Breast Cancer Database



Data Set Characteristics:	Multivariate	Number of Instances:	569	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	32	Date Donated	1995-11-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	1107508



6. 캐글 산탄데르 고객 만족 데이터

▽ kaggle.com/c/santander-customer-satisfaction/data ☆



Santander Customer Satisfaction

Which customers are happy customers?
\$60,000 · 5,122 teams · 4 years ago

Overview Data Notebooks Discussion Leaderboard Rules Late Submission

Data Description

You are provided with an anonymized dataset containing a large number of numeric variables. The "TARGET" column is the variable to predict. It equals one for unsatisfied customers and 0 for satisfied customers.

The task is to predict the probability that each customer in the test set is an unsatisfied customer.

File descriptions

- train.csv - the training set including the target
- test.csv - the test set without the target
- sample_submission.csv - a sample submission file in the correct format



7. 캐글 신용카드 사기 검출

A screenshot of a Kaggle dataset page. The header includes the Kaggle logo, a search bar, and navigation links for Competitions, Datasets, Notebooks, Discussion, Courses, and more. The main title is "Credit Card Fraud Detection" with a subtitle "Anonymized credit card transactions labeled as fraudulent or genuine". Below the title is a thumbnail for the dataset, which is a collage of various credit cards. A yellow circular badge indicates there are 4580 files. The dataset was created by the "Machine Learning Group - ULB" and was last updated 2 years ago (Version 3). Below the thumbnail are tabs for Data (selected), Tasks (9), Kernels (2,334), Discussion (46), Activity, and Metadata. There is also a "Download (144 MB)" button and a "New Notebook" button. At the bottom, there are sections for Usability (8.5), License (Database: Open Database, Contents: Database Contents), and Tags (computing, finance, crime, machine learning, credit cards).



8. 보스턴 주택 가격 데이터

Dataset

UCI ML Datasets

Boston Housing Dataset

Khashayar Baghizadeh Hosseini • updated 3 years ago (Version 1)

Data Tasks Kernels (8) Discussion (1) Activity Metadata Download (35 KB) New Notebook

Usability 8.8 License Database: Open Database, Contents: Database Contents Tags computing, natural and physical sciences, education, society, mathematics

Description

Context



9. 사용자 행동 인식 데이터

<https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones>

Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.



Human Activity Recognition Using Smartphones Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.

Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	10299	Area:	Computer
Attribute Characteristics:	N/A	Number of Attributes:	561	Date Donated	2012-12-10
Associated Tasks:	Classification, Clustering	Missing Values?	N/A	Number of Web Hits:	933580



프로젝트 (숙제) 소개

- **가장 잘 이해하고 있는 데이터 세트 1~2개 선정**
 - 수업에서는 공통 데이터로 붓꽃 (IRIS)을 사용
 - 옵션으로 추가로 1개의 데이터를 선정하여 충분히 이해함
 - 2가지 데이터에 대하여 머신러닝 기법을 적용한 기법 이해함
- **프로젝트**
 - 이용하여 수업에서 배운 모든 방법론을 적용한 1개의 코드를 완성하는 것

03. 실습환경

이홍석 (hsyi@kisti.re.kr)





Installation

[!\[\]\(0ced6b67fbff79b684b4619a328310ee_img.jpg\) TensorFlow](#)

Install Learn API Resources Community Why TensorFlow

[Install TensorFlow](#)

Packages
pip
Docker

Additional setup
GPU support
Problems

Build from source
Linux / macOS
Windows
Raspberry Pi

Language bindings



Missed TensorFlow World? Check out the recap.

[Learn more](#)

Install TensorFlow 2

TensorFlow is tested and supported on the following 64-bit systems:

- Ubuntu 16.04 or later
- Windows 7 or later
- macOS 10.12.6 (Sierra) or later (no GPU support)
- Raspbian 9.0 or later

Download a package

Install TensorFlow with Python's *pip* package manager.

```
# Requires the latest pip  
pip install --upgrade pip
```



Tensorflow : pip 이용한 설치

TensorFlow Install Learn ▾ API ▾ Resources ▾ Community Why TensorFlow ▾

TensorFlow > Install

Install TensorFlow with pip

TensorFlow 2 packages are available

- `tensorflow` —Latest stable release for CPU-only
- `tensorflow-gpu` —Latest stable release with **GPU support** (*Ubuntu and Windows*)
- `tf-nightly` —Preview build (*unstable*). Ubuntu and Windows include **GPU support**.

Older versions of TensorFlow

For the 1.15 release, CPU and GPU support are included in a single package:

- `tensorflow==1.15` —The final 1.x release. Ubuntu and Windows include **GPU support**

For releases 1.14 and older, CPU and GPU packages are separate:

- `tensorflow==1.14` —Release for CPU-only
- `tensorflow-gpu==1.14` —Release with **GPU support** (*Ubuntu and Windows*)



Windows의 시스템 정보 확인

홈

설정 검색

시스템

- 저장 공간
- 태블릿 모드
- 멀티태스킹
- PC에 화면 표시
- 공유 환경
- 클립보드
- 원격 데스크톱
- 정보

정보

디바이스 사양

SAMSUNG PC

디바이스 이름 DESKTOP-LPICGOA

프로세서 Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz
1.99 GHz

설치된 RAM 16.0GB

디바이스 ID 82E601BF-9BBD-4B6F-8DCE-499AD1CDC902

제품 ID 00325-95800-00000-AAOEM

시스템 종류 64비트 운영 체제, x64 기반 프로세서

펜 및 터치 펜과 10개 터치 포인트 지원

이 PC의 이름 바꾸기

Windows 사양

에디션 Windows 10 Home

버전 1903

설치 날짜 2019-08-01



Anaconda를 이용한 설치

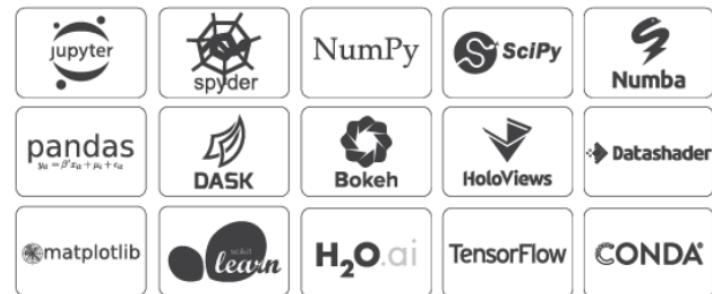
Anaconda Distribution

The World's Most Popular Python/R Data Science Platform

Download

The open-source **Anaconda Distribution** is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 15 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling *individual data scientists* to:

- Quickly download 1,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with **Conda**
- Develop and train machine learning and deep learning models with **scikit-learn**, **TensorFlow**, and **Theano**
- Analyze data with scalability and performance with **Dask**, **NumPy**, **pandas**, and **Numba**
- Visualize results with **Matplotlib**, **Bokeh**, **Datashader**, and **Holoviews**





Anaconda Installation

Anaconda 2019.10 for Windows Installer

Python 3.7 version

[Download](#)

[64-Bit Graphical Installer \(462 MB\)](#)

[32-Bit Graphical Installer \(410 MB\)](#)

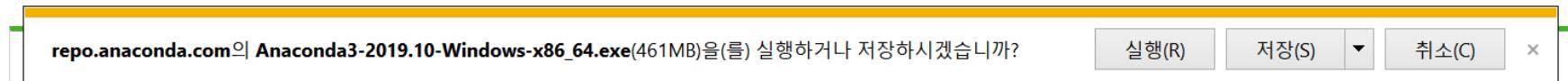
Python 2.7 version

[Download](#)

[64-Bit Graphical Installer \(413 MB\)](#)

[32-Bit Graphical Installer \(356 MB\)](#)

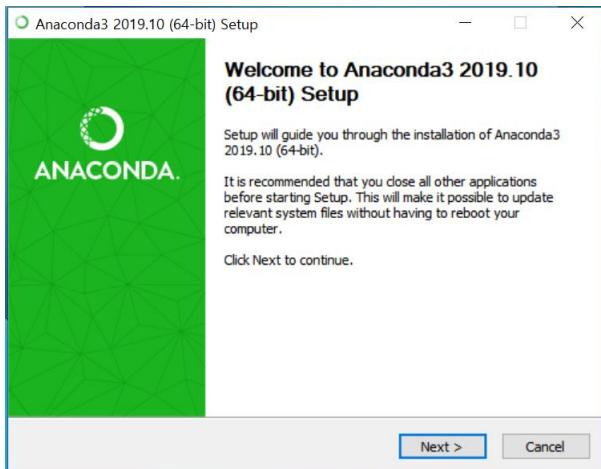
Get Started with Anaconda Distribution



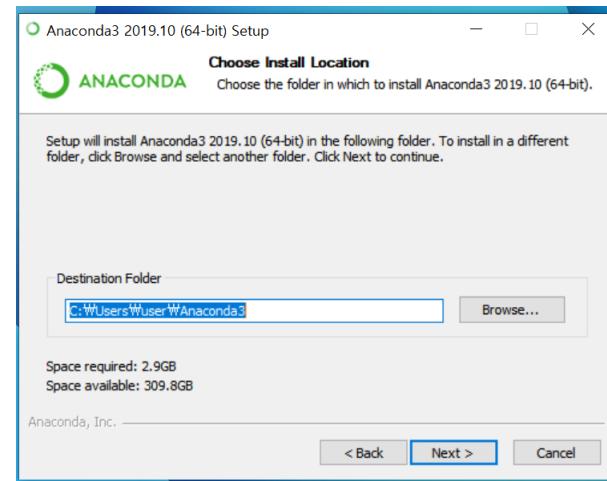


Anaconda를 설치(1)

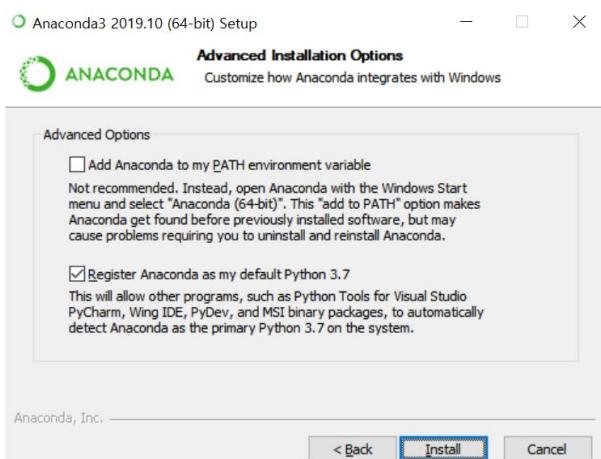
1



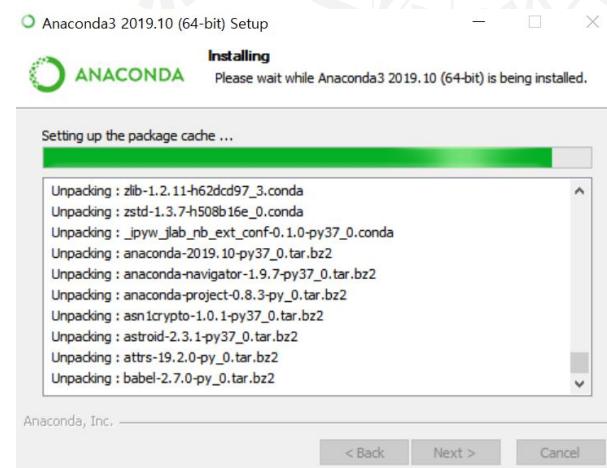
2



3



4





Anaconda Navigator (1)

- Home에서 base(root)
- Environments에서 새로운 가상환경을 생성
 - (각자 생성, 예) tf-cpu

The screenshot shows the Anaconda Navigator application window. On the left, there's a sidebar with 'Home', 'Environments' (which is highlighted with a red circle), 'Learning', and 'Community'. Below the sidebar are 'Documentation' and 'Developer Blog' buttons, along with social media links for Twitter, YouTube, and GitHub. At the bottom of the sidebar are 'Create', 'Clone', 'Import', and 'Remove' buttons. A central panel displays a 'Create new environment' dialog box, also circled in red. The dialog has fields for 'Name' (tf-cpu), 'Location' (set to 'base (root)'), and 'Packages' (Python 3.7 checked, R unchecked). Below the dialog is a list of available packages: asn1crypto and astroid, both of which have checkboxes next to them. To the right of the dialog, the main Navigator interface shows the 'Home' tab selected, with sections for 'Environments', 'Learning', and 'Community'. It also shows a 'Channels' section and a list of applications: 'lab' (JupyterLab) and 'jupyter' (Notebook). A large watermark of various icons like gears, charts, and arrows is overlaid on the right side of the screen.



Anaconda Navigator (2)

- Environments (tf-cpu) 생성 확인과 Update index
 - Not installed에서 tensorflow 검색
 - 버전과 CPU, GPU 확인하고 설치, Keras로 설치

File Help

ANACONDA® NAVIGATOR

Sign in to Anaconda Cloud

Home Environments Learning Community Documentation Developer Blog

Search Environments

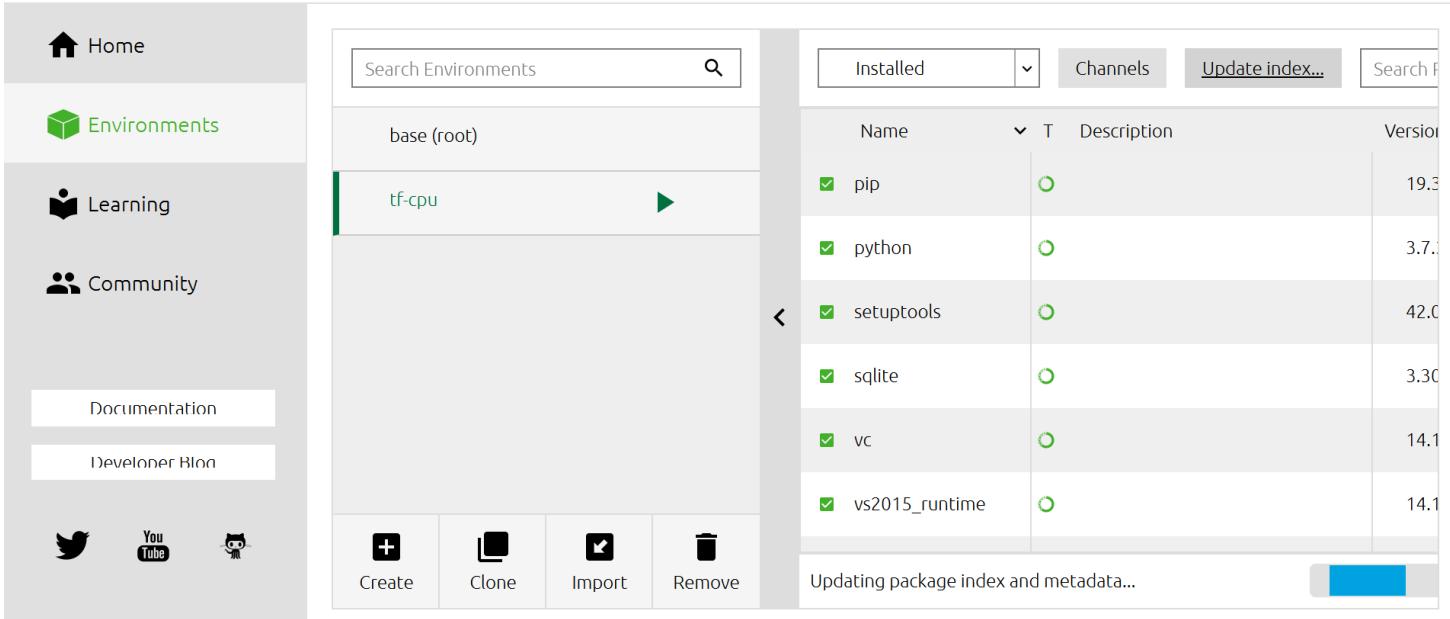
base (root)
tf-cpu

Installed Channels Update index...

Name	T	Description	Version
pip	<input checked="" type="checkbox"/>		19.3
python	<input checked="" type="checkbox"/>		3.7.
setupools	<input checked="" type="checkbox"/>		42.0
sqlite	<input checked="" type="checkbox"/>		3.30
vc	<input checked="" type="checkbox"/>		14.1
vs2015_runtime	<input checked="" type="checkbox"/>		14.1

Updating package index and metadata...

Create Clone Import Remove





Tensorflow version 확인 (2.0.0)

CPU 버전 선택하면, Keras도 CPU (GPU가 있으면 GPU 설치 권장)

➤ Keras 2.3.1, Tensorflow (CPU) 2.0.0

Search Environments Not installed Channels Update index... tensorflow X

base (root)

tf-cpu

Name Description Version

<input checked="" type="checkbox"/> keras	Deep learning library for theano and tensorflow	2.3.1
<input type="checkbox"/> keras-gpu	Deep learning library for theano and tensorflow	2.2.4
<input type="checkbox"/> opt_einsum	Optimizing einsum functions in numpy, tensorflow, dask, and more with contraction order optimization.	3.1.0
<input type="checkbox"/> r-tensorflow		2.0.0
<input type="checkbox"/> sagemaker-tensorflow-container		2.0.7
<input checked="" type="checkbox"/> tensorflow	Tensorflow is a machine learning library.	2.0.0

17 packages available matching "tensor" 2 packages selected Apply Clear

Create Clone Import Remove



중간 확인

- 가상환경(tf-cpu)로 들어가기
- (설치된 것 확인)
 - Tensorboard
 - Pip
 - Numpy, scipy
 - Theano
- 추가 SW 설치
 - Matplotlib,pandas, seaborn, xlrd
- (방법1) navigator 이용
- (방법2) pip install 이용

Anaconda Prompt (Anaconda3)

```
(base) C:\Users\user>activate tf-cpu
```

Package	Version
absl-py	0.8.1
astor	0.7.1
certifi	2019.11.28
gast	0.2.2
google-pasta	0.1.8
grpcio	1.23.0
h5py	2.10.0
Keras	2.3.1
Keras-Applications	1.0.8
Keras-Preprocessing	1.1.0
Mako	1.1.0
Markdown	3.1.1
MarkupSafe	1.1.1
mkl-service	2.3.0
numpy	1.17.3
opt-einsum	0+untagged.49.gdbede45.dirty
pip	19.3.1
protobuf	3.11.0
pygpu	0.7.6
pyreadline	2.1
PyYAML	5.1.2
scipy	1.3.1
setuptools	42.0.1.post20191125
six	1.13.0
tensorboard	2.0.0
tensorflow	2.0.0
tensorflow-estimator	2.0.0
termcolor	1.1.0
Theano	1.0.4
Werkzeug	0.16.0
wheel	0.33.6
wincertstore	0.2
wrapt	1.11.2



Navigator Not installed Search

Install Packages X

20 packages will be installed

	Name	Unlink	Link	Channel
1	matplotlib	-	3.1.2	conda-forge
2	pandas	-	0.25.3	conda-forge
3	seaborn	-	0.9.0	conda-forge
4	xlrd	-	1.2.0	conda-forge
5	*cycler	-	0.10.0	conda-forge
6	*freetype	-	2.10.0	conda-forge
7	*icu	-	64.2	conda-forge
8	*jpeg	-	9c	conda-forge
9	*kiwisolver	-	1.1.0	conda-forge
...

* indicates the package is a dependency of a selected packages

[Cancel](#) [Apply](#)

Sigin to Anaconda Cloud

Not installed ▾ Channels Update index... Search Pac...

Name	Description	Version
_ipyw_jlab_nb_ext...	A configuration metapackage for enabling anaconda-bundled jupyter extensions	0.1.0
_libarchive_static_f...		3.3.3
_libgcc_mutex		0.1
_low_priority		1.0
_mutex_mxnet	Mutex package to pin a variant of mxnet conda package	0.0.40
_py-xgboost-mutex		2.0
_pytorch_select		1.2.0



Jupyter notebook

Anaconda Navigator

File Help

ANA CONDA NAVIGATOR

Sign in to Anaconda Cloud

Applications on tf-cpu

mercurial-app 5.2

jupyter Notebook

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Install Install

Installing application notebook on C:\Users\user\Anaconda3\envs\tf-c...

Home Environments Learning Community Documentation Developer Hint

Twitter YouTube GitHub

Home Page - Select or create a

localhost:8888/tree

https://m.blog.nav...

jupyter

Files Running Clusters

Select items to perform actions on them.

Name	Last Modified	File size
0		
3D Objects	16일 전	
Anaconda3	한 시간 전	
Contacts	16일 전	
Desktop	31분 전	
Documents	3일 전	
Downloads	3일 전	
Favorites	16일 전	
Links	16일 전	
Music	16일 전	

Upload New

Logout





Tensorflow 첫 예제

jupyter

Files Running Clusters

Select items to perform actions on them.

Upload New ▾ Logout

Name: Notebook: Python 3
Other:
Text File
Folder
Terminal

0 / 3D Objects Anaconda3 Contacts Desktop Documents Downloads Favorites

3일 전 3일 전 16일 전

jupyter Untitled (unsaved changes)

File Edit View Insert Cell Kernel Help Trusted Python 3 Logout

In [1]: `import tensorflow as tf`

In [2]: `tf.__version__`

Out[2]: '2.0.0'

In []:



(방법2) pip 이용 tensorflow 설치

- 가상환경 생성 : > conda - n tf-cpu python=3.7
- 가상환경 로그인 > activate tf-cpu
 - 변경내용 확인 : (tf-cpu)>
 - 설치 SW 확인 : (tf-cpu)> pip list
 - SW 설치 : > pip install matplotlib
- 가상환경 로그아웃 > conda deactivate

```
C:\Users\vaio>conda create -n tensorflow python=3.5
Fetching package metadata .....
Solving package specifications: .....

Package plan for installation in environment C:\Users\vaio\Anaconda3\envs\tensorflow:

The following packages will be downloaded:

  package          |      build
  -----|-----
  pip-9.0.1         |    py35_1      1.7 MB

The following NEW packages will be INSTALLED:

  pip:            9.0.1-py35_1
  python:         3.5.2-0
  setuptools:     27.2.0-py35_1
  vs2015_runtime: 14.0.25123-0
  wheel:          0.29.0-py35_0

Proceed ([y]/n)?
```



Tensorflow 첫 테스트 및 버전확인

```
(tf-cpu) C:\Users\user>
(tf-cpu) C:\Users\user>
(tf-cpu) C:\Users\user>python
Python 3.7.3 | packaged by conda-forge | (default, Jul 1 2019, 22:01:29)
n win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import tensorflow as tf
>>> tf.__version__
'2.0.0'
>>>
>>> exit()

(tf-cpu) C:\Users\user>
(tf-cpu) C:\Users\user>
```

Thank You!

www.ust.ac.kr

