

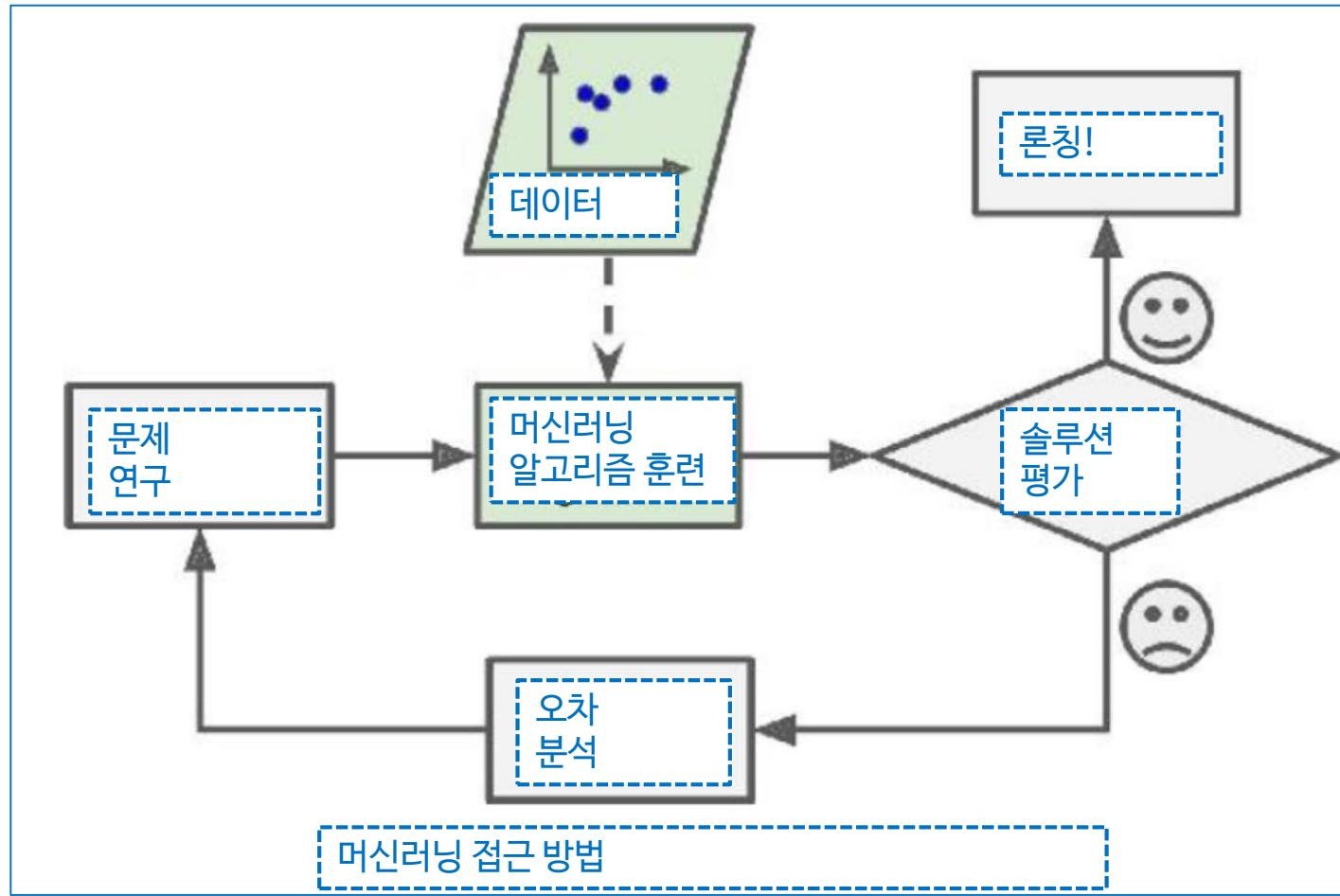
# 1강: 기계학습 생태계의 이해

인공지능 일반강좌 : 기계학습의 이해(L2-1)

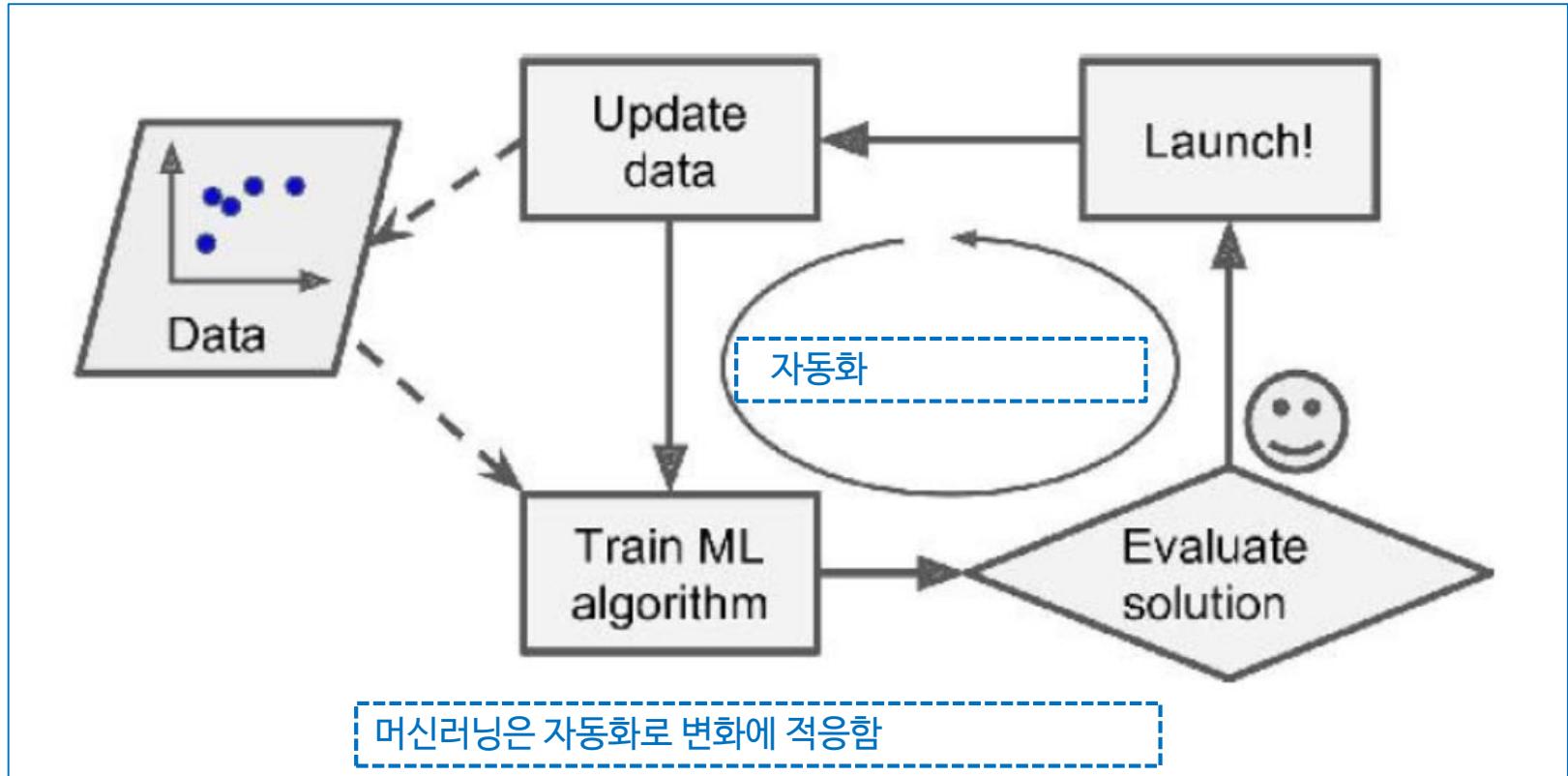
# 머신러닝이란?

- 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

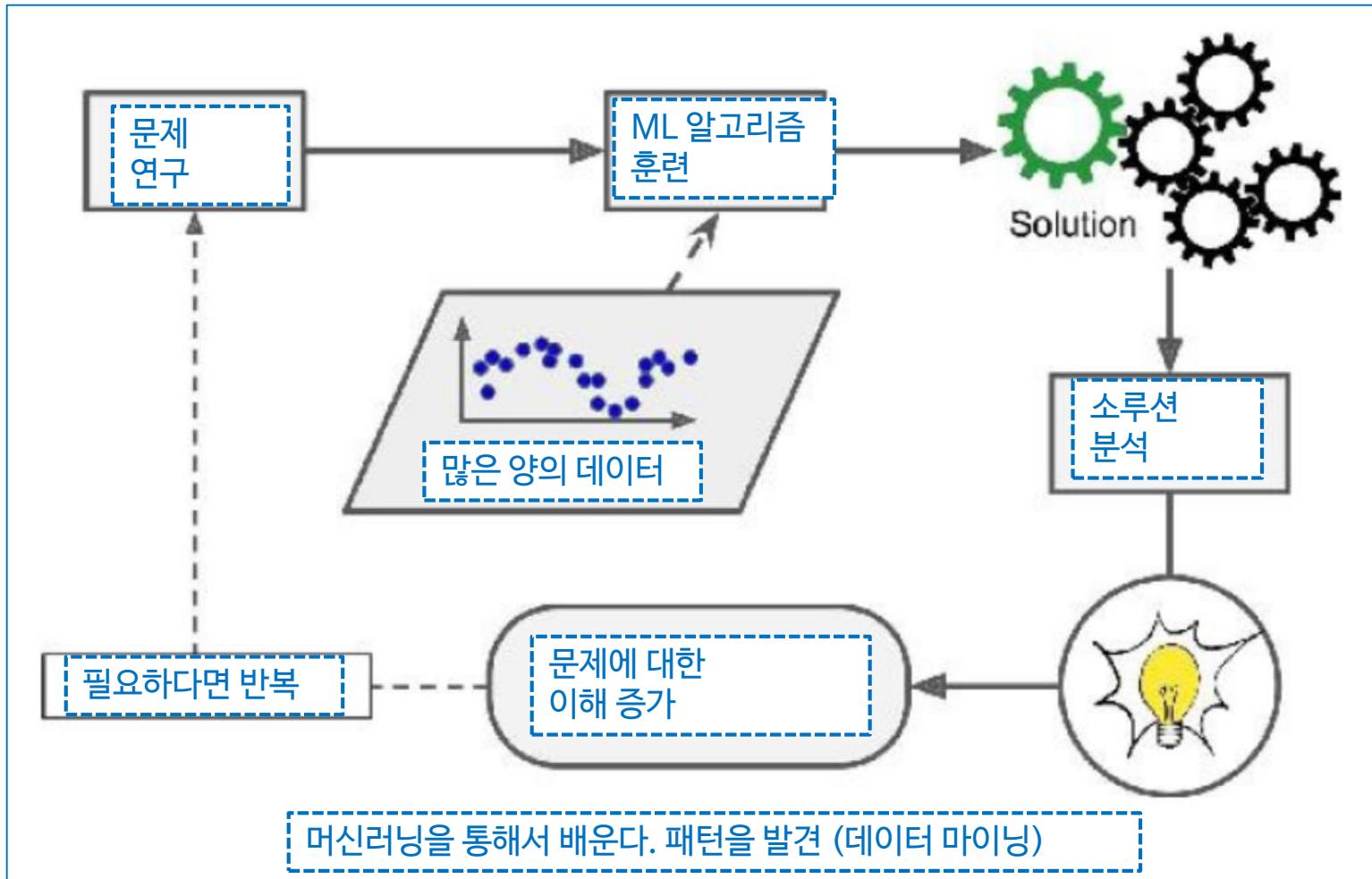
# 왜 머신러닝을 사용하는가? (1)



## 왜 머신러닝을 사용하는가? (2)



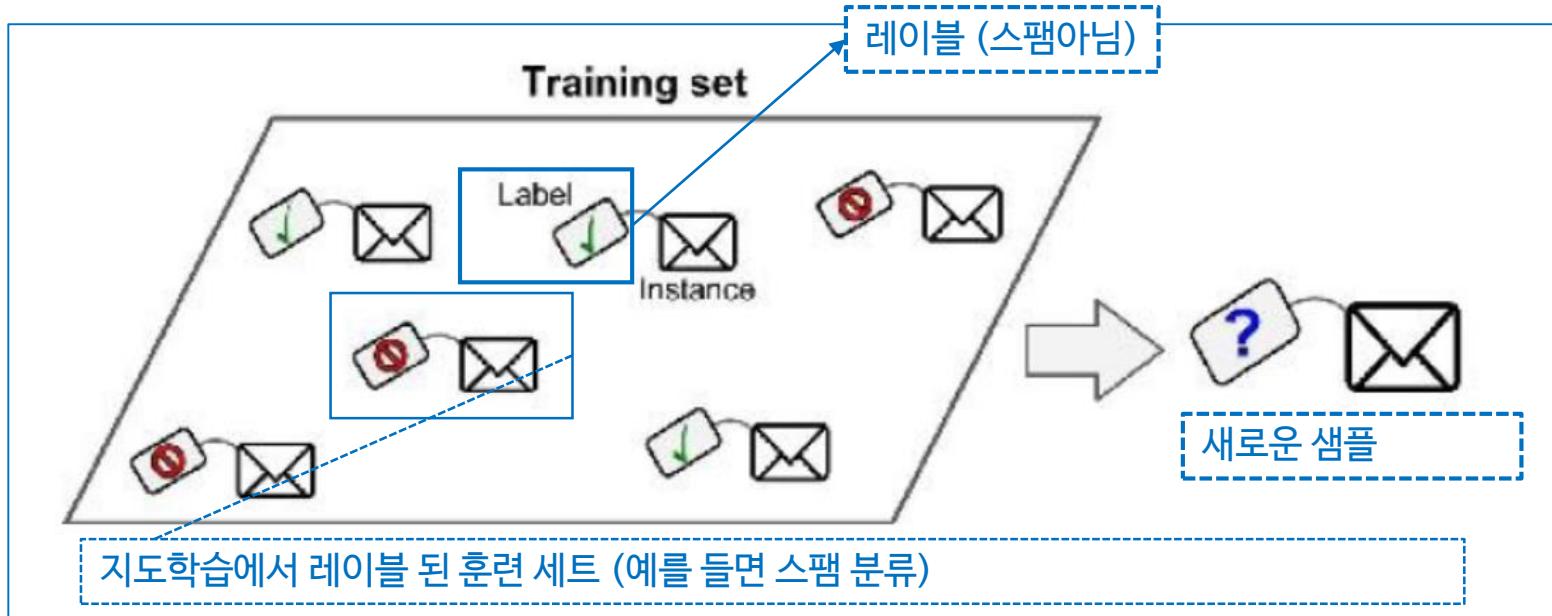
# 왜 머신러닝을 사용하는가? (3)



# 머신러닝 시스템의 종류(1)

- **지도학습**

- ✓ 알고리즘에 주입하는 훈련 데이터에 레이블이라는 원하는 답이 포함됨

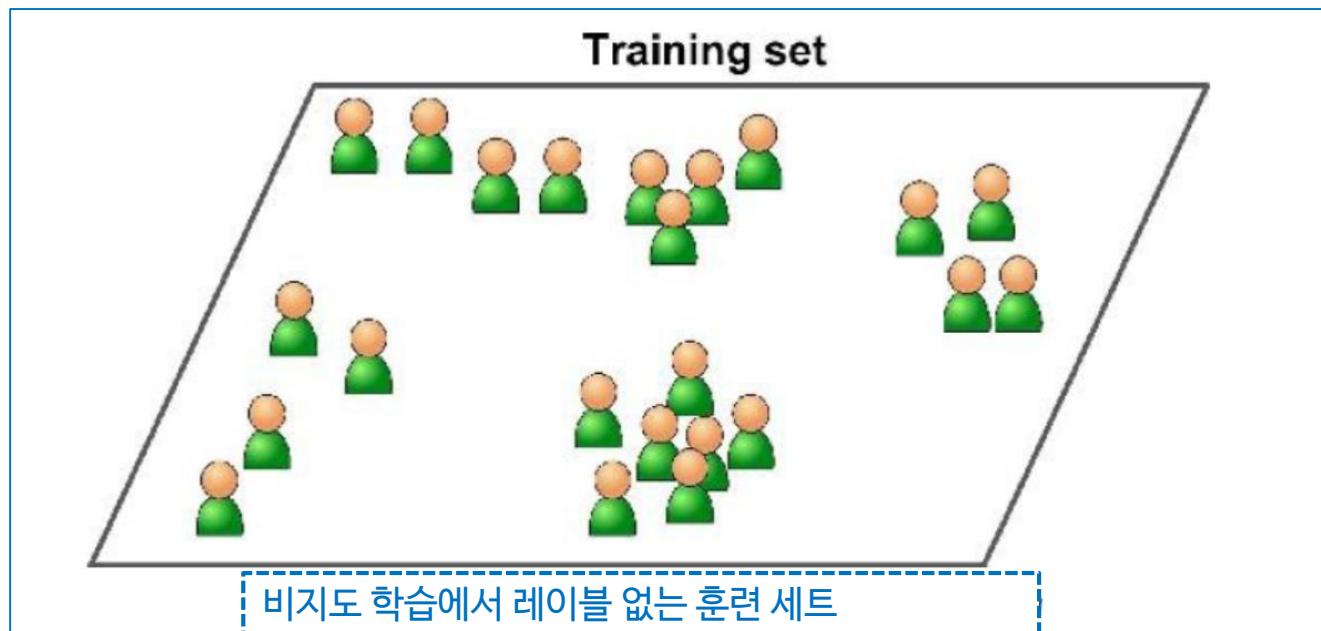


# 지도학습 알고리즘

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

# 비지도학습

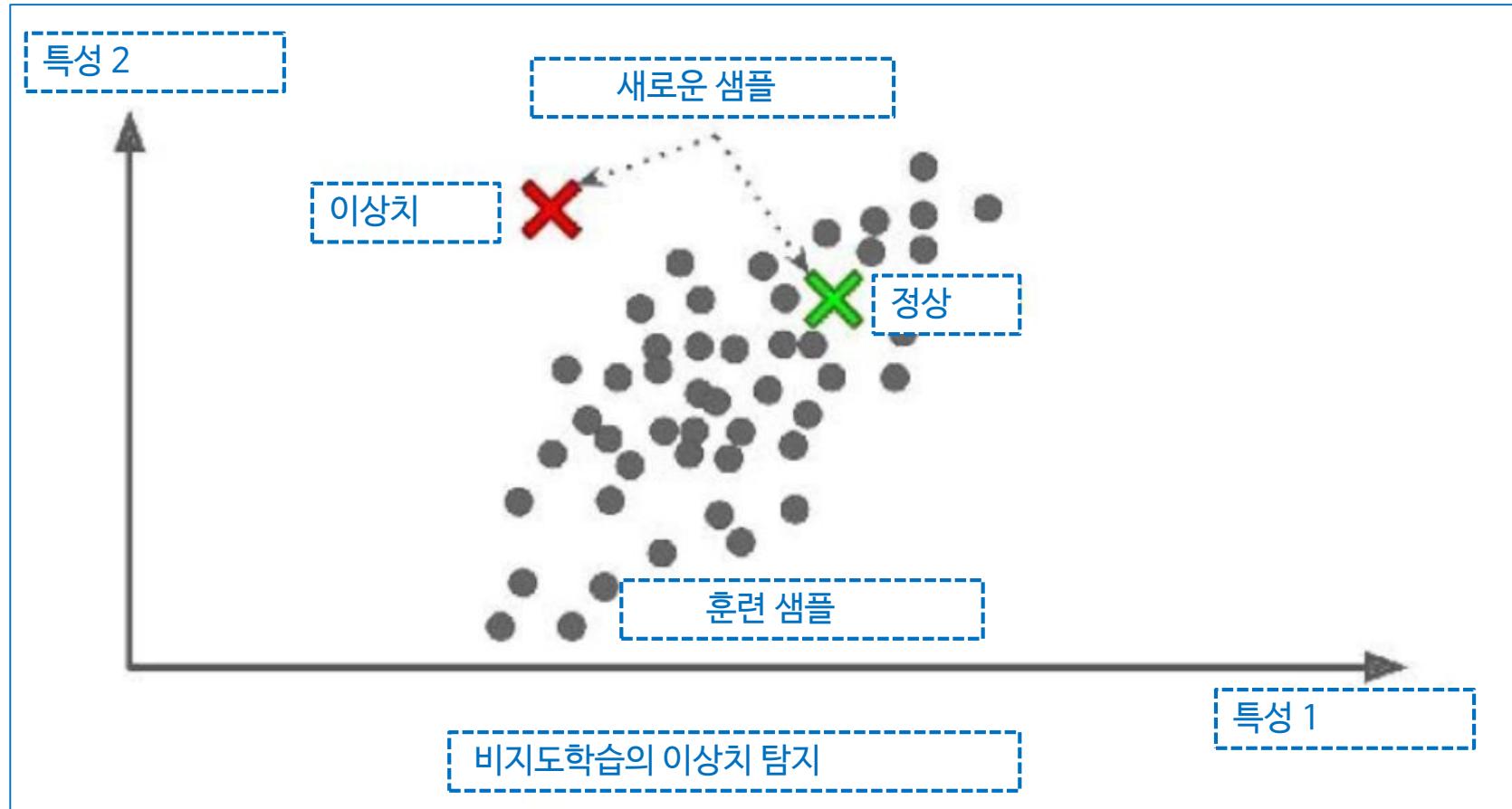
- 비지도학습(Unsupervised Learning)
  - ✓ 훈련데이터에 레이블이 없다.
  - ✓ 시스템이 아무런 도움없이 학습을 해야 한다.
  - ✓ (예) 중간 시험은 봤는데 채점은 하지않고 시험만 계속 봄



# 비지도 학습의 알고리즘

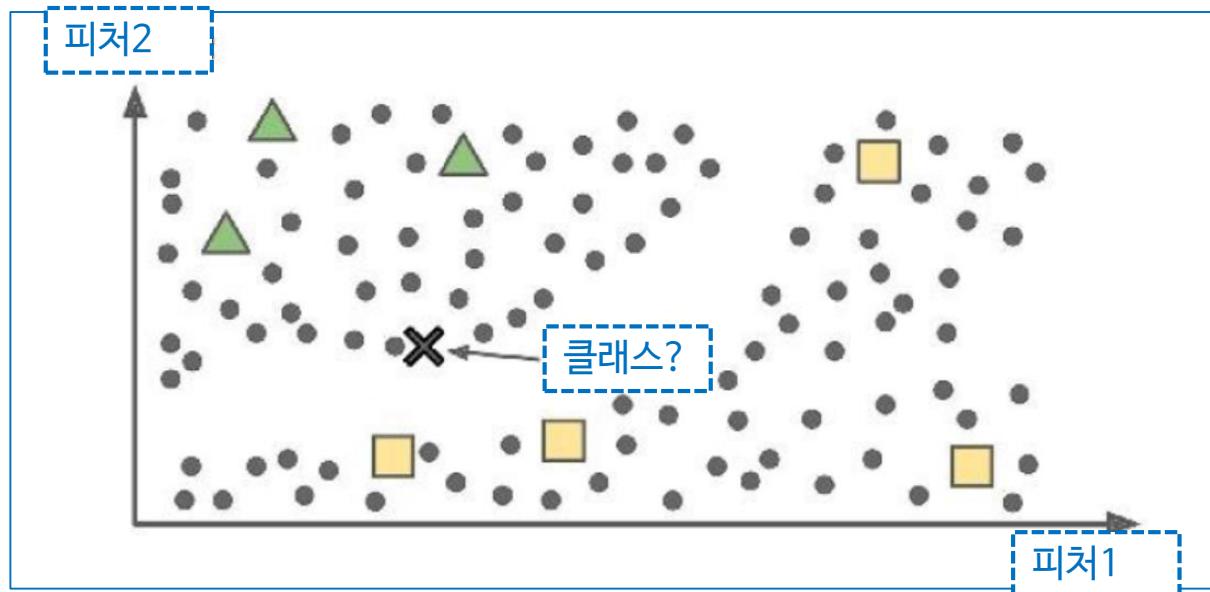
- 군집(Clustering)으로 K-평균
  - ✓ K-Means
- 차원축소로 주성분분석(PCA)
  - ✓ PCA (Principal Component Analysis)
- 이상치 탐지(Anomaly detection)
  - ✓ 부정 거래를 막기 위해 이상한 신용카드 거래를 탐지
  - ✓ 제조 결함을 잡아 내고, 학습 이전에 데이터 셋에서 이상한 값을 자동으로 제거하는 것
  - ✓ 도로에서 주행하는 차량이 차선 위반, 속도위반을 자동으로 검출하는 기술

# 비지도학습의 이상치 탐지

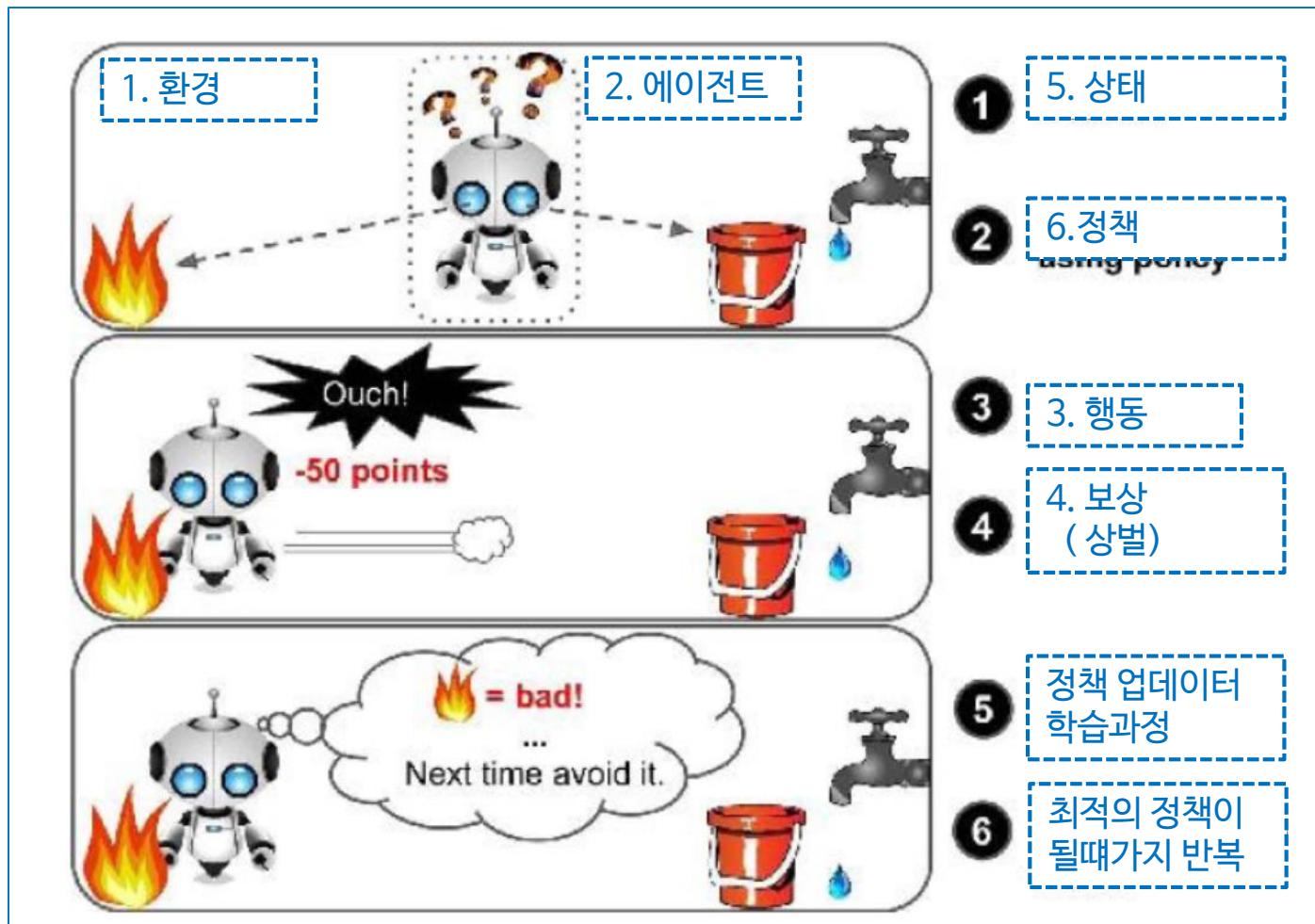


# 준지도학습

- 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

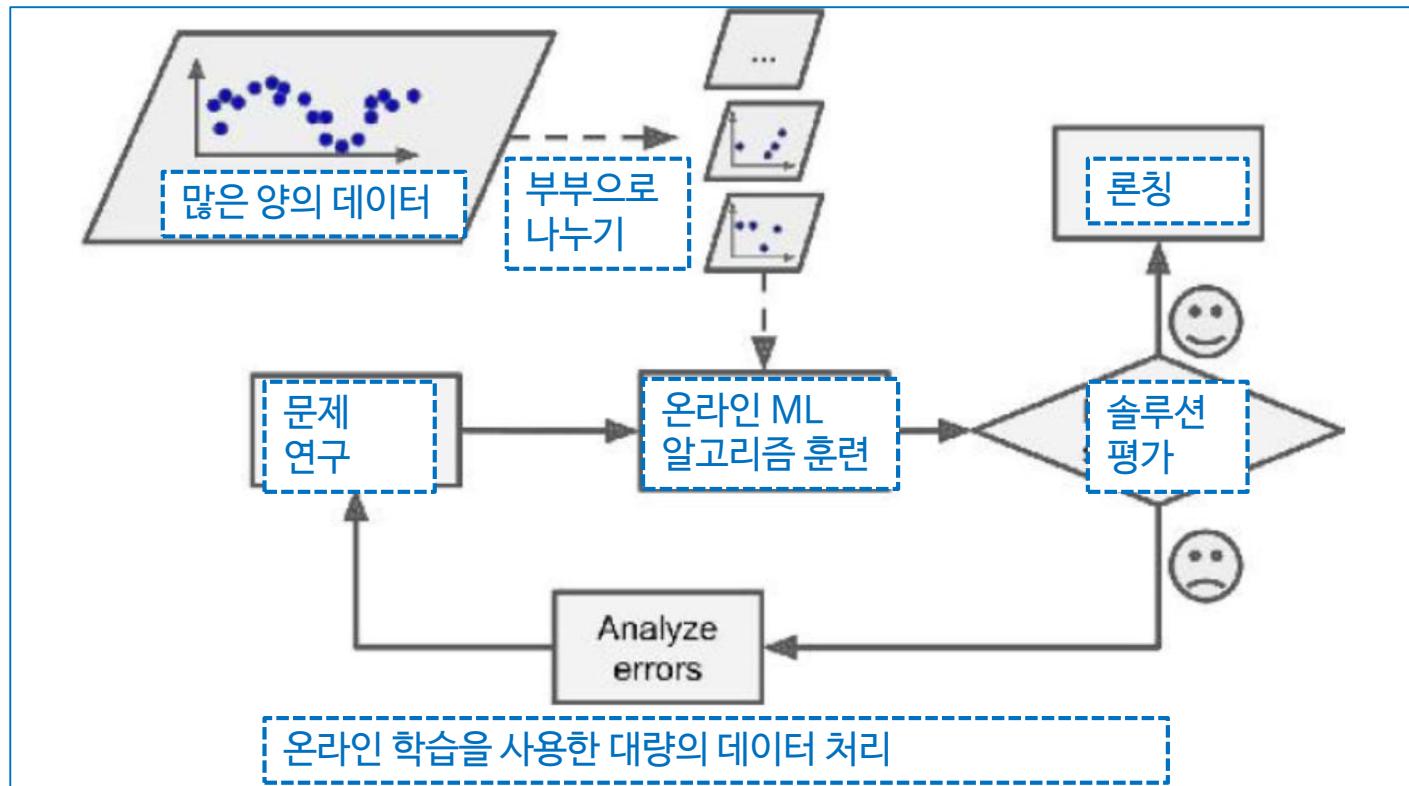


# 머신러닝에서 데이터 학습 방법 (1)

- 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



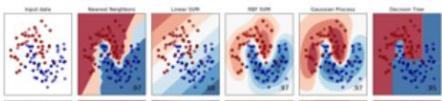
# 머신러닝에 사용하는 패키지

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

# Scikit-learn 사이킷런

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

The screenshot shows the official scikit-learn website at <https://scikit-learn.org/stable/>. The header includes the logo, navigation links for Install, User Guide, API, Examples, and More, and a search bar with a 'Go' button. The main content area features the title "scikit-learn" and subtitle "Machine Learning in Python". Below this are three main 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...  

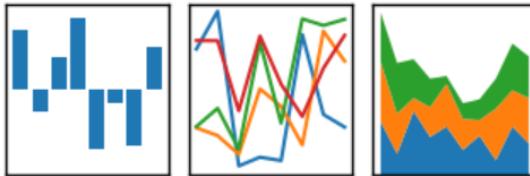

**Clustering**  
Automatic grouping of similar objects into sets.  
**Applications:** Customer segmentation, Grouping experiment outcomes  
**Algorithms:** k-Means, spectral clustering, mean-shift, and more...  


# 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

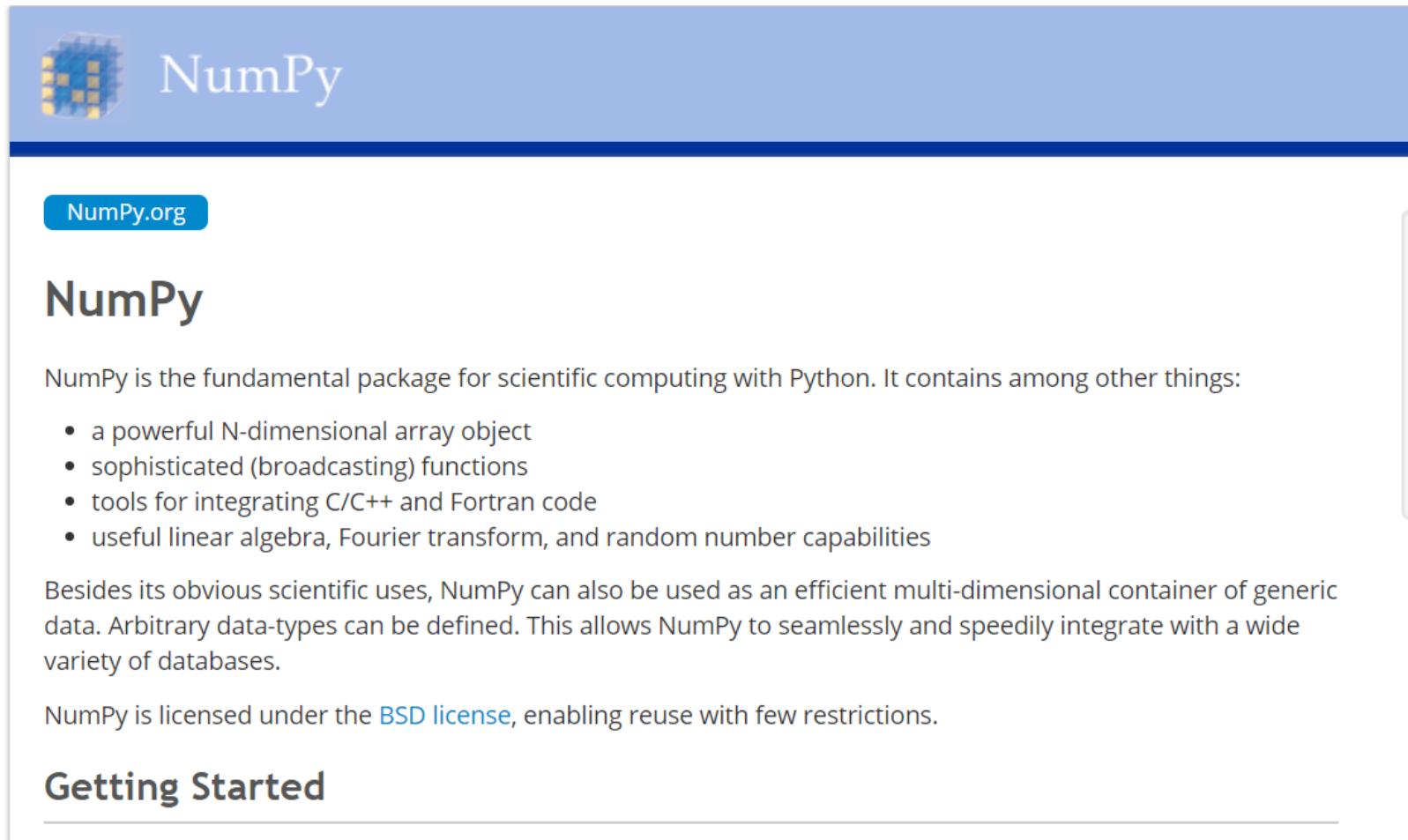
1.0.0 - September 2019

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

#### Previous Releases

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

# Numpy

A screenshot of the NumPy.org homepage. The header features the NumPy logo (a blue and yellow grid icon) and the word "NumPy". Below the header is a navigation bar with a "NumPy.org" button. The main content area has a large title "NumPy". A paragraph describes NumPy as the fundamental package for scientific computing with Python, mentioning its array object, broadcasting functions, tools for integrating C/C++ and Fortran code, and useful linear algebra, Fourier transform, and random number capabilities. Another paragraph discusses its use as an efficient container for generic data and its integration with databases. A note about the BSD license is present, followed by a "Getting Started" section.

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/>

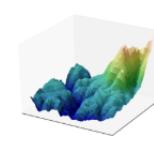
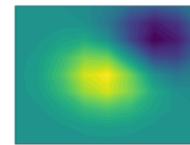
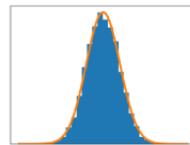
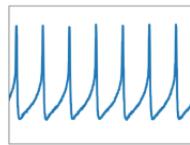


Version 3.1.2

[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.



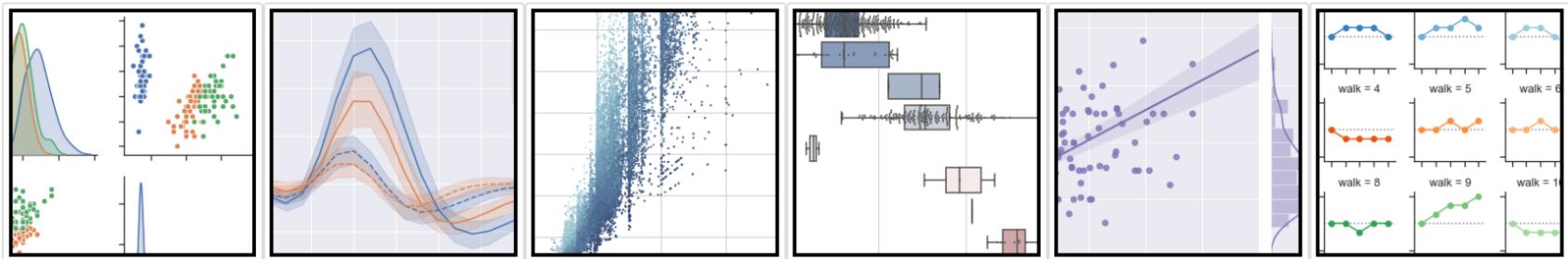
Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code. For examples, see the [sample plots](#) and [thumbnail gallery](#).

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](#)

# 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. Please 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:		Most Popular Data Sets (hits since 2007):	
10-06-2019:	 <a href="#">WISDM Smartphone and Smartwatch Activity and Biometrics Dataset</a>	3091299:	 <a href="#">Iris</a>
09-30-2019:	 <a href="#">Hepatitis C Virus (HCV) for Egyptian patients</a>	1707888:	 <a href="#">Adult</a>
09-23-2019:	 <a href="#">QSAR fish toxicity</a>	1324210:	 <a href="#">Wine</a>
09-23-2019:	 <a href="#">QSAR aquatic toxicity</a>	1123015:	 <a href="#">Wine Quality</a>
09-21-2019:	 <a href="#">Online Retail II</a>	1122923:	 <a href="#">Heart Disease</a>
09-20-2019:	 <a href="#">Human Activity Recognition from Continuous Ambient Sensor Data</a>	1114183:	 <a href="#">Car Evaluation</a>
09-20-2019:	 <a href="#">Beijing Multi-Site Air-Quality Data</a>	1107500:	 <a href="#">Breast Cancer Wisconsin (Diagnostic)</a>
09-20-2019:	 <a href="#">MEx</a>	1097941:	 <a href="#">Bank Marketing</a>
07-30-2019:	 <a href="#">PPG-DaLiA</a>	933570:	 <a href="#">Human Activity Recognition Using Smartphones</a>

# 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)

# 1. 타이타닉 생존 데이터

The screenshot shows the Kaggle website for the "Titanic: Machine Learning from Disaster" competition. At the top left, there's a circular icon with three horizontal bars and the text "Getting Started Prediction Competition". The main title "Titanic: Machine Learning from Disaster" is displayed prominently in large, bold, white font. Below it, a subtitle reads "Start here! Predict survival on the Titanic and get familiar with ML basics". On the left side, there's a small blue square icon with a white letter "k" and the text "Kaggle · 15,671 teams · Ongoing". Below the main title, a navigation bar contains links: Overview (underlined in blue), Data, Notebooks, Discussion, Leaderboard, Rules, Team, and My Submissions. The "Overview" section is currently active. In the "Description" box, there's a yellow hand icon and a blue captain's hat icon followed by the text "Ahoy, welcome to Kaggle! You're in the right place.". The "Evaluation" box contains the text "This is the legendary Titanic ML competition – the best, first challenge for you to di".

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

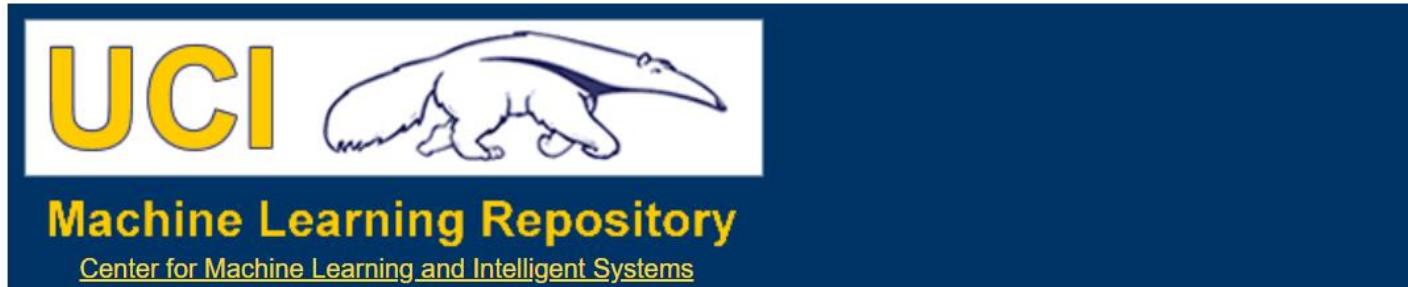
Ahoy, welcome to Kaggle! You're in the right place.

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>



### 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



#### 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

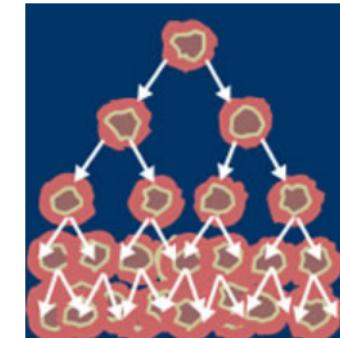
[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))



### 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](https://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

The screenshot shows a web browser displaying the Kaggle website at [kaggle.com/mlg-ulb/creditcardfraud](https://kaggle.com/mlg-ulb/creditcardfraud). The page title is "Credit Card Fraud Detection". The main image is a blurred photograph of several credit cards. On the right side of the image, there is a yellow circular icon with a downward-pointing arrow and the number "4580". Below the image, the text "Anonymized credit card transactions labeled as fraudulent or genuine" is displayed. The dataset was created by "Machine Learning Group - ULB" and was updated 2 years ago (Version 3). The page features a navigation bar with links for Data, 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).

kaggle.com/mlg-ulb/creditcardfraud

kaggle Search Competitions Datasets Notebooks Discussion Courses ...

Dataset

Credit Card Fraud Detection

Anonymized credit card transactions labeled as fraudulent or genuine

Machine Learning Group - ULB • updated 2 years ago (Version 3)

Data Tasks (9) Kernels (2,334) Discussion (46) Activity Metadata Download (144 MB) New Notebook

Usability 8.5 License Database: Open Database, Contents: Database Contents Tags computing, finance, crime, machine learning, credit cards

# 8. 보스턴 주택 가격 데이터

The screenshot shows the UCI ML Datasets page for the Boston Housing Dataset. The page features a background image of a dense urban residential area with multi-story houses. At the top left, there's a 'Dataset' icon. The title 'UCI ML Datasets' and the specific dataset name 'Boston Housing Dataset' are prominently displayed. Below the title, the author's name 'Khashayar Baghizadeh Hosseini' and the update information 'updated 3 years ago (Version 1)' are shown. A navigation bar at the bottom includes links for 'Data' (which is underlined), 'Tasks', 'Kernels (8)', 'Discussion (1)', 'Activity', 'Metadata', 'Download (35 KB)', 'New Notebook', and a three-dot menu. Below the navigation bar, there are sections for 'Usability' (rating 8.8), 'License' (Database: Open Database, Contents: Database Contents), and 'Tags' (computing, natural and physical sciences, education, society, mathematics). There are also 'Description' and 'Context' sections.

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 데이트

<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

# Installation

The screenshot shows the TensorFlow website's "Install" page. The top navigation bar includes links for "Install", "Learn", "API", "Resources", "Community", and "Why TensorFlow". A sidebar on the left contains links for "Install TensorFlow" (which is highlighted in blue), "Packages" (with sub-links for "pip" and "Docker"), "Additional setup" (with sub-links for "GPU support" and "Problems"), "Build from source" (with sub-links for "Linux / macOS", "Windows", and "Raspberry Pi"), and "Language bindings". The main content area features a dark banner at the top with the text "Missed TensorFlow World? Check out the recap." and a "Learn more" button. Below this, the title "Install TensorFlow 2" is displayed. A section titled "TensorFlow is tested and supported on the following 64-bit systems:" lists supported operating systems in two columns:

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

A large heading "Download a package" is followed by the instruction "Install TensorFlow with Python's *pip* package manager." To the right, a code block shows the command: `# Requires the latest pip  
pip install --upgrade pip`.

# Tensorflow : pip 이용한 설치

The screenshot shows the TensorFlow website's navigation bar at the top, with the 'Install' tab selected. Below the navigation, there's a sidebar on the left containing links for 'Install TensorFlow', 'Packages' (which includes 'pip'), 'Docker', 'Additional setup', 'GPU support', and 'Problems'. The main content area is titled 'Install TensorFlow with pip' and states that 'TensorFlow 2 packages are available'. It lists three package options: 'tensorflow' (CPU-only), 'tensorflow-gpu' (with GPU support for Ubuntu and Windows), and 'tf-nightly' (preview build). Below this, it says 'Older versions of TensorFlow' and notes that for version 1.15, CPU and GPU support are included in a single package. It also mentions that for versions 1.14 and older, separate CPU and GPU packages are available.

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

The screenshot shows the Windows System Information window. On the left, there's a sidebar with various options like Home, Settings search, System, Storage, etc. The 'Information' option is highlighted with a red box. The main area displays system details under 'Information' and 'System Configuration'.

**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

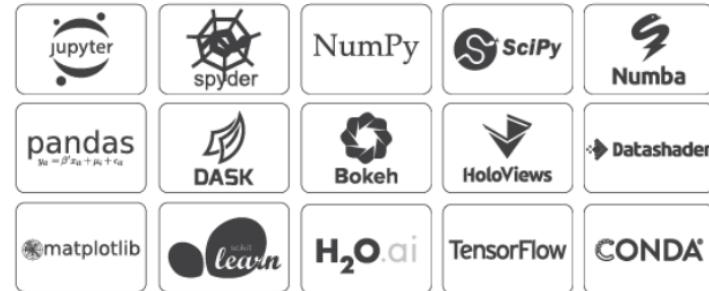
## Anaconda Distribution

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

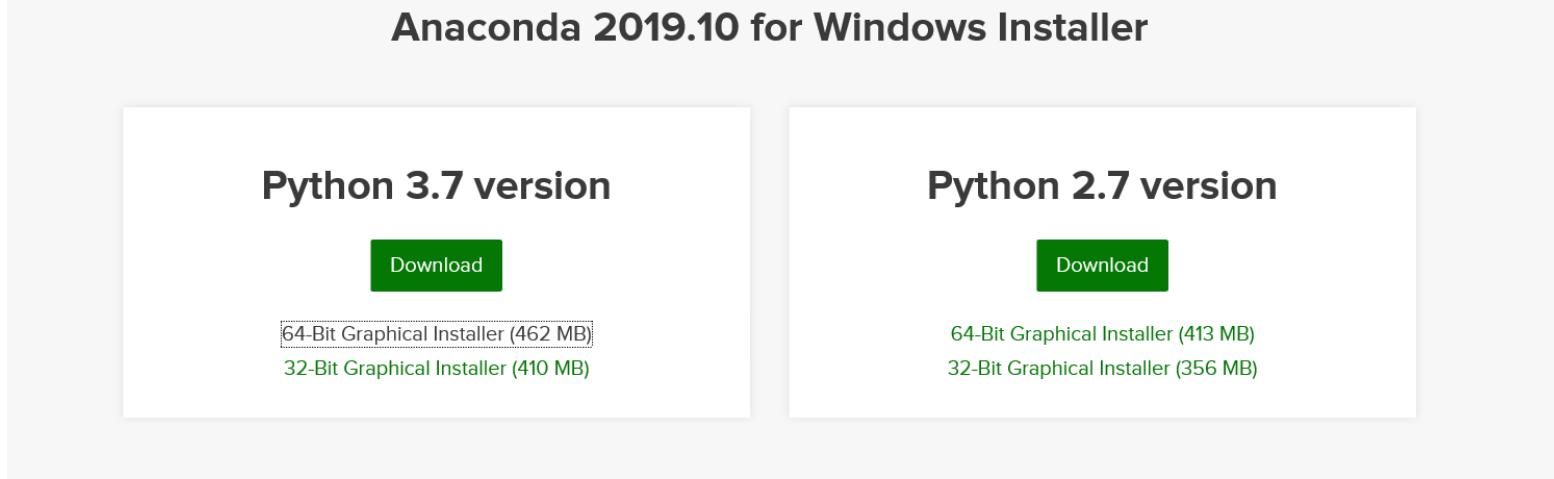
[Download](#)

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:

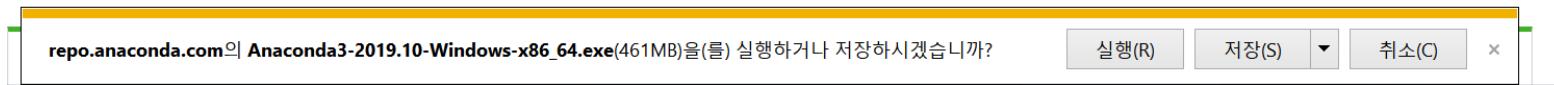
- Quickly download 1,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with [Conda](#)
- Develop and train machine learning and deep learning models with [scikit-learn](#), [TensorFlow](#), and [Theano](#)
- Analyze data with scalability and performance with [Dask](#),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



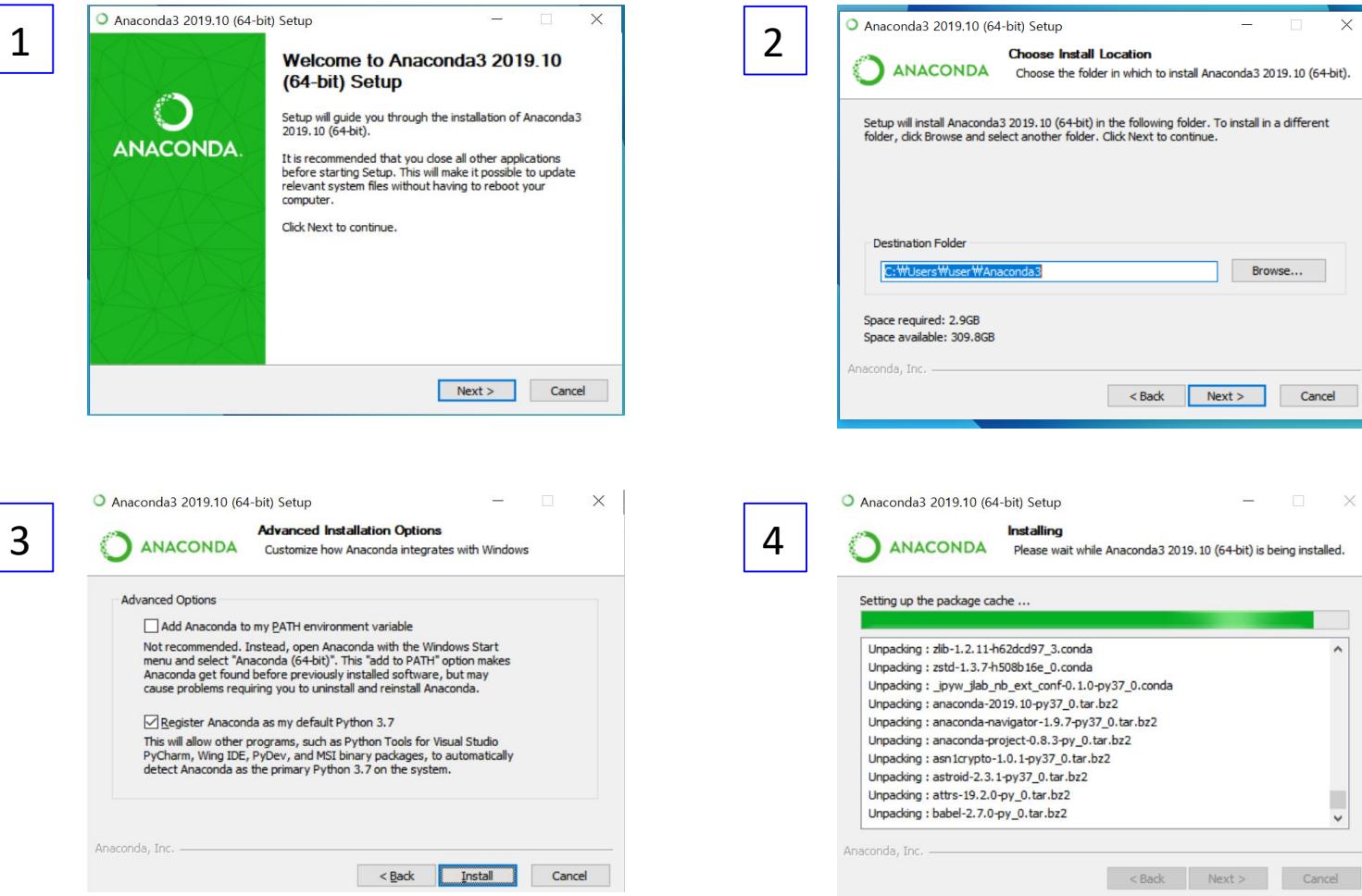
# Anaconda Installation



## Get Started with Anaconda Distribution

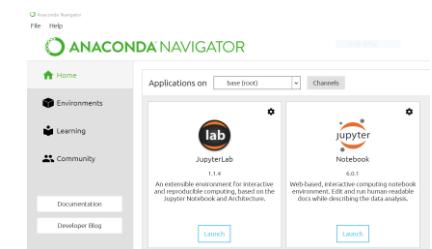
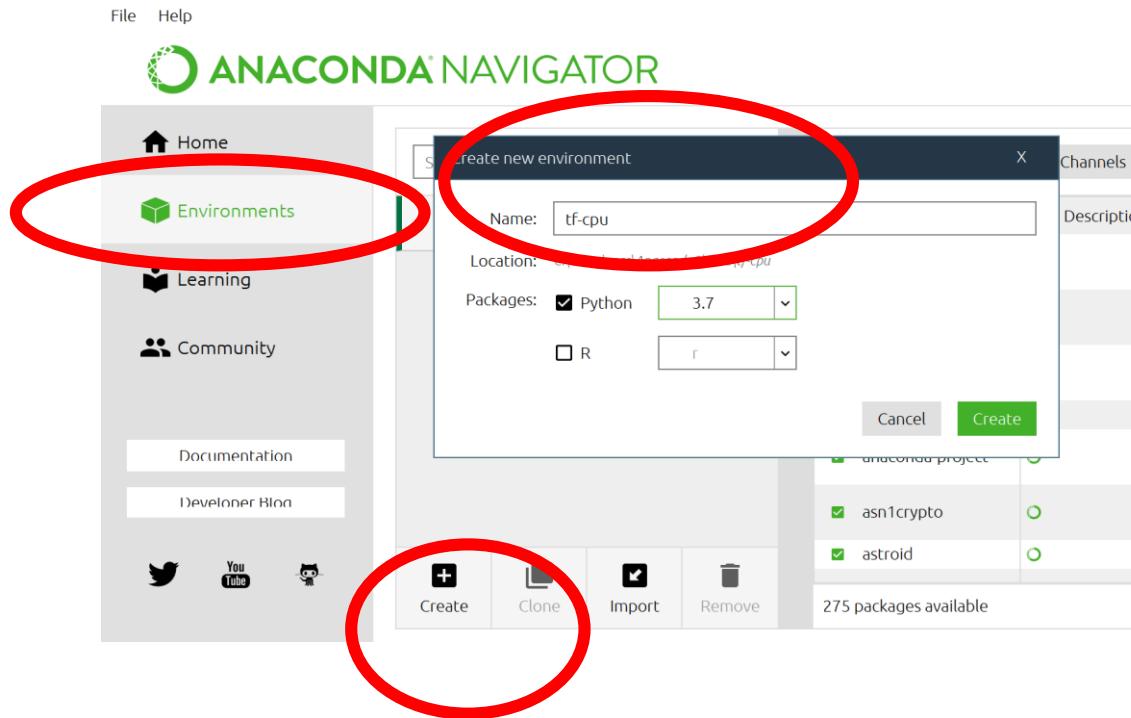


# Anaconda를 설치(1)



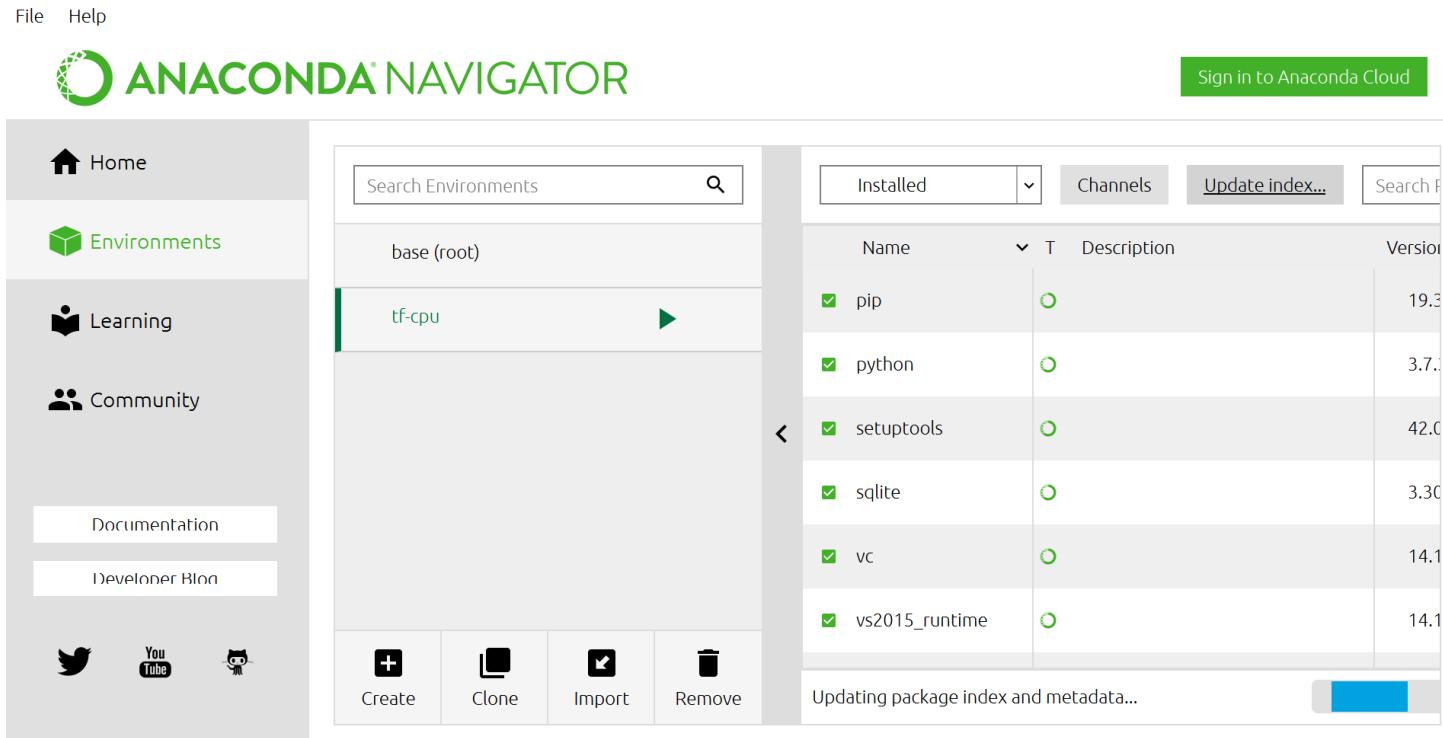
# Anaconda Navigator (1)

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



# Anaconda Navigator (2)

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



# 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

The screenshot shows the Jupyter Notebook environment manager interface. On the left, there's a sidebar with buttons for 'Create', 'Clone', 'Import', and 'Remove'. The main area displays a list of installed packages in the 'tf-cpu' environment. The search bar at the top right contains the text 'tensorflow'. The table has columns for Name, Description, and Version.

Name	Description	Version
keras	Deep learning library for theano and tensorflow	2.3.1
keras-gpu	Deep learning library for theano and tensorflow	2.2.4
opt_einsum	Optimizing einsum functions in numpy, tensorflow, dask, and more with contraction order optimization.	3.1.0
r-tensorflow		2.0.0
sagemaker-tensorflow-container		2.0.7
tensorflow	Tensorflow is a machine learning library.	2.0.0

At the bottom, a message says '17 packages available matching "tensor" 2 packages selected' with 'Apply' and 'Clear' buttons.

# 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-Aplications	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

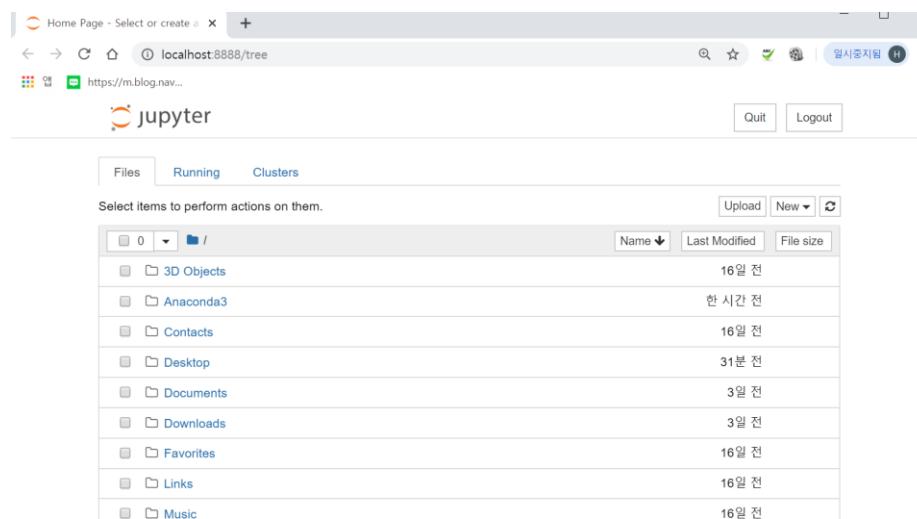
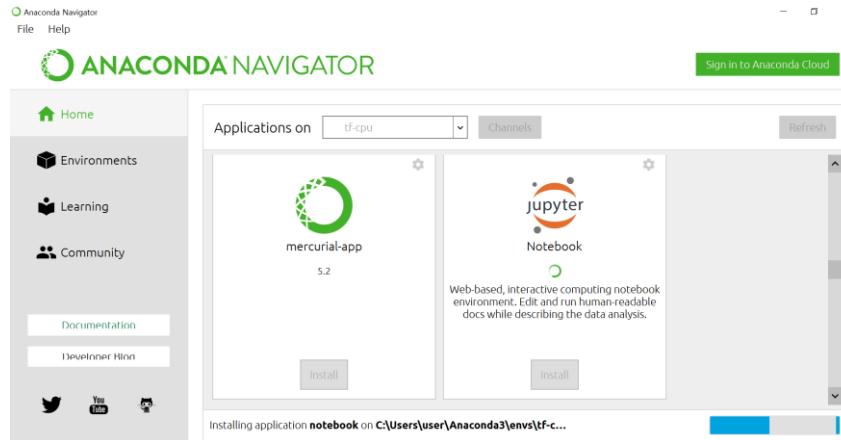
[Cancel](#) [Apply](#)

Sign in to Anaconda Cloud

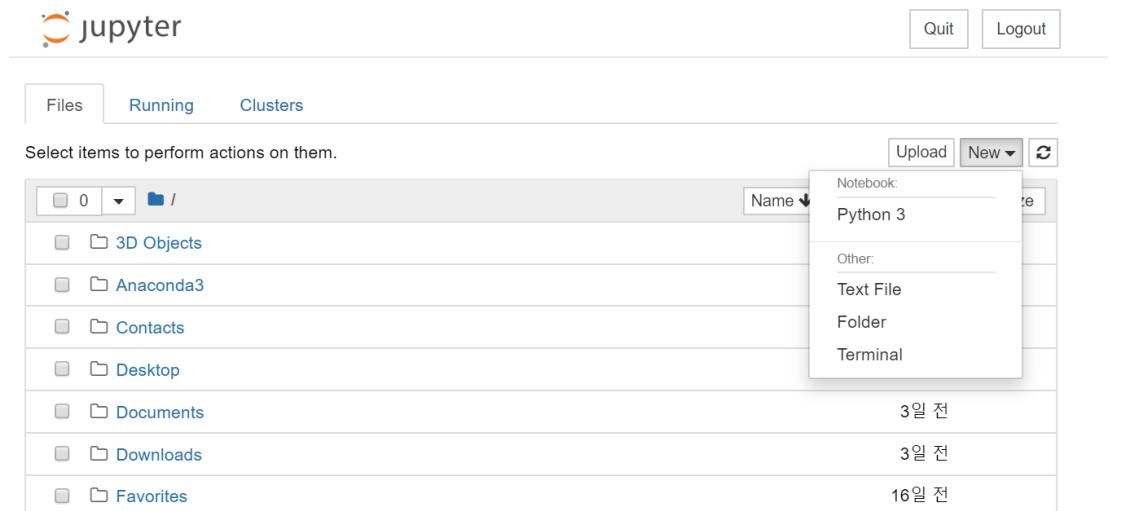
Not installed ▾ Channels Update index... Search Pac...

Name	T	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



# Tensorflow 첫 예제



The screenshot shows the Jupyter Notebook interface with the title 'jupyter Untitled (unsaved changes)'. The toolbar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Help', 'Trusted' status, and a 'Python 3' kernel selector. Below the toolbar is a toolbar with icons for file operations like 'New', 'Open', 'Save', etc., and a 'Run' button. The main area contains a code cell with the following content:

```
In [1]: import tensorflow as tf
In [2]: tf.__version__
Out[2]: '2.0.0'
```

A new code cell is currently being typed, indicated by the cursor in the first cell.

# (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](http://www.ust.ac.kr)