What are your experiences with AI or ML?

What are your expectations for this course?

Slides originally developed by Professor Boyeong Hong

From Data to Machine Learning, and Urban Planning


Urbanization

Urban Problems

Modern Urban Planning Exercises

Challenges


Big Data
&
Computational Capacity

We now have...

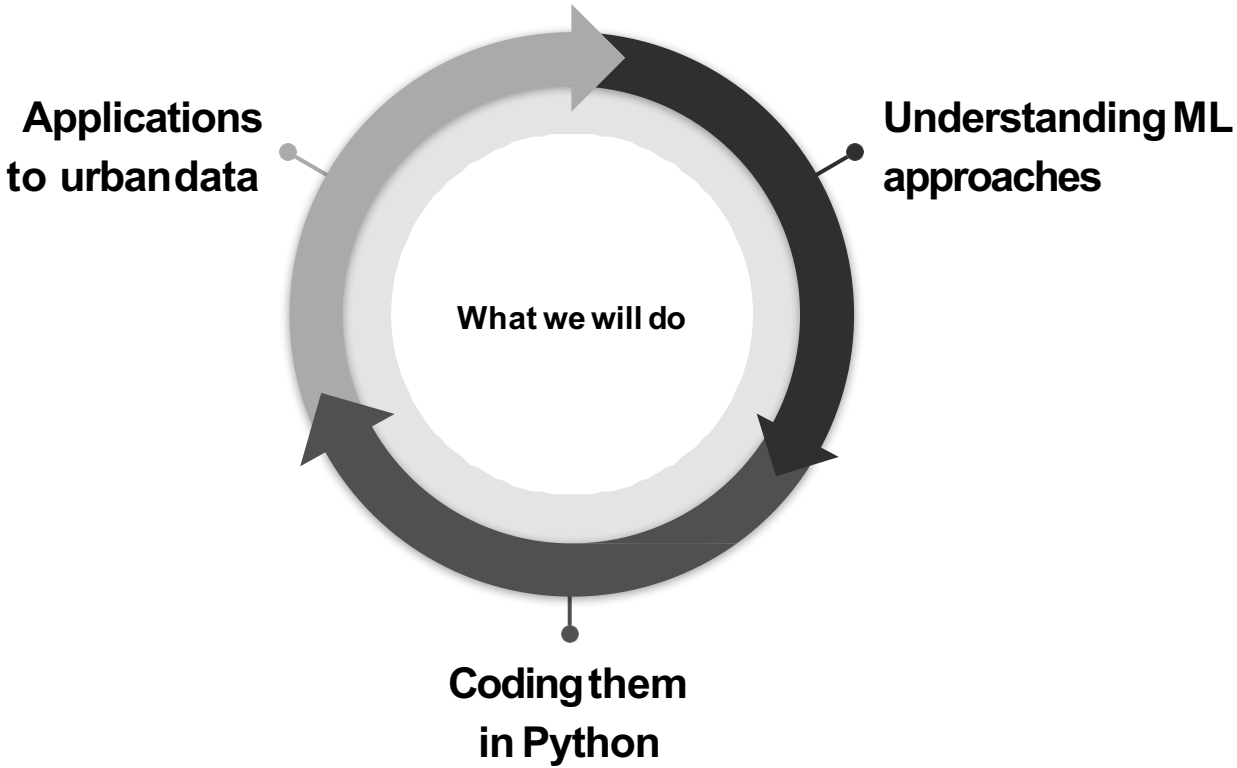
- Innovation, data-driven problem solving, and paradigm shift
- Unsolved persistent urban problems: sustainability, safety, and quality of life

What city governments do

- Apply problem-based data analytics or urban informatics (civic analytics)
- Improve the operation of city services and the policy-making process through a big data analytic approach
- In-depth understanding of urban systems
- Ability to predict and prepare for future scenarios in urban management and planning
- But still exploratory and ad hoc compared to industries

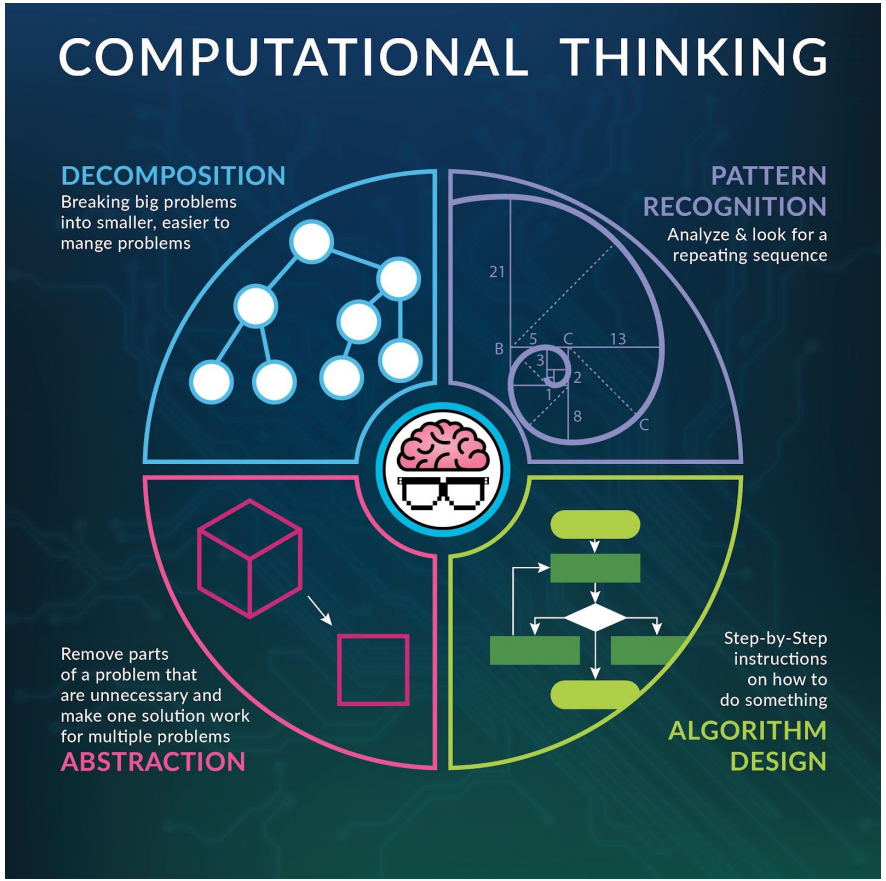
Big urban data and analytics including Machine Learning are NOT ***panacea*** to completely solve current urban problems, but...

- More opportunities
- Support responsive and effective urban systems
- Contribute to positive social impacts and planning strategies

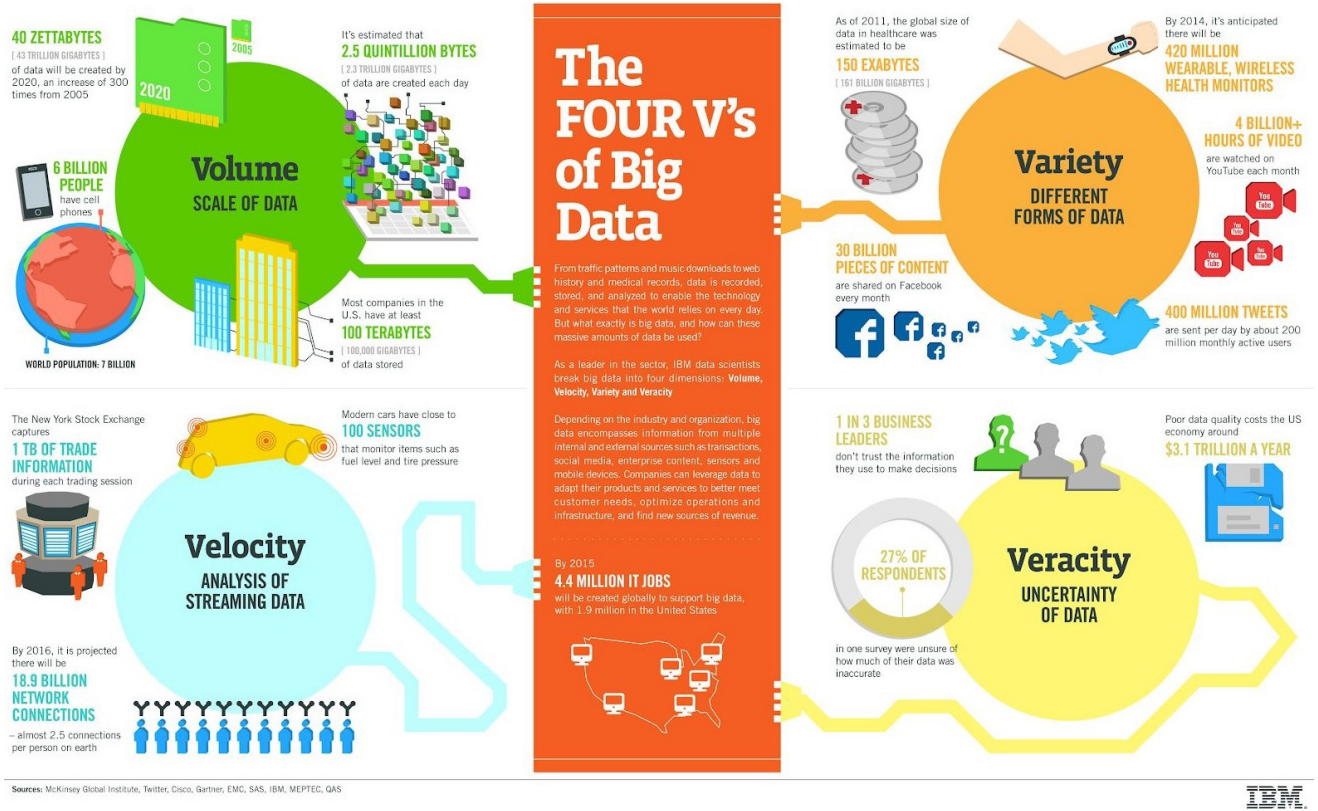

Applications to urban data
Understanding ML approaches
What we will do
Coding them in Python

Computational Thinking

- CT is a set of problem solving methods that involve expressing problems and their solutions in ways that a computer could execute.
- CT is essential to the development of computer applications, but it can also be used to support problem solving across all disciplines, including math, science, and the humanities.

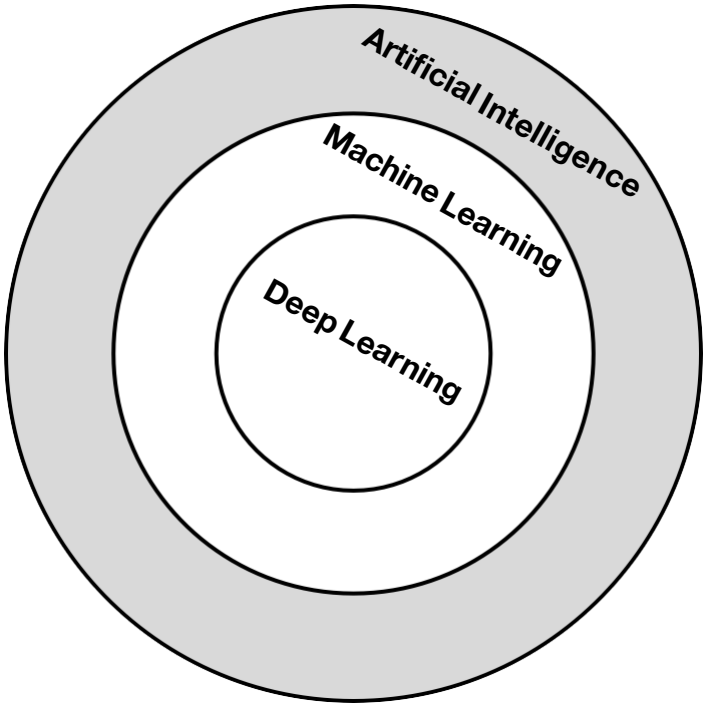

COMPUTATIONAL THINKING
DECOMPOSITION
Breaking big problems into smaller, easier to mange problems
PATTERN RECOGNITION
Analyze & look for a repeating sequence
21
B
5
C
13
3
2
1
8
C
Remove parts of a problem that are unnecessary and make one solution work for multiple problems
ABSTRACTION
Step-by-Step instructions on how to do something
ALGORITHM DESIGN

Big Data


40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005
2005
2020
6 BILLION PEOPLE
have cell phones
WORLD POPULATION: 7 BILLION
Volume
SCALE OF DATA
It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day
Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored
The Four V's of Big Data
From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?
As a leader in the sector, IBM data scientists break big data into four dimensions: Volume, Velocity, Variety and Veracity
Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.
By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States
As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]
Variety
DIFFERENT FORMS OF DATA
By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS
4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month
30 BILLION PIECES OF CONTENT
are shared on Facebook every month
400 MILLION TWEETS
are sent per day by about 200 million monthly active users
The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session
Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure
Velocity
ANALYSIS OF STREAMING DATA
By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth
1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions
Poor data quality costs the US economy around
$3.1 TRILLION A YEAR
27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate
Veracity
UNCERTAINTY OF DATA
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS
IBM

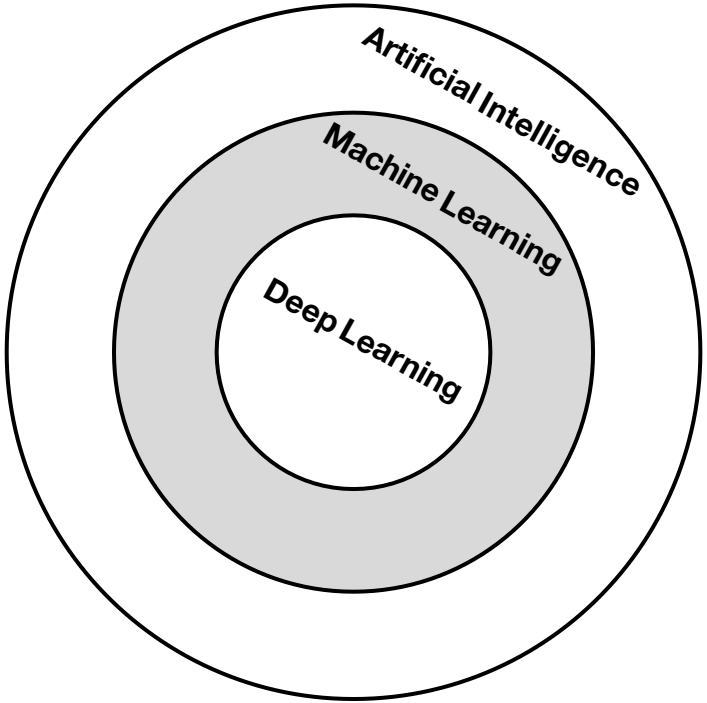
Artificial Intelligence

- Branch of computer science concerned with making computers behave like humans
- Visual perception, speech recognition, decision-making, and translation between languages.



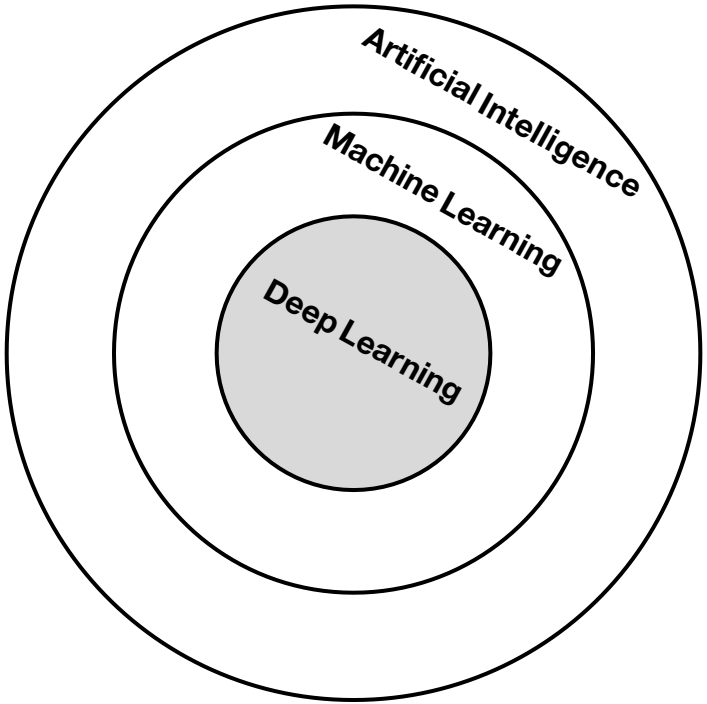
Machine Learning

- Subtopic within AI
- Includes abstruse statistical techniques that enable machines to improve at tasks with experience.
- Extracting knowledge from data



Deep Learning

- Subset of ML
- Composed of algorithms that permit software to train itself to perform tasks
- By exposing multilayered neural networks to vast amounts of data



ML, Improving Performance via Experience

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T as measured by P, improves with experience.”

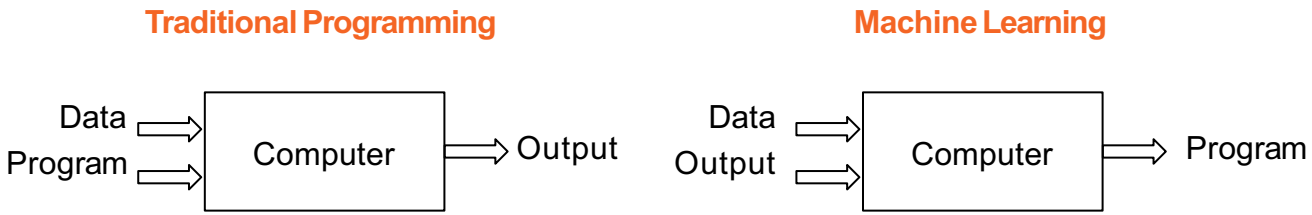
T. Mitchell

Examples of T and P Measures

Task (T)	Performance metric (P)	Experience (E)
Play checkers	% of wins against given opponent	Games previously played with outcomes
Recognize handwritten digits	% of correct recognitions	Set of digit writing with labels
Control a self driving car	Average speed in given conditions provided that safety standards are met	Previous driving record and its evaluation
Predict stock prices	Average prediction accuracy	History of stick prices

Objectives of ML

- Pattern recognition
- Learning strategies of operation
- Prediction



ML Framework

Data preparation	Explore, select, collect, store, and retrive data
Abstraction	Pick up a machine learning model and select performance metric depending on the unknown parameters to learn as well as the dataset to train it
Modeling	Look for the set of model parameters that optimize the given performance metric
Evaluation and Generalization	Evaluate different models and finally pick the best one to turn into a reusable form

ML Application Examples

- Classifying spam emails
- Netflix recommendations
- Reading checks
- Cancer diagnosis
- Weather forecast
- Credit card fraud detection
- Predicting customer demand
- Message auto correction
- Document classification

ML Urban Examples

- Predicting real estate prices
- Classification of land types by satellite images
- Slum area detection
- Sound recognition for emergency response
- Self-driving vehicles
- Identifying at-risk buildings
- Building energy usage prediction
- Activity detection from social media
- Estimating parking capacity
- Disaster response
- Population estimation

Types of Machine Learning

Machine Learning

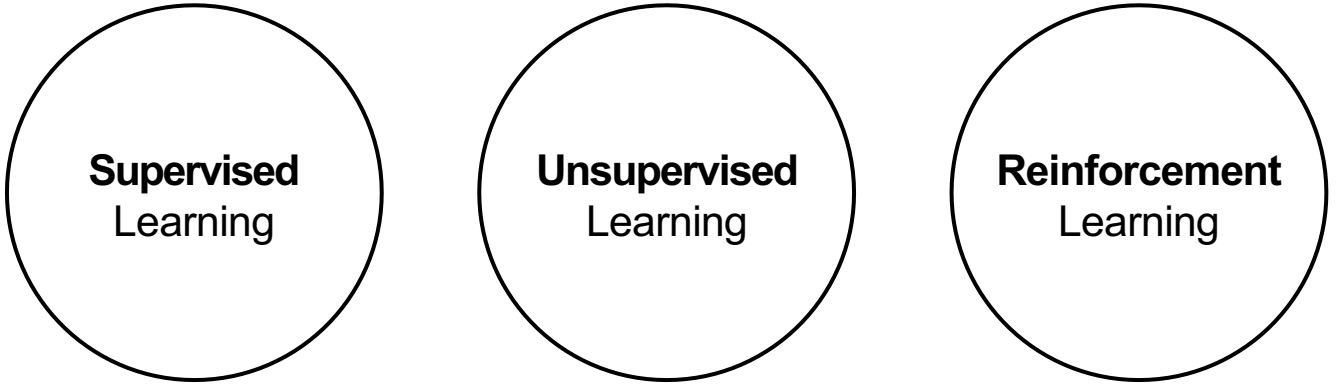
- Extracting knowledge from data
- Programming computers to optimize a performance criterion using example data or past experience
- Extracting automatically the algorithm for the task
- Developing a good learner that accurately predicts an outcome

Different roles of ML

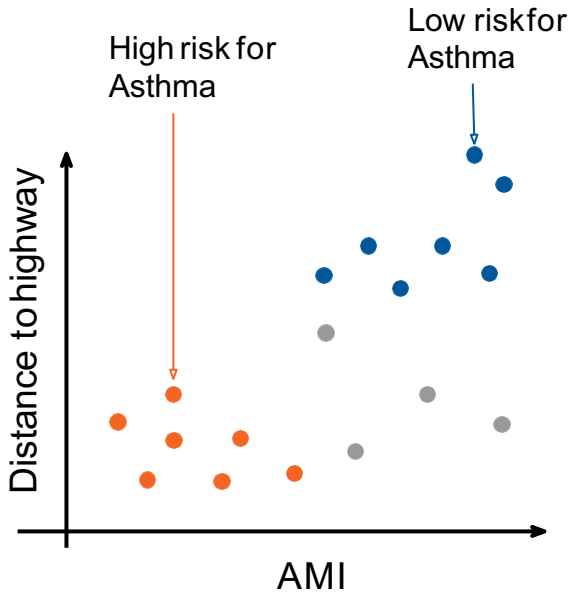
- Descriptive - learning association
- Predictive

- Example 1
  - Identifying recycling behavior across the city
  - Predicting community level diversion rate next year
- Example 2
  - Understanding consumers' shopping pattern
  - Predicting potential customers

Types of ML


Supervised
Learning
Unsupervised
Learning
Reinforcement
Learning

Supervised Learning



Supervised Learning

- **Input/output pairs**
- Learn the mapping from the input to the output whose **correct values are provided by a supervisor**
- Easy to measure performance
- $y = g(x|\theta)$ where $g()$ is the model and θ is its parameters
- Regression or classification cases

Examples of Supervised Learning

- Identifying the zip code from handwritten digits
- Determining whether a tumor is benign based on a medical image
- Detecting credit card fraud detection
- Predicting real estate prices based on property characteristics
- Forecasting household solid waste generation
- Detecting unusual building energy consumption

Unsupervised Learning

Data sample

Cluster/Group

Unsupervised Learning

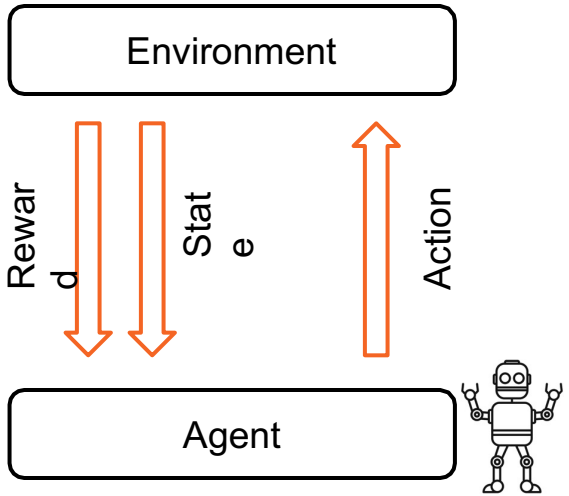
- **Only the input data** is known
- No known output data is given to the algorithm
- Find the **regularities/patterns** in the input
- Harder to understand and evaluate

Examples of Unsupervised Learning

- Document clustering
- Segmenting customers into groups with similar preferences
- Detecting abnormal access patterns to a website
- Identifying evacuation pattern across the neighborhoods
- Categorizing neighborhoods based on service requests

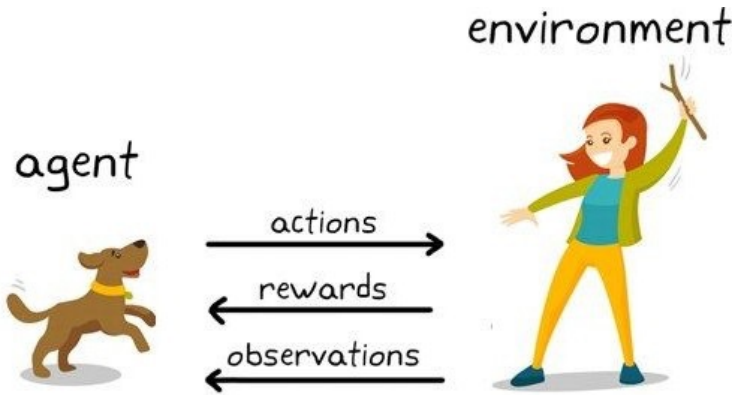
Reinforcement Learning

- Take actions in an environment in order to maximize some notion of cumulative reward
- The sequence of correct actions to reach the goal



Example of Reinforcement Learning

- Educating pets
- Game playing



Lab01 - Intro to Python for ML1

Why Python?

- Has become the most popular programming language for many data science applications (http://pypl.github.io/PYPL.html)
- Combination of the power of general-purpose programming languages and the ease of use of domain-specific scripting languages like MATLAB of R
- General- and special-purpose functionality: data loading, visualization, statistics, and more
- Ability to interact directly with the code, using a terminal or other tools like the **Jupyter Notebook**
- Quick and easy interaction and iteration

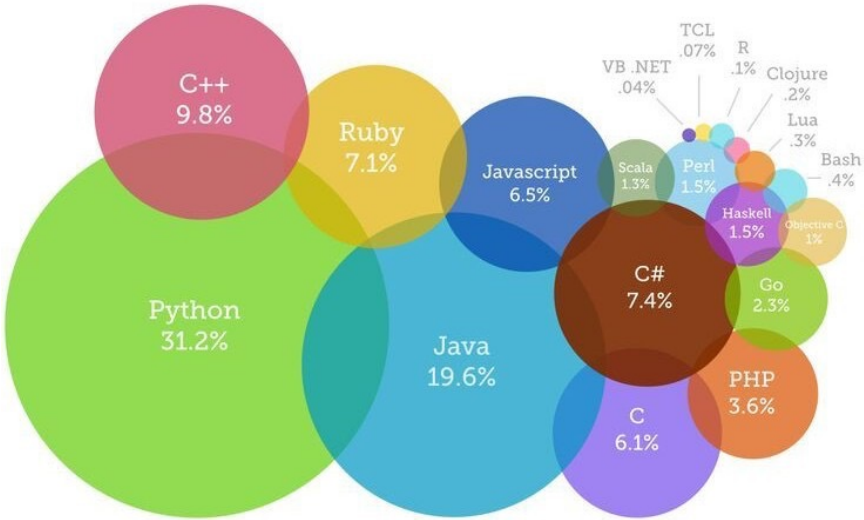
Primary Programming Language: Python

Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

KLINT FINLEY BUSINESS 03.02.2020 02:15 PM

Python Is More Popular Than Ever

Despite a rocky transition from Python 2 to Python 3, developers are still flocking to the programming language.



Reproducibility

- Direct interaction
- Automation
- Reusability
- Reproducibility

- Ability of an algorithm to produce repeated measurements
- When you share the pipeline you would like to make sure that your pipeline will produce the same output for other users when provided with the same input data
- Important to develop clear, transparent, and shareable pipeline and resources
  - Data
  - Metadata
  - Script
  - Documentation

How to write beautiful Python code

https://realpython.com/python-pep8/

—

What YouNeed

- Python3.7: programming language
- Jupyter: has a role of a specific software or an application to communicate with the computer
- Required basic packages
- Terminal or Bash
- ...

Install **Anaconda**

Step 1. Go to this link

https://docs.anaconda.com/anaconda/install/

Installation

Review the system requirements listed below before installing Anaconda Individual Edition. If you don't want the hundreds of packages included with Anaconda, you can install Miniconda, a mini version of Anaconda that includes just conda, its dependencies, and Python.

Tip

Looking for Python 3.5 or 3.6? See our FAQ.

Step 2.

Check system requirements first.

System requirements

- License: Free use and redistribution under the terms of the End User License Agreement - Anaconda® Individual Edition.
- Operating system: Windows 8 or newer, 64-bit macOS 10.13+, or Linux, including Ubuntu, RedHat, CentOS 6+, and others.
- If your operating system is older than what is currently supported, you can find older versions of the Anaconda installers in our archive that might work for you. See Using Anaconda on older operating systems for version recommendations.
- System architecture: Windows- 64-bit x86, 32-bit x86; MacOS- 64-bit x86; Linux- 64-bit x86, 64-bit Power8/Power9.
- Minimum 5 GB disk space to download and install.

On Windows, macOS, and Linux, it is best to install Anaconda for the local user, which does not require administrator permissions and is the most robust type of installation. However, if you need to, you can install Anaconda system wide, which does require administrator permissions.

- Installing on Windows
- Installing on macOS
- Installing on Linux
- Installing on Linux POWER
- Installing in silent mode
- Installing for multiple users
- Verifying your installation
- Anaconda installer file hashes
- Updating from older versions
- Uninstalling Anaconda

Installation

Review the system requirements listed below before installing Anaconda Individual Edition. If you don't want the hundreds of packages included with Anaconda, you can install Miniconda, a mini version of Anaconda that includes just conda, its dependencies, and Python.

Tip

Looking for Python 3.5 or 3.6? See our FAQ.

System requirements

- License: Free use and redistribution under the terms of the End User License Agreement - Anaconda® Individual Edition.
- Operating system: Windows 8 or newer, 64-bit macOS 10.13+, or Linux, including Ubuntu, RedHat, CentOS 6+, and others.
- If your operating system is older than what is currently supported, you can find older versions of the Anaconda installers in our archive that might work for you. See Using Anaconda on older operating systems for version recommendations.
- System architecture: Windows- 64-bit x86, 32-bit x86; MacOS- 64-bit x86; Linux- 64-bit x86, 64-bit Power8/Power9.
- Minimum 5 GB disk space to download and install.

On Windows, macOS, and Linux, it is best to install Anaconda for the local user, which does not require administrator permissions and is the most robust type of installation. However, if you need to, you can install Anaconda system wide, which does require administrator permissions.

Step 3.

Click for an instruction to install Anaconda on your machine

- Installing on Windows
- Installing on macOS
- Installing on Linux
- Installing on Linux POWER
- Installing in silent mode
- Installing for multiple users
- Verifying your installation
- Anaconda installer file hashes
- Updating from older versions
- Uninstalling Anaconda

Step 4. **ForWindows**

Installing on Windows

1. Download the Anaconda installer. **Open link in a new tab to keep the instruction page**
2. RECOMMENDED: Verify data integrity with SHA-256. For more information on hashes, see What about cryptographic hash verification?
3. Double click the installer to launch.

> ⓘ **Note**
>
> To prevent permission errors, do not launch the installer from the Favorites folder.

> ⓘ **Note**
>
> If you encounter issues during installation, temporarily disable your anti-virus software during install, then re-enable it after the installation concludes. If you installed for all users, uninstall Anaconda and re-install it for your user only and try again.

4. Click Next.
5. Read the licensing terms and click “I Agree”.
6. Select an install for “Just Me” unless you’re installing for all users (which requires Windows Administrator privileges) and click Next.
7. Select a destination folder to install Anaconda and click the Next button. See FAQ.

> ⓘ **Note**
>
> Install Anaconda to a directory path that does not contain spaces or unicode characters.

> ⓘ **Note**
>
> Do not install as Administrator unless admin privileges are required.

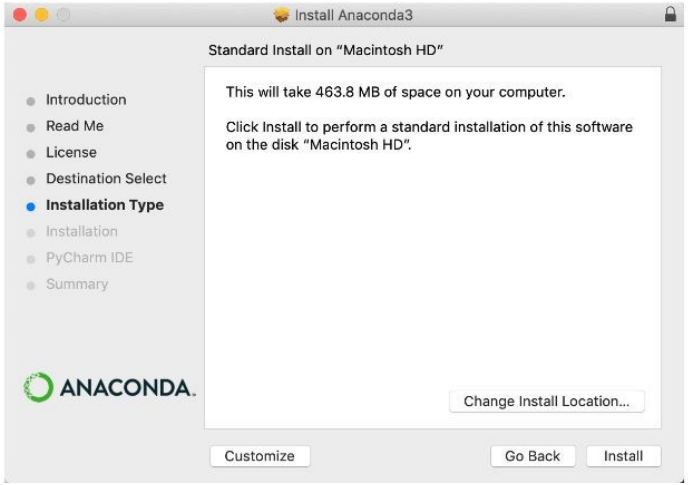
Step 4-1. **For macOS**

Installing on macOS

You can install Anaconda using either the graphical installer (“wizard”) or the command line (“manual”) instructions below. If you are unsure, choose the graphical install.

macOS graphical install

1. Download the graphical macOS installer for your version of Python. **Open link in a new tab to keep the instruction page**
2. RECOMMENDED: Verify data integrity with SHA-256. For more information on hashes, see What about cryptographic hash verification?
3. Double-click the downloaded file and click continue to start the installation.
4. Answer the prompts on the Introduction, Read Me, and License screens.
5. Click the Install button to install Anaconda in your ~/opt directory (recommended):



Using the command-line install

Follow the instruction to install Anaconda using the command-line

Step 4-2. For macOS

Use this method if you prefer to use a terminal window.

1. In your browser, download the command-line version of the macOS installer for your system.

2. RECOMMENDED: Verify data integrity with SHA-256. For more information on hash verification, see cryptographic hash validation.

   - Open a terminal and run the following:

```
shasum -a 256 /path/filename
```

> **Note**
>
> Replace `/path/filename` with your installation's path and filename.

3. Install for Python 3.7 or 2.7:

   - For Python 3.7 enter the following:

```
bash ~/Downloads/Anaconda3-2020.02-MacOSX-x86_64.sh
```

   - For Python 2.7, open the Terminal.app or iTerm2 terminal application and then enter the following:

```
bash ~/Downloads/Anaconda2-2019.10-MacOSX-x86_64.sh
```

> **Note**
>
> Include the `bash` command regardless of whether or not you are using the Bash shell.

> **Note**
>
> Replace `~/Downloads` with your actual path and `Anaconda3-2020.02-MacOSX-x86_64.sh` with actual name of the file you downloaded.

4. The installer prompts "In order to continue the installation process, please review the license agreement." Click Enter to view the license terms.

Individual Edition

Your data science toolkit

With over 20 million users worldwide, the open-source Individual Edition (Distribution) is the easiest way to perform Python/R data science and machine learning on a single machine. Developed for solo practitioners, it is the toolkit that equips you to work with thousands of open-source packages and libraries.

Download

Click this

Open Source

Anaconda Individual Edition is the world's most popular Python distribution platform with over 20 million users worldwide. You can trust in our long-term commitment to supporting the Anaconda open-source ecosystem, the platform of choice for Python data science.

Conda Packages

Search our cloud-based repository to find and install over 7,500 data science and machine learning packages. With the conda-install command, you can start using thousands of open-source Conda, R, Python and many other packages.

Manage Environments

Individual Edition is an open source, flexible solution that provides the utilities to build, distribute, install, update, and manage software in a cross-platform manner. Conda makes it easy to manage multiple data environments that can be maintained and run separately without interference from each other.

Step 5. **For Windows & macOS**

(for macOS users who want to use graphical installer)

Step 5. For Windows & macOS

Anaconda Installers

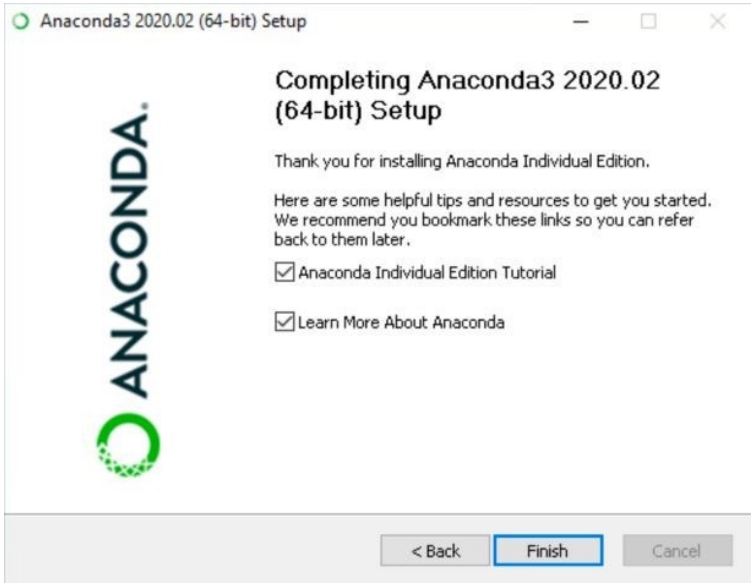
Download Anaconda Installers based on your OS

Windows	MacOS	Linux
Python 3.8	Python 3.8	Python 3.8
64-Bit Graphical Installer (466 MB)	64-Bit Graphical Installer (462 MB)	64-Bit (x86) Installer (550 MB)
32-Bit Graphical Installer (397 MB)	64-Bit Command Line Installer (454 MB)	64-Bit (Power8 and Power9) Installer (290 MB)

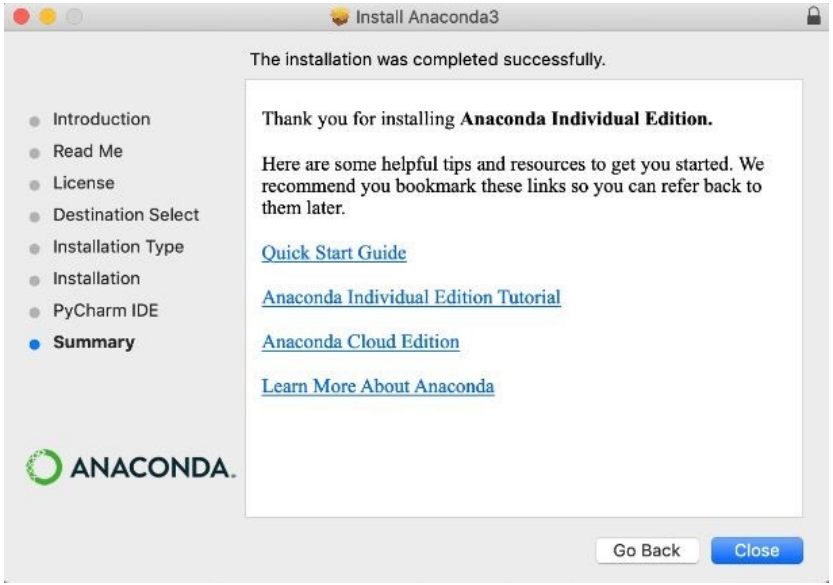
Please follow the instruction and install Anaconda.

Anaconda should be installed to do the first homework.

Windows



macOS



Other Platforms

- Google CoLab (https://colab.research.google.com/notebooks/)