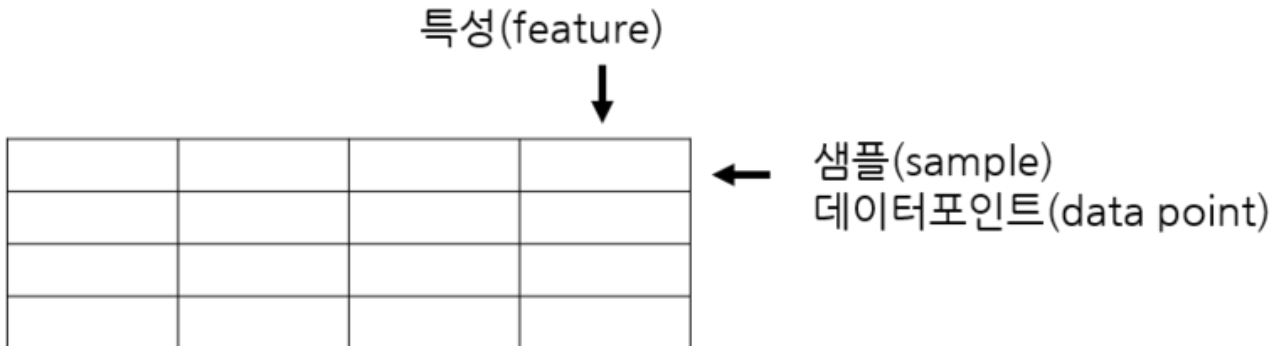


1.1 머신러닝

- 지도학습(supervised learning): 이미 알려진 사례를 바탕으로 일반화된 모델을 만들어 의사결정 프로세스를 자동화하는 것
 - 비지도학습(unsupervised learning): 알고리즘에 입력은 주어지지만 출력은 제공되지 않음
- 지도학습과 비지도학습의 공통점: 컴퓨터가 인식할 수 있는 형태로 입력 데이터를 준비하는 것이 중요!



참고) 특성추출(Feature extraction), 특성공학(Feature engineering)

- 고민해야 할 점
- 1) 어떤 질문에 대한 답을 원하는가? 가지고 있는 데이터가 원하는 답을 줄 수 있는가?
 - 2) 내 질문을 머신러닝의 문제로 가장 잘 기술하는 방법은 무엇인가?
 - 3) 문제를 풀기에 충분한 데이터를 모았는가?
 - 4) 내가 추출한 데이터의 특성은 무엇이며 좋은 예측을 만들어낼 수 있을 것인가?
 - 5) 머신러닝 애플리케이션의 성과를 어떻게 측정할 수 있는가?
 - 6) 머신러닝 솔루션이 다른 연구나 제품과 어떻게 협력할 수 있는가?

1.2 파이썬

- 데이터 적재, 시각화, 통계, 자연어 처리, 이미지 처리 등에 필요한 라이브러리 보유
- 대화형 프로그래밍(Jupyter Notebook)

1.3 Sickit-learn

- 파이썬 머신러닝 라이브러리 (Numpy, Scipy 이용)

- 알고리즘 설명 문서: <https://scikit-learn.org/stable/documentation.html>

Documentation scikit-learn: machine learning in Python — scikit-learn 0.21.3 documentation

Documentation of scikit-learn 0.21.3 Quick Start A very short introduction into machine learning problems and how to solve them using scikit-learn. Presents basic concepts and conventions. User Guide The main documentation. This contains an in-depth descriptio...

scikit-learn.org

- 사용자 가이드: https://scikit-learn.org/stable/user_guide.html

User guide: contents — scikit-learn 0.21.3 documentation

User Guide 1. Supervised learning 1.1. Generalized Linear Models 1.2. Linear and Quadratic Discriminant Analysis 1.3. Kernel ridge regression 1.4. Support Vector Machines 1.5. Stochastic Gradient Descent 1.6. Nearest Neighbors 1.7. Gaussian Processes 1.8. Cross decomposition 1.9. Naive Bayes 1.10. D...

scikit-learn.org

- API 문서: <https://scikit-learn.org/stable/modules/classes.html>

API Reference — scikit-learn 0.21.3 documentation

API Reference This is the class and function reference of scikit-learn. Please refer to the full user guide for further details, as the class and function raw specifications may not be enough to give full guidelines on their uses. For reference on concepts repeated across the API, see Glossary of Co...

scikit-learn.org

● 개발환경

1) Anaconda: <https://www.anaconda.com/>



Anaconda | The World's Most Popular Data Science Platform

Anaconda is the standard platform for Python data science, leading in open source innovation for machine learning. Develop, manage, collaborate, and govern at scale with our enterprise platform.

www.anaconda.com

2) Numpy: <https://numpy.org/>



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...

numpy.org

- 파이썬으로 과한 계산 시 꼭 필요한 패키지
- 다차원(n-차원) 배열인 ndarray 클래스 이용 (배열의 모든 원소는 동일한 데이터 타입이어야 함)

3) Scipy: <https://www.scipy.org/scipylib/>

SciPy library — SciPy.org

SciPy.org SciPy library The SciPy library is one of the core packages that make up the SciPy stack. It provides many user-friendly and efficient numerical routines such as routines for numerical integration, interpolation, optimization, linear algebra and statistics. For tutorials, reference documen...

www.scipy.org

- 과학 계산용 함수를 모아놓은 파이썬 패키지
- 희소행렬(sparse matrix): 0 을 많이 포함한 2 차원 배열 저장 시 사용 (scipy.sparse)
- 참고사이트: <http://scipy-lectures.org/>

Scipy Lecture Notes — Scipy lecture notes

Scipy Lecture Notes One document to learn numerics, science, and data with Python Tutorials on the scientific Python ecosystem: a quick introduction to central tools and techniques. The different chapters each correspond to a 1 to 2 hours course with increasing level of expertise, from beginner to e...

scipy-lectures.org

4) Matplotlib: <https://matplotlib.org/>

Matplotlib: Python plotting — Matplotlib 3.1.1 documentation

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 graph...

matplotlib.org

- 파이썬의 대표적인 과학 계산용 그래프 라이브러리

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

Python Data Analysis Library — pandas: Python Data Analysis Library

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 wor...

pandas.pydata.org

- 데이터 처리와 분석을 위한 파이썬 라이브러리
- DataFrame 이라는 데이터 구조를 기반으로 사용
- 전체 배열의 원소가 동일한 타입이어야 하는 Numpy 와 달리 각 열의 타입이 달라도 됨
- 파일과 데이터베이스에서 데이터를 읽어들이 수 있음