

Machine Learning from Scratch with Python

본 강의는 [TEAMLAB](#)과 [Inflearn](#)이 함께 구축한 데이터 사이언스 과정의 두 번째 강의인 [밑바닥 부터 시작하는 머신러닝 입문](#)입니다. [밑바닥부터 시작하는 머신러닝 입문](#)은 Part I과 Part II로 구성되어 있습니다.

본 강의는 [TEAMLAB](#)과 [Inflearn](#)이 함께 준비한 [WADIZ 펀딩](#)의 지원을 받아제작되었습니다.

아래 목록에 대한 강의를 개발할 예정입니다.

- [데이터 과학을 위한 파이썬 입문](#) - 개발 완료
- Machine Learning from Scratch with Python Part I - 본과정
- Machine Learning from Scratch with Python Part II

또한 기존 K-MOOC 과정은 아래 목록을 참고하시기 바랍니다.

- [K-MOOC: 데이터 과학을 위한 파이썬 입문](#)
- [Operation Research with Python Programming](#)

Course overview

- 강좌명: 밑바닥 부터 시작하는 머신러닝 입문(Machine Learning from Scratch with Python)
- 강의자명: 가천대학교 산업경영공학과 최성철 교수 (sc82.choi@gachon.ac.kr, Director of [TEAMLAB](#))
- Facebook: [Gachon CS50](#)
- Email: teamlab.gachon@gmail.com

Course Info

- 본 과정은 머신러닝에 대한 기초개념과 주요 알고리즘들에 대해 이해하고 구현하는 것을 목적으로 함
- 본 과정을 통해 수강자는 데이터 과학에서 사용되는 다양한 용어에 대한 기본적인 이해를 할 수 있음
- 본 과정의 기본적인 구성은 알고리즘에 대한 설명, [Numpy](#)를 사용한 구현, [Scikit-Learn](#)을 사용한 패키지 활용으로 이루어져 있음
- 수강자는 머신러닝에서 주로 사용되는 알고리즘을 구현하기 위해 고등학교 수준의 통계학과 선형대수학의 이해가 필요함
- 수강자는 본 과정을 통해 Numpy, Pandas, Matplotlib, Scikit-Learn 등 데이터 분석을 위한 기본적인 파이썬 패키지를 이해하게 됨

Course Contents

Chapter 1 - Introduction to Machine Learning

- Chapter Intro
- Machine Learning Overview - [\[강의영상\]](#), [\[강의자료\]](#)
- An Understanding of the Data Keywords - [\[강의영상\]](#), [\[강의자료\]](#)
- How to Learn Machine Learning - [\[강의영상\]](#), [\[강의자료\]](#)
- Types of Machine Learning - [\[강의영상\]](#), [\[강의자료\]](#)
- A History of Data Analysis: In Perspective of Business - [\[강의영상\]](#), [\[강의자료\]](#)
- Environment Setup
 - Python Ecosystem for Machine Learning - [\[강의영상\]](#), [\[강의자료\]](#)
 - Pycharm 설치(Mac) - [\[강의영상\]](#), [\[강의자료\]](#)
 - How to use Jupyter Notebook - [\[강의영상\]](#), [\[강의자료\]](#)
- 참고자료
 - 가상환경과 Package 활용하기 - [강의영 상](#), [강의자료](#)

Chapter 2 - Warm Up Section: An understanding of data

Lecture

- Chapter Intro
- The concepts of a feature - [강의영상](#), [강의자료](#)
- Data types - [강의영상](#), [강의자료](#)
- Loading data with pandas - [강의영상](#), [강의자료](#)
- Representing a model with numpy - [강의영상](#), [강의자료](#)

Chapter 3 - Numpy Section

Lecture

- Chapter Intro - Numpy
- Numpy overview
- ndarray
- Handling shape
- Axis 이해하기 - [강의영상_테스트](#)
- Indexing & Slicing
- Creation functions
- Opertaion functions
- Array operations
- Comparisons
- Boolean & fancy Index
- Numpy data i/o
- Lab assignment

Supplements

- TF-KR 첫 모임: Zen of NumPy - [발표자료](#), [강의영상](#) (하성주, 2016)

Chapter 4 - Pandas Section #1

Lecture

- Chapter Intro - Pandas
- Pandas overview
- Series - [강의영상](#), [강의자료](#), [code](#)
- DataFrame - [강의영상](#), [강의자료](#), [code](#)
- Selection & Drop
- Dataframe operations
- lambda, map apply
- Pandas built-in functions
- Lab assignment

Chapter 5 - Pandas Section #2

Lecture

- Groupby
- Pivot table & Crosstab
- Merg & Concat
- Lab assignment
- Database connection & Persistence
- Data Cleaning Problem Overview - [강의영상](#) [강의자료](#)
- Missing Values - [code](#)
- Categorical Data Handling - [code](#)
- Feature Scaling - [강의영상](#), [강의자료](#), [code](#)

Chapter 7 - Matplotlib Section & Miniproject

Lecture

- Chapter overview - Matplotlib overview
- Basic functions & operations
- Graph
- Matplotlib with pandas
- Casestudy - KaggleProblems
- Miniproject - Preprocessing works for House Price Problems

Chapter 8 - Linear Regression

Lecture

- Linear regression overview
- Cost functions
- Linear Equality
- Gradient descent approach
- Linear regression with gradient descent
- Linear regression with Numpy
- Multivariate linear regression models
- Multivariate linear regression with NumPy
- Lab Assignment

Chapter 9 - Linear Regression extended

Lecture

- Overfitting - bias vs. variance
- Regularization - L1 and L2
- Implementation of generalization with NumPy
- Linear regression with sklearn
- Polynomial regression
- sklearn SGD family
- Performance measure
- Training, test and Validation concepts

Chapter 10 - Logistics Regression

Lecture

- Logistic regression overview - 강의자료, [code](#)
- Sigmoid function - 강의자료, [code](#)
- Cost function - 강의자료, [code](#)
- Logistic regression implementation with Numpy- 강의자료, [code](#)
- Maximum Likelihood estimation - 강의자료
- Regularization problems
- Logistic regression with sklearn

Chapter 11 - Logistics Regression extended

Lecture

- Softmax function for Multi-class classification - [강의자료](#)
- Cross entropy loss function - [강의자료](#)
- Softmax regression - [강의자료](#)
- Performance measures for classification
- ROC Curve & AUC
- Hyperparameter searching
- Data sampling method
- Handling imbalanced dataset - Oversampling, Undersampling, and SMOTE

Chapter 12 - Naive Bayesian Classifier

Lecture

- Probability overview - [강의자료](#)
- Bayes theorem - [강의자료](#)
- Single variable bayes classifier - [강의자료](#), [code](#)
- Naive bayesian Classifier - [강의자료](#), [code](#)
- NB classifier with sklearn - [code](#)
- Gaussian Normalization for Naive Bayesian

Chapter 13 - Decision Tree

Lecture

- Decision tree overview - [강의자료](#)
- The concept of entropy - [강의자료](#)
- The algorithm of growing decision tree - [강의자료](#)
- ID3 & Information gain - [강의자료](#)
- CART & Gini Index - [강의자료](#)
- Decision Tree with sklearn - [\[강의자료\]](#)
- Handling a continuous attribute - [강의자료](#)
- Decision Tree for Regression - [강의자료](#)
- Tree pruning - [강의자료](#)
- Regression Tree with sklearn - [코드](#)

Chapter 14 - How to improve a performance of your model

Lecture

- Chapter intro
- Ensemble model overview
- Random Forest
- Boosting, Bagging, AdaBoost
- Implementing ensemble classifier with sklearn
- Gradient boosting - XGBoost, GBM & LightGBM
- Stacking
- Feature Engineering
- Hyperparameter searching advanced
- Hyperparameter searching with Parallel training
- AutoML

Chapter 15 - Support Vector Model

Lecture

Chapter 16 - Neural Network

Lecture

참고자료

- [Machine Learning](#) (Coursera) by Andrew Ng
- [모두를 위한 딥러닝](#) by Sung Kim
- [C++로 배우는 딥러닝](#) by Sung Kim
- Machine Learning From Scratch[<https://github.com/eriklindernoren/ML-From-Scratch>]

Textbooks

- Reading materials
- [밑바닥부터 시작하는 데이터 과학](#)(조엘 그루스, 2016)
- [파이썬 머신러닝](#)(세바스티안 라슈카, 2016)
- [Hands-On Machine Learning with Scikit-Learn and TensorFlow](#)(Aurélien Géron, 2017, [PDF](#))
- [Data Mining: Concepts and Techniques](#)(Jiawei Han, Micheline Kamber and Jian Pei, 2011, [PDF](#))

- Supplementary textbooks
- [파이썬 라이브러리를 활용한 데이터 분석](#)(웨스 맥키니, 2013)
- [머신러닝 인 액션](#)(피터 해링턴, 2013)
- [데이터 과학 입문](#)(레이철 슈트 | 캐시 오닐, 2014)
- [머신러닝 인 파이썬](#)(마이클 보울즈, 2015)
- [머신러닝 이론 입문](#)(나카이 에츠지, 2016)

Prerequisites - 수강전 이수 또는 수강중 들었으면 하는 교과들

- 입문 수준의 통계학
 - [세상에서 가장 쉬운 통계학](#)(고지마 히로유키, 2009)
 - [세상에서 가장 쉬운 베이지통계학입문](#)(고지마 히로유키, 2017)
 - [확률과통계](#)(한양대학교 이상화 교수, 2014)
 - Reading Materials: Data Science from the Scratch - Ch.5, Ch.6, Ch.7
- 고교 이과 수준의 선형대수학 (Matrix와 Vector의 기본개념은 Review 필요)
 - [Essence of linear algebra](#)(3Blue1Brown, 2017)
 - [Linear Algebra](#)(Khan Academy)
 - [선형대수학](#)(한양대 이상화 교수, 2013) - Advance Course
 - Reading Materials - Data Science from the Scratch - Ch.4
- 고교 이과 수준의 미적분학 (개념에 대한 이해 필요)
 - [Essence of calculus](#)(3Blue1Brown, 2017)
- 파이썬 기초
 - [데이터 과학을 위한 파이썬 입문](#) (TEAMLAB, 2017)
- Git
 - [Pro Git](#) (스캇 샤콘 | 벤 스트라움, 2016)
 - [Git & Github](#) (TEAMLAB, 2016)
 - [Git 강의](#) (생활코딩, 2014)