

수업계획서

2022년 1학기

○ 과목명 : 파이썬 머신러닝 (GITA315)	○ 담당강사 : 최명환 교수
○ 학 점 : 2 학점	○ 시 간 : 목 18:30 ~ 20:10
○ 강의실 : AS 호	
○ 조 교 : TBD	

1. 교과목표

머신러닝 이론과 기술을 소프트웨어적으로 구현하는 기술을 학습함으로써 그 이론과 구현 기술에 대해 이해를 더욱 깊게 하고 활용 능력을 배양하도록 한다. 이를 위해서 파이썬 언어를 사용하며 이를 통해 머신러닝 기술이 실제로 구현되는 방식과 이에 관련된 자료는 어떻게 처리되는 지를 학습한다.

이 강의는 파이썬, scikit-learn, 그리고 TensorFlow로 배우는 머신러닝과 딥러닝의 전반부에 해당하는 강의이며 여러 머신러닝 기술 각각에 대해 기초 이론을 적절한 수준으로 소개하고 예제를 통해 이들을 프로그램으로 구현하고 동작 특성을 파악하도록 한다. 이는 머신러닝 기술에 대한 이해를 높이고 구현과 활용 기술을 습득할 수 있도록 할 것이며 더욱 복잡한 딥러닝에 대한 이해와 구현 기술 습득의 기반이 될 것이다.

2. 수업형태

강의 : 100%

- 수업은 100% 강의로 진행되나 강의 내용을 무난히 파악할 수 있기 위해서는 Python에 대한 충분한 이해가 필요합니다.
- 강의는 "동영상 강의 업로드" 방식과 "zoom을 통한 실시간 강의" 방식으로 진행될 예정입니다. 그리고 실시간 강의는 사전에 안내해 드립니다.

3. 강의계획

가. 교과 개요

주	교수내용	수업형태	비고
1	Overview: giving computers the ability to learn from data	강의	intro
2	Training simple machine learning algorithms for classification	강의	
3	Training simple ML algorithms for classification - implementation	강의	
4	Machine learning classifiers using scikit-learn	강의	
5	Machine learning classifiers using scikit-learn – implementation	강의	
6	Building good training data: Data preprocessing	강의	
7	Compressing data via dimensionality reduction	강의	
8	Midterm exam	시험	
9	Compressing data via dimensionality reduction	강의	
10	Model evaluation and hyperparameter tuning	강의	
11	Ensemble learning	강의	
12	Ensemble learning	강의	
13	Applying machine learning to sentiment analysis	강의	
14	Embedding a machine learning model into a Web application	강의	
15	Predicting continuous target variables with regression analysis, Clustering analysis (if time permits)	강의	
16	Final exam	시험	

나. 교과 내용

Chapter 1: Giving computers the ability to learn from data

- Building intelligent machines to transform data into knowledge
- Using Python for machine learning

Chapter 2: Training simple machine learning algorithms for classification

- Artificial neurons
- Implementing a perceptron learning algorithm in Python
- Adaptive linear neurons and the convergence of learning

Chapter 3: Tour of machine learning classifiers using scikit-learn

- Modeling class probabilities via logistic regression
- Maximum margin classification with support vector machines
- Solving nonlinear problems using a kernel SVM
- Decision tree learning
- K-nearest neighbors

Chapter 4: Building good training datasets – Data Preprocessing

- Dealing with missing data

- Handling categorical data
- Selecting meaningful features
- Assessing features importance with random forests

Chapter 5: Compressing data via dimensionality reduction

- Unsupervised dimensionality reduction via principal component analysis
- Supervised data compression via linear discriminant analysis
- Using kernel principal component analysis for nonlinear mappings

Chapter 6: Learning best practices for model evaluation and hyperparameter tuning

- K-fold cross-validation
- Debugging algorithms with learning and validation curves
- Fine-tuning machine learning models via grid search

Chapter 7: Combining different models for ensemble learning

- Learning with ensembles
- Combining classifiers via majority vote
- Bagging
- AdaBoost

Chapter 8: Applying machine learning to sentiment analysis

- Preparing the IMDb movie review data for text processing
- Introducing the bag-of-words model
- Training a logistic regression model for document classification

Chapter 9: Embedding a machine learning model into a Web application

- Saving the current state of a trained machine learning model
- Developing a web application using the Flask web framework

Chapter 10: Predicting continuous target variables with regression analysis

- Simple and Multiple linear regression
- Implementing an ordinary least squares linear regression model
- Dealing with nonlinear relationships using random forests

Chapter 11: Working with unlabeled data – clustering analysis

- Grouping objects by similarity using k-means
- Organizing clusters as a hierarchical tree
- Locating regions of high density via DBSCAN

4. 교재 및 자료

가. 교재 :

S. Raschka and V. Mirjalili, Python Machine Learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow2, 3rd Edition, Packt, 2019.

나. 참고자료 및 도서 :

a) M. Mohri et al., Foundations of Machine Learning, The MIT Press, 2012.

a) Tom M. Mitchell, Machine Learning, McGraw-Hill, 1977.

다. 강의자료: 서강대학교 사이버캠퍼스에서 다운로드할 수 있음.

5. 평 가

가. 중간고사 : 30%

나. 기말고사 : 40%

다. homeworks: 30%

▶ 상황에 따라 중간고사를 보지 않을 경우에는 평가 반영 비율이 변경될 예정입니다.

▶ 중간고사, 기말고사 방식은 추후 안내 예정입니다.

6. 기 타

면담: 이메일로 면담 시간 정할 수 있습니다 (mchoi@sogang.ac.kr)