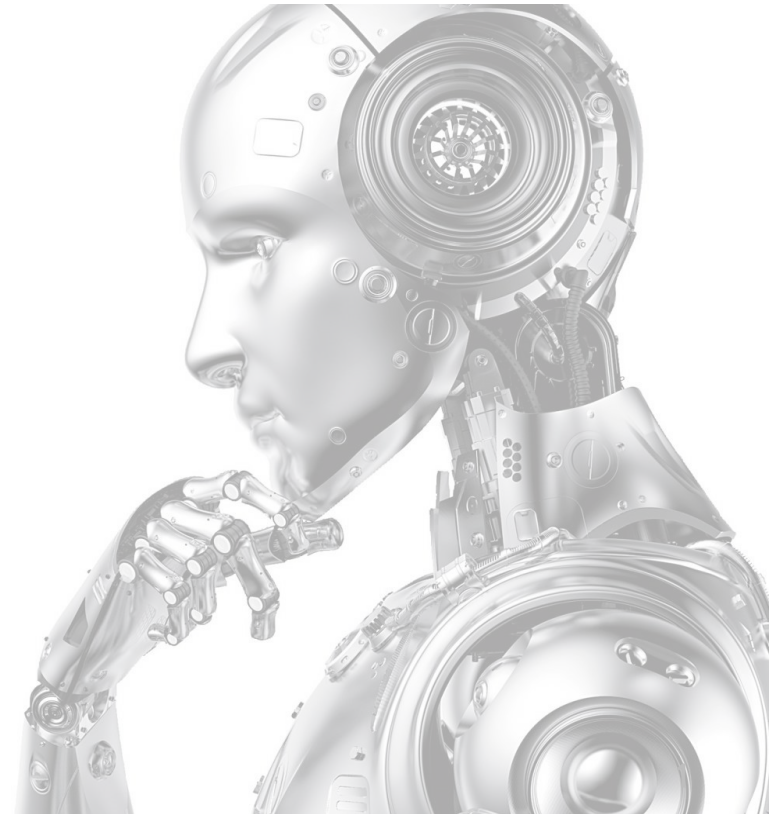


2021년 'LG전자 BS본부 DX아카데미'

머신러닝 이해와 활용 과정

학습안내



머신러닝 이해와 활용 과정

머신러닝의 개념과 작동원리를 배우고, 핵심 알고리즘을 구현해 봄으로써 추후 현업의 AI문제해결과제를 기획해 본다.

■ 교육개요

목표	<ul style="list-style-type: none"> 머신러닝의 본질을 이해하고, 머신러닝 기반 문제해결 프로젝트 프로세스를 이해한다. 다양한 머신러닝 알고리즘의 이론을 학습하고 구현해 봄으로써 각 알고리즘이 데이터를 이해하는 방식과 그에 따라 어떻게 다른 추론이 결정되는지 배운다. 다양한 알고리즘을 배우는 것을 넘어서 분석 상황에 따라 맞는 알고리즘을 선택할 수 있는 능력을 배양한다.
교육 대상	<ul style="list-style-type: none"> ‘파이썬 활용 데이터 분석’ 3일 과정 정도 수강한 분 파이썬으로 탐색적 데이터 분석은 하시는 분
교육 방식	<ul style="list-style-type: none"> 디지털 대면 강의 (Webex 접속)
교육 일시	<ul style="list-style-type: none"> 1차: 10/19(화)~10/22(금) 2차: 11/9(화)~11/12(금)
준비물	<ul style="list-style-type: none"> 사전설치: ①아나콘다 ②파이썬 + 주피터 노트북 설치 (①, ②중 한가지 설치) 사전가입: Github, colab 가입

■ 사전학습 방법

✓북러닝

- 도서명: ‘나의 첫 머신러닝/딥러닝’
- 머신러닝을 활용하고자 하는 분이라면 필수적으로 알아야 하는 기본적인 머신러닝, 딥러닝 알고리즘과 그에 필요한 지식을 단계별로 학습하고 실습하기에 좋은 책입니다.
- 1~4장까지 정독하시고(4장 일부 챕터 제외), 예제 파일을 다운로드 받아 코드의 전체 흐름을 한번 살펴보시길 추천드립니다.

✓파이썬 복습

- 파이썬 활용 데이터 분석을 배운지 6개월 이상이 되었다면, pandas 관련 예제를 한번 돌려 보고 오시기 바랍니다.

사전학습: 복러닝

[고급] 머신러닝 이해와 활용 과정

- **복러닝 의미:** 아래 1~ 4장의 내용을 정독하시고, 머신러닝에 자주 등장하는 필수 개념과 주요 알고리즘을 잘 이해하는 것은 매우 중요함
- [참고] 예제코드 주소: GitHub 저장소: <https://github.com/wikibook/machine-learning/>
ZIP 형식으로 다운로드: <https://github.com/wikibook/machine-learning/archive/master.zip>



저자: 허민석
실리콘밸리에서 머신러닝 개발
자로 활동 중

구분	모듈	상세내용	비고
1장	개발자가 처음 만난 머신러닝의 세계	1.1 머신러닝이란? 1.2 프로젝트 과정 미리보기 1.3 실습의 중요성	
2장	실습 준비	2.1 예제 코드 2.2 구글 코랩(Google Colaboratory)	
3장	자주 등장하는 머신러닝 필수 개념	3.1 지도학습과 비지도학습 3.2 분류와 회귀 3.3 과대적합과 과소적합 3.4 혼동 행렬 3.5 머신러닝 모델의 성능 평가 3.6 k-폴드 교차 검증	
4장	머신러닝 알고리즘 실습	4.1 머신러닝 알고리즘 실습 개요 4.2 k-최근접 이웃(k-Nearest Neighbor, kNN) 4.4 의사결정 트리 4.7 군집화 4.8 선형회귀 4.9 로지스틱 회귀	[읽지 않아도 되는 챕터] 4.3 서포트 벡터 머신 4.5 나이브 베이즈 4.6 앙상블 4.10 주성분 분석

1일차

머신러닝의 이해

: 머신러닝의 작동원리

구분	모듈	학습목표	학습내용	예제
오전 (4H)	오리엔테이션	머신러닝을 배워 앞으로 활용하려는 분야 및 과정에 대한 기대사항으로 자기 소개	- 과정 소개 및 강의 환경 점검, 강의 도구 사용 방법 소개 (WebEX) - 자기 소개 (담당 업무, 머신러닝을 활용하려는 분야 등)	-
	머신러닝 개요	자주 등장하는 머신러닝의 필수 개념을 이해하고, 다양한 머신러닝의 알고리즘을 언제, 어떻게 적용해야 하는지에 대해 배운다	- 머신러닝의 개념 - 머신러닝 알고리즘의 종류 - 지도학습 vs 비지도학습 - 지도학습 프로세스 - 데이터셋 분할: Hold-out validation, k-folds Cross validation	
오후 (4H)			- 회귀모형의 성능 평가: MSE, RMSE, MAE, MAPE - 분류모형의 성능 평가: Confusion Matrix, F1 Score, ROC, AUC	붓꽃 데이터셋
	데이터 전처리	데이터를 본격적으로 분석하기 이전에 주요 패키지를 사용하여 기초적인 전처리(preprocessing)를 하는 방법을 배운다	- 데이터 표준화 및 원핫 인코딩 - 데이터 균형화 - 불균형된 데이터셋 처리: SMOTE - 비유사도의 척도 - 다양한 거리 계산법: 맨하탄, 유클리드, Gower	

2일차

머신러닝의 주요 알고리즘

: 회귀분석

구분	모듈	학습목표	학습내용	예제
오전 (4H)	선형회귀분석	선형회귀분석은 연속형 목표변수 를 추정하는 지도학습 알고리즘이며, 입력변수마다 목표변수에 미치는 영향을 회귀계수로 제공하므로 모형을 쉽게 이해할 수 있다.	<ul style="list-style-type: none"> - 선형회귀분석의 개요 - 선형회귀분석 알고리즘 상세 - 회귀계수의 추정 - 모형 적합 및 결과 확인 - 회귀모형 성능 평가 - 정규화 회귀: Ridge, LASSO 	중고차 가격 추정 회귀 모형
오후 (4H)	로지스틱회귀분석	로지스틱회귀분석은 명목형 목표변수 를 분류하는 지도학습 알고리즘이며, 입력변수마다 목표변수에 미치는 영향을 회귀계수로 제공하므로 모형을 쉽게 이해할 수 있다.	<ul style="list-style-type: none"> - 로지스틱회귀분석의 개요 - 로지스틱회귀분석 알고리즘 상세 - 가능도 함수 - 모형 적합 및 결과 확인 - 분류모형 성능 평가 	개인대출 이용 분류 모형

3일차

머신러닝의 주요 알고리즘

: 나무모형

구분	모듈	학습목표	학습내용	예제
오전 (4H)	의사결정나무	의사결정나무는 분류와 회귀에 사용되는 대표적인 지도학습 알고리즘으로, 분류모형의 경우 명목형 목표변수를 분리하는 최적의 분리규칙을 제공한다.	<ul style="list-style-type: none"> - 의사결정나무의 개요 - 의사결정나무 알고리즘 상세 - 정지규칙 - 모형 적합 및 결과 확인 - 가지치기 - 모형 재적합 - 분류모형 성능 평가 	개인대출 이용 분류 모형
오후 (4H)	회귀나무	연속형 목표변수를 추정하는 의사결정나무 알고리즘을 회귀나무라고 하며, 목표변수의 분산을 최소로 하는 최적의 분리규칙을 제공한다.	<ul style="list-style-type: none"> - 회귀나무의 개요 - 회귀나무 알고리즘 상세 - 정지규칙 - 모형 적합 및 결과 확인 - 가지치기 - 모형 재적합 - 회귀모형 성능 평가 	중고차 가격 추정 회귀 모형

4일차

머신러닝의 주요 알고리즘

: 랜덤 포레스트, 그래디언트 부스팅

구분	모듈	학습목표	학습내용	예제
오전 (4H)	랜덤 포레스트	<p>의사결정나무 알고리즘은 한 개의 나무모형을 제시하지만 랜덤 포레스트는 수 백, 수 천 개의 나무모형을 동시에 만들어 결과를 통합하므로 성능이 우수한 모형을 제공한다.</p> <p>분류모형과 회귀모형을 모두 만들 수 있다.</p>	<ul style="list-style-type: none"> - 배깅 vs 부스팅 - 랜덤 포레스트의 개요 - 랜덤 포레스트 알고리즘 상세 - 모형 적합 및 결과 확인 - 분류모형 성능 평가 - 최적 파라미터 탐색: 그리드 서치 - 최적 모형 성능 평가 	<p>개인대출 이용 분류 모형 + 중고차 가격 추정 회귀 모형</p>
오후 (4H)	그래디언트 부스팅	<p>그래디언트 부스팅은 성능이 낮은 나무모형의 약점을 순차적으로 개선하는 방식으로 수 백, 수 천 개의 나무모형을 결합하는 알고리즘으로, 랜덤 포레스트보다 조금 더 성능이 우수한 모형을 제공한다. 분류모형과 회귀모형을 모두 만들 수 있다.</p>	<ul style="list-style-type: none"> - 그래디언트 부스팅의 개요 - 그래디언트 부스팅 알고리즘 상세 - 모형 적합 및 결과 확인 - 분류모형 성능 평가 - 최적 파라미터 탐색: 그리드 서치 - 최적 모형 성능 평가 	<p>개인대출 이용 분류 모형 + 중고차 가격 추정 회귀 모형</p>

- **학력**
 - 한양대학교 경영학과 MIS 전공 박사(데이터 마이닝 연구) 수료 (2020)
 - 서울대학교 경영학과 마케팅 전공 석사 (2002)
 - 고려대학교 식량자원학과 학사 (2000)
- **경력**
 - 하나금융경영연구소, 데이터 사이언티스트 (2013~2019)
 - 하나카드, 데이터 분석가 (2010~2013)
 - 현대캐피탈, 데이터 분석가 (2009~2010)
 - 삼성화재해상보험, 데이터 분석가 (2002~2009)
- **강의분야**
 - R 프로그래밍 기초, 시각화, 통계 분석, 머신러닝, 웹 크롤링, 텍스트 마이닝 및 업무자동화(RPA)
 - Python 프로그래밍 기초, 크롤링, 기초 수학 등
 - 다양한 머신러닝 알고리즘을 활용한 분류/회귀 모형 개발
 - 데이터 시각화 및 탐색적 데이터 분석
- **강의**
 - SK하이닉스, LG전자, LS그룹, 삼성전기, LG인화원, GS SHOP, SK C&C, SK이노베이션, PoscoICT, ktds, KB 국민카드, GS 칼텍스, 한국언론재단 소속 언론사 등
- **저서**
 - 효율적인 R 프로그래밍 (2018)
 - 포워드 2019 미래를 읽다(2018)