

M1399.000500: 통계적 기계학습

M1399.000500: Statistical Machine Learning

2024년 1학기

교수	김지수, jkim82133@snu.ac.kr	조교	박석훈, shrdid@snu.ac.kr
수업시간	화, 목 09:30 - 10:45	수업장소	25동 210호
면담시간 (교수)	목 11:00 - 13:00	면담장소 (교수)	25동 335호
면담시간 (조교)	금 11:00 - 12:00	면담장소 (조교)	25동 305호
홈페이지	https://jkim82133.github.io/M1399.000500/2024S/		
eTL	https://myetl.snu.ac.kr/courses/256735		

수업 소개

통계적 기계학습(Statistical Machine Learning) 방법은 데이터과학 및 인공지능 분야에서 핵심 방법론으로 사용되고 있다. 본 강의에서는 통계적 기계학습 방법론을 소개하고 기본 이론을 배운다. 주로 지도학습(supervised learning)을 위주로 데이터마이닝에서 널리 쓰이는 회귀(regression) 및 분류(classification) 알고리즘 및 기반 통계적 이론을 배운다. 모형을 평가하는 여러 개념을 배우고, 고급 데이터마이닝 기법인 앙상블 기법(bagging, boosting 등)을 배운다. 커널을 이용한 방법론 및 이론적 기반인 Reproducing Kernel Hilbert Space를 배운다. 또한, 비지도학습(unsupervised learning)의 예로 군집분석(clustering)을 배운다. 심층학습(deep learning) 관련 통계 이론도 간단히 배운다.

- 키워드: 지도학습(supervised learning), 비지도학습(unsupervised learning), 회귀(regression), 분류(classification), 앙상블 기법(ensemble methods), 군집분석(clustering), 심층학습(deep learning)

수업 목표

- 통계적 기계학습(Statistical Machine Learning)을 접하고 이해한다.
- 통계적 기계학습에 어떤 통계적 도구가 쓰이는지 이해한다.
- 상황에 맞게 적절한 데이터마이닝 내지는 기계학습 방법론을 선택할 수 있다.
- 기계학습 알고리즘에 따라 적절한 통계적 분석을 할 수 있다.

선수과목

- 확률(probability)의 기본적 개념에 최소한 학부 수준으로는 익숙해야 하며, 학부에서는 확률의 개념 및 응용(326.211), 측도이론과 확률(M1407.002500), 실변수함수론(881.425), 대학원에서는 확률론 1(326.513), 실해석학(3341.503) 등과 같은 과목에서 다룬다.
- 수리통계(mathematical statistics)의 기본적 개념에 최소한 학부 수준으로는 익숙해야 하며, 수리통계 1(326.311), 수리통계(M1399.000900) 등과 같은 과목에서 다룬다.

- 선형사상, 기저, 차원 등 선형대수(linear algebra)의 기본 개념에 익숙해야 하며, 선형대수학 1(300.203A), 선형대수학(881.007) 등과 같은 과목에서 다룬다.

위의 필수과목들 외에 다음의 개념을 접해보면 좋으나 필수는 아니며, 수업에서 사용하는 개념들은 수업에서 정의한다.

- 회귀분석을 접해보면 좋으며, 회귀분석 및 실습(326.313) 등과 같은 과목에서 다룬다.
- 볼록 최적화(convex optimization)를 접해보면 좋으며, 최적화의 수학적 이론 및 계산(3341.454) 등과 같은 과목에서 다룬다.

교재 및 참고문헌

정해진 교재의 내용을 처음부터 끝까지 따라가진 않지만, 다음 책들을 참조한다.

- Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Second Edition, 2009. <https://hastie.su.domains/ElemStatLearn/>
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning: With Applications in R, 2021. <https://statlearning.com>
- Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, Foundations of Machine Learning, Second Edition, 2018. <https://cs.nyu.edu/~mohri/mlbook/>

그 외에 각 주제에 맞춰서 참조하는 많은 참고문헌들이 있다.

평가 방법

- 최종학점은 다음과 같이 계산한다: 과제 40%, 중간 25%, 기말 25%, 수업 참여 10%

과제

- 과제는 4번 제출할 예정이며 과제는 과제 제출 마감 날짜 23:59까지 eTL로 제출하거나 서면으로 제출해야 한다.
- 과제를 기한 내로 완성하지 못했을 경우 10%의 점수를 감점하고 1주일의 시간을 더 쓸 수 있다. 1주일 후까지도 제출하지 못하면 0점 처리된다.

중간고사와 기말고사

중간고사와 기말고사 일정은 다음과 같다.

- 중간고사: 4/26(금)
- 기말고사: 6/14(금)

프로그래밍 언어

과제 중에 프로그래밍을 이용하는 문제가 나올 수 있다. 프로그래밍 언어는 수강생이 편한 언어를 고를 수 있으며, 수업 시간 중에 R 과 파이썬으로 예시를 들어줄 예정이다.

과제 협력에 관한 원칙

과제를 동료 학생과 같이 협력하는 것을 장려한다. 하지만 답안지는 반드시 본인 스스로 작성해야 한다. 도움을 받았을 때는 도움을 준 학생의 이름을 과제 제출 시 명시해야 한다. 그렇지 않으면 같은 답안을 제출한 과제는 모두 0점 처리한다.

일정표

아래의 일정표는 잠정적으로 작성한 것으로, 최신 일정표는 항상 홈페이지에서 확인한다.

날짜	주제	비고
1주 (3/5, 3/7)	Introduction to statistical learning, Overview of Supervised Learning	
2주 (3/12, 3/14)	Overview of Supervised Learning, Linear classifier	
3주 (3/19, 3/21)	Shrinkage estimator	
4주 (3/26, 3/28)	Basis expansion and Kernel methods	과제1 마감(3/26)
5주 (4/2, 4/4)	Basis expansion and Kernel methods	
6주 (4/9, 4/11)	Ensemble	
7주 (4/16, 4/18)	Ensemble	과제2 마감(4/16)
8주 (4/23)	Model assessment and selection	중간고사 (4/26)
9주 (4/30, 5/2)	Function estimation on high dimensions	
10주 (5/7, 5/9)	Empirical risk minimization	
11주 (5/14, 5/16)	Kernel machines	과제3 마감(5/14)
12주 (5/21, 5/23)	Clustering	
13주 (5/28, 5/30)	Deep learning: Introduction	
14주 (6/4)	Other topics	과제4 마감(6/4)
15주 (6/11)	Other topics	기말고사 (6/14)

장애학생 지원사항

장애유형	강의수강 관련	과제 및 평가 관련
시각장애	교재 제작(디지털교재, 점자교재, 확대교재 등), 대필도우미 허용	과제 제출기한 연장, 과제 제출 및 응답 방식의 조정, 평가 시간 연장, 평가 문항 제시 및 응답 방식의 조정, 별도 고사실 제공
지체장애	교재 제작(디지털교재), 대필도우미 및 수업보조 도우미 허용	
청각장애	대필 및 문자통역 도우미 활동 허용, 강의 녹취 허용	
건강장애	질병 등으로 인한 결석에 대한 출석 인정, 대필도우미 허용	
학습장애	대필도우미 허용	
지적장애	대필도우미 및 수업 멘토 허용	개별화 과제 제출 및 대체 평가 실시
자폐성장애		