

321.621A: 통계이론세미나 - 위상구조의 통계적 추정

321.621A: Seminar in Recent Development of Statistical Theories - Statistics on Topological Structure

2023 가을학기

교수	김지수, jkim82133@snu.ac.kr	조교	신민섭, jpoth1729@snu.ac.kr
수업시간	화, 목 09:30 - 10:45	수업장소	25동 210호
면담시간	화 16:00 - 18:00 또는 약속	면담장소	25동 212호
홈페이지	https://jkim82133.github.io/321.621A/2023F/		
eTL	https://myetl.snu.ac.kr/courses/247782/		

수업 소개

위상수학은 수학의 한 분야로, 국소적인 부분들이 대역적으로 어떻게 연결되어 있는지 연구한다. 자료를 분석할 때 이런 위상학적 특성을 이용하는 것을 위상 자료 분석(Topological Data Analysis, TDA)이라고 한다. TDA를 활용하면 기존의 분석 방법과 다른 정보를 주기도 하고, 분석 모형을 잡음(noise)에 강건(robust)하게 만들어 주기도 한다. 이 수업에서는 이런 TDA를 통계적으로 분석하는 것을 주목표로 한다. 예를 들어, TDA에서 다루는 위상 구조들을 자료로부터 일치성(consistency) 있게 추정할 수 있는지, 불확실성을 신뢰 집합(confidence set)으로 정량화할 수 있는지, 위상 구조 추정의 어려움을 미니맥스(minimax)로 정량화할 수 있는지 등과 같은 통계적 질문을 공부해 본다. 이를 위해, TDA에서 다루는 여러 위상구조들을 알아보고, 자료가 주어졌을 때 이런 위상구조에 어떤 통계적 분석을 할 수 있는지 알아본다.

- 키워드: 위상 자료 분석(Topological Data Analysis), 잡음(noise), 강건성(robustness), 일치성(consistency), 신뢰 집합(confidence set), 미니맥스(minimax)

수업 목표

- 위상 자료 분석(Topological Data Analysis, TDA)을 접하고 이해한다.
- TDA를 통계적으로 분석하는 데에 어떤 위상적 도구와 통계적 도구가 쓰이는지 이해한다.
- 상황에 맞게 적절한 TDA 방법론을 선택할 수 있다.
- 통계 모델에 TDA를 활용했을 때 그에 해당하는 통계적 분석을 할 수 있다.

선수과목

- 선형사상, 기저, 차원 등 선형대수(linear algebra)의 기본 개념에 익숙해야 하며, 선형대수학 1(300.203A), 선형대수학(881.007) 등과 같은 과목에서 다룬다.
- 연속함수(continuous function)의 개념에 익숙해야 하며, 해석개론 및 연습 1(M1407.000600), 해석개론 1(3341.201), 해석개론(881.008), 위상수학개론 1(881.401) 등과 같은 과목에서 다룬다.

- 측도론(measure theory)의 개념에 익숙해야 하며, 확률론 1(326.513), 실변수함수론(881.425), 실해석학(3341.503) 등과 같은 과목에서 다르다.

이 수업은 통계학과 위상수학에 걸쳐 있어, 관련 과목들이 통계학과와 수리과학부에 걸쳐 분산되어 있다. 위의 필수과목을 제외하고는 관련 과목들을 전부 들었다고 가정하지는 않으며, 수업에서 사용하는 개념들은 수업에서 정의한다.

교재 및 참고문헌

정해진 교재의 내용을 처음부터 끝까지 따라가진 않지만, 다음 두 서베이 논문을 주로 참조한다.

- Larry Wasserman, Topological Data Analysis, 2018. <https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-031017-100045/>
- Frédéric Chazal, Bertrand Michel, An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists, 2021. <https://www.frontiersin.org/articles/10.3389/frai.2021.667963/>

또한, 수업 전반적으로 다음 문헌들을 참조할 수 있다.

- Herbert Edelsbrunner, John Harer, Computational Topology: An introduction, 2010.
- Allen Hatcher, Algebraic topology, 2002. <https://pi.math.cornell.edu/~hatcher/AT/ATpage.html>
- Gunnar Carlsson, Mikael Vejdemo-Johansson, Topological Data Analysis with Applications, 2022.

그 외에 각 주제에 맞춰서 참조하는 많은 참고문헌들이 있다.

평가 방법

- 최종학점은 다음과 같이 계산한다: 과제 40%, 프로젝트 40%, 수업 참여 20%

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.
- 과제 1회당 배점은 최소 100점 이상이다.
- 최종학점에 과제 점수는 4번 제출한 점수의 합산과 100점 중 낮은 걸로 반영된다. 예를 들어, 과제 점수가 90, 40, 80, 0일 경우, 반영되는 점수는 $\min\{90 + 40 + 80 + 0, 100\} = 100$ 점이다.

프로젝트

이 수업의 프로젝트에서는 수업과 관련 있으면서 관심 있는 주제를 골라, 팀을 짜서 관련 논문을 읽고, 중요한 이론적 결과들을 정리해 최대 8쪽 분량의 보고서로 작성하고 5분 길이로 발표한다. 논문 1개만 다뤄도 되고 논문의 일부만 다뤄도 된다. 이 프로젝트는 새로운 연구를 수행하는 게 아닌 기존 연구를 이해하는 것을 목표로 한다.

보고서에는 배경지식, 중요한 이론적 결과들, 증명의 간단한 열개가 포함되어야 한다. 3단계에 걸쳐서 제안서, 중간 보고서, 최종 보고서를 제출하고 10분 이내의 길이로 발표한다.

- 제안서: 프로젝트 제안서에는 (1) 프로젝트 제목, (2) 팀 구성원, (3) 프로젝트에서 다루는 문제의 정확한 명세, (4) 프로젝트가 다루는 범위, (5) 읽을 참고문헌 등이 포함되어야 한다.
- 중간 보고서: 중간 보고서에는 (1) 서론, (2) 지금까지 한 부분, (3) 앞으로 해야 할 부분, (4) 팀원 간 분배가 포함되어야 한다.
- 최종 보고서: 최종 보고서는 NeurIPS 형식으로 작성해야 하며 최대 8쪽까지 허용한다. 다음과 같은 내용이 구조로 드러나게 작성해야 한다.
 1. 서론: 문제의 동기와 해당 분야의 요약
 2. 표기법과 가정

3. 중요 결과들
4. 증명 열개
5. 결론: 결과에 대한 주석과 미해결 문제들

- 발표는 10분 이내의 길이로 하며, 문제의 동기와 중요 결과들을 중심으로 발표한다.
- 점수 평가에는 최종 보고서와 발표만 들어간다. 즉, 제안서와 중간 보고서는 제출할 경우 첨삭을 받을 수는 있지만 점수에 반영되지는 않으며, 제출하지 않더라도 점수에 불이익은 없다.

과제와 프로젝트

- 기본적으로는 과제와 프로젝트를 둘 다 한다.
- 프로젝트를 선호할 경우, 프로젝트를 두 개 하는 것으로 과제를 대체할 수 있다.
- 프로젝트가 부담스러운 경우, 교수와의 면담을 통해 프로젝트를 대신해 과제를 추가로 하는 것으로 프로젝트를 대체할 수 있다. 이 경우, 최종학점에 과제와 프로젝트 점수는 4번 제출한 점수의 합산과 200점 중 낮은 걸로 반영된다. 예를 들어, 과제 점수가 90, 40, 80, 0일 경우, 반영되는 점수는 $\min\{90 + 40 + 80 + 0, 200\} = 200$ 점이다.

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 (9/5, 9/7)	Introduction, Review on Statistics	
2주 (9/12, 9/14)	Review on Geometry and Topology	
3주 (9/19, 9/21)	Density Clustering	
4주 (9/26)	Metric spaces, Covers, and Simplicial Complexes	과제1 마감(9/26)
5주 (10/5)	Mapper	
6주 (10/10, 10/12)	Mapper and Reach	제안서 마감(10/12)
7주 (10/17, 10/19)	Geometric Reconstruction and Estimating Intrinsic Dimension	과제2 마감(10/19)
8주 (10/24, 10/26)	Homology Inference	
9주 (10/31, 11/2)	Persistent Homology	
10주 (11/7, 11/9)	Stability of Persistent Homology	과제3 마감(11/9)
11주 (11/14, 11/16)	Consistency of Persistent Homology	중간 보고서 마감(11/16)
12주 (11/21, 11/23)	Confidence Set of Persistent Homology	
13주 (11/28, 11/30)	Statistics on Persistence Landscape	과제4 마감(11/30)
14주 (12/5, 12/7)	Other topics	
15주 (12/12, 12/14)	Project Presentations	최종 보고서 마감(12/14)

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		