

위상학적 자료 분석(Topological Data Analysis)의
통계적 추정 및 기계학습의 응용



Speaker : Dr. Jisu Kim

Affiliation : Inria Saclay, France

* 본 강의는 한국어로 진행됩니다.

* This talk will be given in Korean.

June 2, 2022 16:00 - 17:15

Zoom <https://unist-kr.zoom.us/j/87540455917>



Host

Prof. Pilwon Kim Email : pwkim@unist.ac.kr

Abstract

위상학적 자료 분석(Topological Data Analysis)은 포괄적으로 자료에서 위상학적인 특성을 추출하는 분석 방법을 통칭한다. 대표적으로는 군집 나무(cluster tree)와 persistent homology가 있는데, 군집 나무는 서로 비슷한 자료끼리 묶어서 군집을 만들고, persistent homology는 자료를 여러 해상에서 관측하고 지속성 있게 나타나는 위상 특성을 추려낸다. 이와 같은 위상학적 자료 분석은 유용한 정보를 제공하는데, 자료의 과학적인 정보를 전달해 주기도 하고, 자료의 추가적인 특성으로 학습 문제에 활용할 수도 있다. 이때, 자료로부터 위상학적 자료 분석을 수행하면 자료 분포의 임의성에 의해 오차가 생기는데, 이를 통계적으로 정량화할 수 있다. 이 발표에서 위상학적 자료 분석을 어떻게 통계적으로 추정하고 기계학습에 응용하는지 소개한다. 첫째, 자료의 분포 함수로부터 계산한 군집 나무(cluster tree)의 통계적 추정을 알아본다. 둘째, 자료의 분포 함수의 persistent homology의 통계적인 추정을 알아본다. 셋째, persistent homology에서 어떻게 특성(feature)을 만들어 기계학습(machine learning)에 활용할 수 있는지 소개한다. 넷째, persistent homology를 포함해서 위상학적 자료 분석(topological data analysis)을 계산할 수 있게 해주는 프로그램인 R package TDA를 간단히 소개한다.