University of California, Irvine
Statistics Department
Seminar

A Statistical Pipeline for Feature Selection
and Association Mapping with 3D Shapes

Lorin Crawford
Assistant Professor of Biostatistics
Brown University

Thursday, May 2, 2019
4 p.m., 6011 Bren Hall
(Bldg. #314 on campus map)

It has been a longstanding challenge to implement an analogue of variable selection with 3D shapes as the covariates in a regression model. Here, we present SINATRA, a statistical pipeline for sub-image selection where the goal is to identify the physical features of 3D shapes that best explain the variation between two trait classes. A key insight is that, one can use tools from differential topology to transform objects represented as meshes into a collection of vectors (with little to no loss of information about their natural structure). Since this topological transformation is invertible, SINATRA uses an interpretable machine learning algorithm to map significant vector indices back onto the original shapes — thus, highlighting the physical 3D shape characteristics that best distinguish members in the two groups. As an application of our pipeline, we conduct feature selection on a dataset consisting of mandibular molars from five different genera of New World Monkeys and examine the physical properties of their teeth that best characterize their dietary regimens.