

MASTER'S THESIS PRESENTATION

Selective Inference for Clustering with Variable Selection and Unknown Variance

WHEN

April 27, 2022

11:00AM

WHERE

Jones Laboratory, Room 303

**Young-Joo Yun, MS candidate**

Selective inference allows for performing valid inference when parts of the procedure, such as the choice of hypothesis, depend on the data set to be used for testing. An example of such a setting would be testing whether two samples belong to the same true cluster, where the cluster assignments of the samples have been produced by a clustering algorithm. Gao et al. (2020) provide a framework for doing selective inference for this setting, in the case where the data are assumed to be generated from a Gaussian distribution with a known isotropic covariance matrix. In our work, we provide two extensions of the work by Gao et al. (2020): we consider the case where a variable selection method is applied before running a clustering algorithm and the case where the isotropic covariance matrix is unknown. We provide the expressions for the p-values for both settings and present relevant simulation results.

stat.uchicago.eduTHE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS