STATISTICS SEMINAR

Huizi Wang, Ph.D. Candidate

COMPARISON OF CLASSIFICATION METHODS FOR PATHOGEN DETECTION WITH HIGH-DIMENSIONAL MI-FI DATA

Friday, April 12, 2024
2:30 – 3:30 pm
310 MSCS
ZOOM Link for Applied M.S. Students
https://okstate-edu.zoom.us/j/92298224591

Abstract: Microbe Finder (Mi-Fi) is a diagnostic tool developed by researchers at Oklahoma State University which uses high-throughput sequencing technology to measure the abundance of a pathogen in a sample. The basic goal is to determine if a sample should be classified as pathogenic or healthy using Mi-Fi data. The challenge is that pathogen abundance is measured as the number of "e-probe hits" for at least 10 and up to 10,000 "e-probes", depending on the pathogen of interest, and there are a relatively few number of samples to train a classifier. Hence this is a high-dimension (HD) classification problem with count data. This talk first reviews the simple classification approaches for a variable called "total score" and shows an assessment of the simple classification approaches. Second, this talk reviews the two main approaches for HD classification problem. In particular, high-dimensional discriminant analysis (DA) methods, such as Poisson linear discriminant analysis, and high-dimensional multiple logistic regression (MLR) approaches are reviewed. It is shown analytically that an MLR model can incorporate common DA model features, like over-dispersion and zero-inflation. However, in practice a wide variety of parameter estimators are available, depending on the statistical model and type of shrinkage penalty. This talk concludes with a comparison of these methods as applied to some real Mi-Fi data. The focus is on methods that have small classification error and allow for model / "e-probe" selection.

Refreshments immediately following seminar in Room 309 MSCS.