Mathematics, Statistics, and Computer Science **@ UIC**

Statistics and Data Science Seminar

Penalized Linear Discriminant Analysis for Family Studies

Yixin Fang (NYU Langone Medical Center)

Abstract: In family studies with multiple continuous phenotypes, we are interested in finding linear combinations of the phenotypes with large heritabilities, which can be considered as new phenotypes for genetic analysis. The problem can be recast as linear discriminant analysis (LDA). When the number of phenotypes is large, LDA is not appropriate for two reasons: the standard estimate for the within-family covariance matrix is singular, and it is difficult to interpret the newly defined phenotypes. Here we propose a novel version of penalized LDA, with an L_1^2 penalty in the denominator of the Rayleigh quotient. Besides overcoming the above two problems, the proposed method has at least three advantages compared with the existing regularization methods. First, it solves the singularity problem and achieves the sparsity property simultaneously. Second, the method is scale-invariant. Third, the consistency can be proved. We evaluate the performances of the method using simulations and two family studies.

Wednesday, February 22 at 4:00 PM in SEO 636