

When: Friday 15:00 – 16:00
Where: HRBB 204
Coordinator: Xiaoning Qian (xqian@ece.tamu.edu)

Speaker: Prof. Irina Gaynanova

Assistant Professor
Department of Statistics
Texas A&M University



Title: High-dimensional classification via sparse linear projections with application to gene expression data

Date: 10-14-2016

Abstract: Traditional classification methods, such as discriminant analysis and logistic regression, often face difficulties when applied to modern high-dimensional datasets due to the presence of spurious correlations and over-selection of relevant measurements. In the literature, these problems are typically addressed separately, leading to suboptimal misclassification rates and computationally challenging methodology. In this talk I will discuss how to address these problems jointly via sparse linear projections. The use of projections is motivated by Fisher's discriminant analysis, where the goal is to maximize the distances between the classes in the projected space. These ideas can be extended to the high-dimensional data using the convex optimization framework, thus leading to computationally efficient classification procedure. The proposed method achieves optimal variable selection and classification rates when the covariance matrices between the classes are equal, and has competitive performance when this assumption is violated. I will conclude by using sparse linear projections to classify patients with prostate cancer and Crohn's disease.

Biography: Gaynanova received her Diploma with Honors in Applied Mathematics and Computer Science from Lomonosov Moscow State University and her Ph.D. degree from Cornell University in Statistics. She is currently an Assistant Professor in the Department of Statistics at Texas A&M University. Her research interests include analysis of high-dimensional data, multivariate analysis, statistical methods for analyzing biological data and machine learning. She has worked on a variety of applied problems such as classification of leukemia patients based on DNA methylation profiles, control of false discovery rates in sample size calculations and study of antibiotic molecular actions based on the metabolic profiles.