

Perspectives on Statistical Machine Learning with Applications

Feryaal F. Ahmed & S. Ejaz Ahmed

Abstract: There are hosts of buzzwords in today's data-centric world, and especially in digital and print media. We encounter data in every walks of life, and for analytically and objectively-minded people, data is everything. However, making sense of the data and extracting meaningful information from it may not be an easy task. We come across buzzwords such as big data, high dimensional data, data streaming, data science, and open data without a proper definition of such words. The rapid growth in the size and scope of data sets in a host of disciplines has created a need for innovative statistical strategies analyzing such data. For example, many private and public agencies are using sophisticated data mining strategies and/or big data analytics to reveal patterns based on collected information. Some examples of big data that have prompted demand are digital marketing, customer service standards, gene expression arrays, social network modeling, clinical, genetics and phenotypic data.

Open science data is particularly lucrative in medicine since clinical data is complex and cumbersome to collect often delaying the progression of medical research. Accessible databases have become a saving grace for many researchers. In genomics, free public archives such as ClinVar, provide reports of the relationships among human genetic variants and phenotypic expressions. However, the level of confidence in the accuracy of claims of clinical significance are reliant on the supporting evidence and the origin of the reports submissions. These variants can be classified as benign, likely benign, indeterminate, likely pathogenic, and pathogenic. From laboratory to laboratory, variant misclassification causes uncertainty in the interpretation of variant impact on disease.

The purpose of this talk is bi-fold. First, we focus on We focus on estimation of model parameters and prediction based on high dimensional data (HDD). In classical regression context, we define HDD where number of predictors (p) are larger than the sample size (n). In the 2nd part of the talk we investigates which ClinVar human genetic variants' features or combinations thereof will help researchers predict classification conflicts with the most suitable model. We present a case study which uses statistical and machine learning techniques to investigate which ClinVar human genetic variants' features or combinations thereof will help researchers predict classification conflicts with the most suitable model ranging from logistic regression to Gradient Boosting Machine.

Bio:

Feryaal Ahmed is a second-year PhD student in Management Science at Ivey Business School, Western University, Canada. Her research focus is in revenue management and big data optimization methods. She has a Master of Science in Management from Ivey Business School and an Honours Bachelors of Science in Biophysics from Brock University, Canada.

Dr. S. Ejaz Ahmed is Professor of Statistics and Dean of the Faculty of Math and Science at Brock University, Canada. Previously, he was Professor and Head of the Mathematics and Statistics Department at the University of Windsor, Canada and University of Regina, Canada as well as Assistant Professor at the University of the Western Ontario, Canada. He holds adjunct professorship positions at many Canadian and International universities. He has supervised numerous Ph.D. and Master Students and organized several international workshops and conferences around the globe. He is a Fellow of the American Statistical Association. His areas of expertise include big data analysis, statistical learning, and shrinkage estimation strategy. Having authored several books, he edited and co-edited several volumes and special issues of scientific journals. He is Technometrics Review Editor for past ten years. Further, he is Editor and associate editor of many statistical journals. Overall, he published more than 175 articles in scientific journals and reviewed more than 100 books. Having been among the Board of Directors of the Statistical Society of Canada, he was also Chairman of its Education Committee. Moreover, he was Vice President of Communications for The International Society for Business and Industrial Statistics (ISBIS) as well as a member of the "Discovery Grants Evaluation Group" and the "Grant Selection Committee" of the Natural Sciences and Engineering Research Council of Canada.

For more information see: <https://brocku.ca/mathematics-science/mathematics/directory/syed-ejaz-ahmed/>