

ORAL COMMUNICATION (MATHEMATICS AND MEDICINE)

Multiple penalized regression models for stable selection in high dimensional dataANA HELENA TAVARES^a, VERA AFREIXO^b, GABRIELA MOURA^c^a Institution: Center for Research & Development in Mathematics and Applications (CIDMA), Agueda School of Technology and Management, University of Aveiro, Portugal

E-mail: ahtavares@ua.pt

^b Institution: Center for Research & Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, Portugal

E-mail: vera@ua.pt

^c Institution: Genome Medicine Lab, Department of Medical Sciences, iBiMED - Institute of Biomedicine, University of Aveiro, Portugal

E-mail: gmoura@ua.pt

Abstract

The main goal of this talk is to discuss a stable procedure to perform variable selection in high dimensional data, as in genome-wide association studies. We propose a variable selection procedure that combines the result of multiple penalized regression models, and define a weighted formulation to identify the most important predictor variables. Variables that are selected more often in higher quality models have greater potential to be selected for the final model. The procedure is applied as a proof of concept to the Alzheimer's Disease Neuroimaging Initiative (ADNI) public dataset in order to identify genetic variants (SNPs) that are associated with Alzheimer's Disease.

References

- [1] P. Waldmann, G.M.B. Gredler, C. Fürst, and J. Sölkner: *Evaluation of the lasso and the elastic net in genome-wide association studies*. *Frontiers in Genetics*, **4** (2013).
- [2] V. Afreixo, A. H. Tavares, V. Enes, M. Pinheiro, L. Rodrigues, and G. Moura: *Stable Variable Selection Method with Shrinkage Regression Applied to the Selection of Genetic Variants Associated with Alzheimer's Disease*. *Applied Sciences*, **14** (6): 2572 (2024).