

Centro de Investigação em Matemática e Aplicações
Departamento de Matemática
Programa de Doutoramento em Matemática

Seminário (online)

11 de novembro de 2020

Feature selection for marine species origin
prediction

Regina Bispo

(Professora Auxiliar, r.bispo@fct.unl.pt)

Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade
Nova de Lisboa

Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa

Abstract The framework of general linear models (GLM) is typically used to model a response variable as a function of a set of p associated covariates, based on a random sample with n elements, whenever the parent population is described by a distribution within the exponential family. Fisher information matrix (FIM), arises in this context as the inverse of FIM yields the Cramer-Rao Lower Bound for the covariance matrix of the estimated parameters. However, when dealing with high dimensional data ($p > n$), the specification matrix is not a full-rank matrix, FIM is not invertible and, consequently, the estimation process fails to converge. In this talk, I'll discuss some methods that allow to overcome this problem, by selecting explanatory variables (features) and, hence, enable model fitting despite the original data high-dimensionality. An applied example to a marine biology dataset will be discussed.

Keywords: Feature selection, Fisher information matrix, general linear models, high-dimensional data.

Acknowledgements

This talk has been partially supported by Centro de Investigação em Matemática e Aplicações (CIMA), through the Project UID/04674/2020 of FCT-Fundação para a Ciência e a Tecnologia, Portugal.

RB work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

