DA: Ecological and Evolutionary Inference Using Supervised Discriminant Analysis

Xinghu Qin[1,1], Mary Wu[2], T. Ryan Lock[3], and Robert Kallenbach[4]

[1]University of St Andrews
[2]AI Genomics
[3]University of Missouri
[4]University of Missouri Division of Plant Sciences

June 19, 2020

Abstract

With the rapid and large production of biological data (phenotypic traits, genomes, and simulated DNA), traditional statistic-based approaches may not meet the demands of ecological or evolutionary inferences. To mitigate this issue, we propose supervised visual and statistical machine learning approaches to do biological, evolutionary, and demographic inference. We introduce five supervised learning approaches (DAPC, DAKPC, LFDA, LFDAKPC, KLFDA) into ecology and evolution within the same discriminant analysis family, but with different linear and non-linear properties. We tested their performance and expected to find the optimal method for biological, evolutionary, and demographic inference. Applicable examples of such methods include species classification, population structure identification, and demography inference. We applied these five supervised learning techniques to simulated spatially-structured demographic scenarios along with realistic ecological and genetic data to elucidate their power and practicability in pattern inference. LFDA shows the highest discriminatory power in demographic inference. However, KLFDA outperforms other methods in population structure identification. DAPC and DAKPC differentiated species traits well when applied to real datasets. These approaches assess the structure of the data without model assumptions and show the potential to identify complex demographic histories and subtle population structure. We have made the DA package available at https://github.com/xinghuq/DA. We recommend users choose these machine learning approaches appropriately depending on their scientific questions and target data.

Hosted file

Evolutionary and biological inference_MER_update_12_06_2020.pdf available at https://authorea.com/users/304610/articles/460985-da-ecological-and-evolutionary-inference-using-supervised-discriminant-analysis