



Statistical Species Identification

Måns Karlsson and Ola Hössjer

Juni 2019

Abstract

Identification of taxa can be significantly assisted by statistical classification in two major ways. First, one may use a statistical model to determine taxon of subjects based on various characteristics or traits. Secondly, when faced with a collection of subjects with common traits measured, one may determine combinations of traits that signify each taxon in question. To this end, we present a general Bayesian approach to classification of observations based on traits, whose measurements follow some (latent) multivariate Gaussian distribution, but might be truncated or even missing and allow for the traits to depend on covariates. It is inspired by liability threshold modelling and Bayesian Quadratic Discriminant Analysis. The approach is paired with two decision rules: one for which classification is forced, and one that allows for uncertainty of classification, including all categories whose posterior probability ratio, compared to the most likely taxon, exceeds a given threshold. Both of these decision rules are evaluated using blockwise Gibbs sampling. Then we show how the reward function corresponding to these two rules can be used for model selection in terms of blockwise cross validation. Finally, we exemplify our approach on a data set over four morphologically similar *Acrocephalus*-genus warblers.