

Using boosted decision trees for star-galaxy separation

Penélope Etayo-Sotos^{1,2} and Ignacio Sevilla-Noarbe²

¹ UCM, Universidad Complutense de Madrid

² CIEMAT, Departamento de Investigación Básica, Avda. Complutense 40, 28040, Madrid, Spain.

Abstract

We present an application of a particular machine-learning method (Boosted Decision Trees, BDT) to separate stars and galaxies from their catalog characteristics. This application is based on the BDT implementation in the Toolkit for Multivariate Analysis (TMVA) for ROOT, a physics analysis package widely used in high energy physics. The main goal is to improve from simple thresholding cuts on standard separation variables that may be affected by local effects such as blending, badly calculated background levels or which do not include information in other bands. We explain the basics of decision trees and the training sets used for the cases that we analyze. The improvements are shown using the Sloan Digital Sky Survey Data Release 7. With this method we have reached an efficiency of 99% with a contamination level of less than 0.45%.