

*Gerardo Sanz (University of Zaragoza)*

*Luis M. Esteban (University of Zaragoza)*

## Binary Classification Models in Oncology

Topic 2 – Learning more from what we already know

Keywords: Binary classification, step-by-step algorithm, prostate cancer, classification accuracy, survival

### Introduction

The binary classification problems are among the most common in oncology. To send a patient to surgery or another alternative treatment is not an easy decision and the predictive models can help the physician. Combining data of several markers for classifying of patients is a major issue in Oncology. To tackle this problem some approaches have been proposed in the literature but many of them are parametric, which is a severe restriction in real situations. That is why distribution-free approaches become more appealing, but the computing limitations had restricted their use to models with only two variables. We have analyzed a step-by-step algorithm when more than two variables are available. The algorithm is implemented in R and its performance has been analyzed through extensive simulations with satisfactory results. The algorithm was also applied to a prostate cancer database. Once the predictive model is built and validated, the next step is to define a threshold above which one individual is classified as class 1 and below as class 2. To establish a good threshold is a key issue, but unfortunately is an unsolved problem. We have introduced a proposal based on the estimate, through kernel functions, of the density functions of the populations. This procedure has advantages on the classical methods based on classification risk groups, such as the National Cancer Comprehensive Network – NCCN. The building of models has also been considered in survival analysis.