

Heart Disease Machine Learning 1

Jose Pérez Cano and Álvaro Ribot Barrado

GCED – FIB – UPC

March 23, 2020

1. Problem

The database we've chosen to work with is about the presence or absence of heart disease in a sample of the population. The dataset consists of 75 attributes, one of them is the presence of heart disease in a scale from 0 to 4. It also contains missing values and dummy variables created to replace some sensible data,

2. Reason

We wanted to work with a medical dataset because it is the type of dataset we feel is more practical. At first we have a pool of several problems from which this one got more variables to work with although less patients. This way we got to practice not only the machine learning algorithm but the problem of having too many dimensions and very few patients, which seems more real for us.

3. References

The creators and all the information about the dataset can be found in this link. It has a long list of papers which cite the dataset.