Echocardiographic diagnosis of rheumatic disease

Proposal of the statistical methodology

Eduardo Elias Ribeiro Junior Henrique Aparecido Laureano
Maria do Carmo Pereira Nunes

Faculdade de Medicina
Universide Federal de Minas Gerais - UFMG

13 de setembro de 2016

Roteiro

1. Introduction and goals

2. Classification Models

3. Evaluate and compare the classifiers

Introduction and goals

Problem

- Today we don't have a well-defined score for predicting the progression of rheumatic disease

Problem

- Today we don't have a well-defined score for predicting the progression of rheumatic disease
- Based on empirical characteristics, patients are classified into three categories:
  - Definitive
  - Borderline
  - Normal

Problem

- Today we don't have a well-defined score for predicting the progression of rheumatic disease
- Based on empirical characteristics, patients are classified into three categories:
  - Definitive
  - Borderline
  - Normal

Goal

Reduce the number of patients classified as borderline

Classifiers

2012 World Heart Federation criteria

Advantages:

- Simple;
- Standard echocardiographic diagnosis.

Disadvantages:

- Subjective criteria;
- In practice, unused some indications.

Data-based classifiers

Advantages:

- Not subject to define the variables;
- Informe the variables importance;
- Classify based in "scores".

Disadvantages:

- Difficulty in understanding by non-statisticians;
- Equalized database.

Classification Models

Outcomes

The response variable in database should be defined as:

- The extremes classification (Normal and Defined) using 2012 World Heart Federation criteria; and/or
- The patient status about progression of the rheumatic heart disease, but long time follow up is necessary to observe the progression.

Classifiers

Simple and easy interpretation:

- Logistic Regression
- Discriminant Analysis Based (Linear, Quadratic, etc.)
- K-Nearest Neighbor (KNN)
- Decisions Trees

More complex and hard or not interpretation, but effective:

- Bagging Decision Trees
- Random Forests
- Support Vector Machines

Evaluate and compare the classifiers

Data test

Classifiers:

- 2012 World Heart Federation criteria
- Modified 2012 World Heart Federation criteria
- Data-based classifiers

Steps:

1. Fit the classifiers using database ???, with 2012 World Heart Federation criteria (Normal and Definitive classification).
2. Using the fit classifiers, classify individuals in the diseased database.

Compare measures

- Predict error (number of incorrect prediction)
- ROC Curve (Receiver Operating Characteristic)
  - Confusion matrix (sensitivity and specificity)
- AUC Area under curve