

Métodos Quantitativos

Prof. Dr. A. L. Korzenowski

Aula 08: Support Vector Machines

Today we are going to discuss the support vector machine (SVM), an approach for classification that was developed in the computer science community in the 1990s and that has grown in popularity since then. The support vector machine is a generalization of a simple and intuitive classifier called the *maximal margin classifier*. This classifier unfortunately cannot be applied to most data sets, since it requires that the classes be separable by a linear boundary. Support vector machine is a further extension of the support vector classifier in order to accommodate non-linear class boundaries. Support vector machines are intended for the binary classification setting in which there are two classes. In this class we will also discuss extensions of support vector machines to the case of more than two classes.

Relationship to Logistic Regression

When SVMs were first introduced in the mid-1990s, they made quite a splash in the statistical and machine learning communities. This was due in part to their good performance, good marketing, and also to the fact that the underlying approach seemed both novel and mysterious. The idea of finding a hyperplane that separates the data as well as possible, while allowing some violations to this separation, seemed distinctly different from classical approaches for classification, such as logistic regression and linear discriminant analysis. Moreover, the idea of using a kernel to expand the feature space in order to accommodate non-linear class boundaries appeared to be a unique and valuable characteristic.

Considere o modelo de regressão logística $f(x) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ como

$$\underset{\beta_0, \beta_1, \dots, \beta_p}{\text{minimize}} \left\{ \sum_{i=1}^n \max[0, 1 - y_i f(x - 1)] + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

onde λ é um parâmetro de sintonia não negativo. Quando λ é grande então $\beta_0, \beta_1, \dots, \beta_p$ são pequenos, mais violações nas margens das áreas de classificação serão toleradas e uma pequena variância mas com alto viés de classificação irá resultar. Quando λ é pequeno então poucas violações nas margens irão ocorrer; isto aumenta a variância mas reduz o viés do classificador. O termo $\lambda \sum_{j=1}^p \beta_j^2$ é uma penalidade e atua no papel de controlar o *trade-off* viés-variância no *support vector classifier*.

A equação acima possui uma forma do tipo “*Loss + Penalty*” que pode ser escrita como:

$$\underset{\beta_0, \beta_1, \dots, \beta_p}{\text{minimize}} \{L(\mathbf{X}, \mathbf{y}, \beta) + \lambda P(\beta)\}$$

Aplicação no caso de classificação binária

nononono

```
#Limpa o Workspace  
rm(list=ls())  
  
#Habilita o pacote car  
if (!require(car)) install.packages('car')  
require(car)  
attach(mtcars)
```

““