Master BeNeFri in Computer Science

Course: Statistical Learning Methods
Spring 2016

Exercise #11. Classification: Support Vector Machines

1. Download from folder Exercise#9 on ILIAS website the dataset `Vertebral` dataset (filename: VertebralData.2C.txt) containing various biomedical variables that can be used to predict the orthopedic class of the patient (variable `Status`) and read file VertebralDescription.pdf. You have 310 observations (patients) with six predictors. We have no missing value.

Compare the predictions you obtain with different variants of the SVM package with the dataset VertebralData.2C.txt for variable `Status` (`Normal / Abnormal`). Use a fair methodology to compare the classifiers (and explain your choice). The medical organization want a (fair) error rate estimation at max 5%.

Can you estimate the error rate for these strategies? Which parameter values are the best? Can you satisfy the medical criterion?

2. Download from folder Exercise#8 on ILIAS website the dataset `Cancer dataset` (filename: Cancer.txt) and read the description CancerDescription.pdf.

Compare the predictions you obtain with different variants of the SVM package for the category `diagnosis` (`malignant / benign`). Use a fair methodology to compare the classifiers (and explain your choice).

Can you estimate the error rate for these strategies? Which parameter values are the best?