

Aprendizagem Automática

Assignment 1 - Classifiers

Andrea Mascaretti N52222

Daniel Pimenta N45404

October 20, 2017

Abstract

The main goal of this assignment was the parameterization, fitting and comparison of Logistic Regression, K-nearest Neighbours and Naive Bayes classifiers.

The data set used was the banknote authentication which was obtained from the UCI machine learning repository.

To achieve our goals we used the Spyder IDE with programming language Python 3.6 and some of its modules such as *Panda* to load and manage our data, *NumPy* which is the fundamental package for scientific computing with Python and was used for various mathematical calculations and managing multidimensional arrays, *sklearn* (*scikit-learn*) which has efficient tools for data mining and data analysis used in our assignment to preprocess data (shuffle, standardize, split), fold our data into k stratified folds, calculate cross validation score and fit our data on Logistic Regression and K-nearest Neighbours classifiers. Another imported module was Matplotlib.Pyplot which provides a MATLAB-like plotting framework used to build the error plots.

1 Introduction

WHAT WE INITIALLY HAD,
WHAT WAS ASKED TO DO,
HOW WE PLAN ON DOING IT

2 Classifiers

Explain what are the three parameters that were optimized and their effects on their respective classifiers. This will also require a brief explanation of how each classifier works.

Explain the method by which the optimal values were found, noting the differences between the errors and the importance of leaving out a test set for the final evaluation

The report should show the error plots but no other plots are necessary

Logistic Regression

K-nearest Neighbours

Naive Bayes

3 Comparison

McNemar's test TALK ABOUT WHAT IS MCNEMAR TEST AND HOW IT WAS IMPLEMENTED AND USED

Estimate the true error for each classifier, compare the classifiers with McNemar's test and discuss which, if any, is better for this application

4 Conclusion

WHAT WAS ACHIEVED AND WHAT WE LEARNT

5 Bibliography

- ...