Machine learning
Report 4

Sylwester Piątek

24 January 2019

Contents

1 Introduction

In this report I will present the performance of SVM model using MNIST data. The dataset contains pictures with hand-written numbers (0-9). This method is considered to be a good method for pictures classification, like for example face recognition. The models will be tested on two datasets - one of them has pictures of size 8×8 and the second one 28×28.

2 Small pictures

Test set contains approximately 25% of whole 1797 records. The model having parameters $gamma = 0,001$ and $C = 100$ has accuracy 0.9689.

3 Large pictures

The dataset with larger pictures contains 10,000 records. I divided it into training set (8,000 records) and test set. The model having default values of γ and C had 10.85% accuracy which is close to a result of a random classification. I checked performance of the SVM model on this dataset with 3 different γ and 3 different values of C parameter. Amid these values the best accuracy was in the models with $\gamma = 4 \cdot 10^{-8}$. The C parameter did not have a huge impact on the results, but the models with values 100 and 1000 gave better results than the ones with $C = 10$.

4 Summary

The default value for a model built using $svm.SVC$ function is not a good choice in case of this dataset. In such case different values should be used to build a good model. The value of γ shouild be very small to give good results.

$\gamma \backslash C$	10	100	1000
10^{-4}	0.1085	0.1085	0.1085
$4 \cdot 10^{-8}$	0.9505	0.9515	0.9515
10^{-9}	0.9045	0.9295	0.9285

Table 1: Accuracy of the SVM model with various γ and C