Pattern Classification Using SVM

Mikhail Podvalkov

Due to April 12, 2022

1. Introduction

The data for learning was a CIFAR-10 which consists of color images divided into three parts training, validation, and testing. The computer assignments aim to design support vector machines. The computer assignments created an SVM in MATLAB code [1] with different functions. Moreover, the cross-validation was defined and found kernel parameters and optimum C. The best model was trained on the test dataset compared to CNN from previous assignments. The computer assignments provided the benchmark of all machines.

2. Theory

The SVM is the machine learning method whose goal is to find the optimal hyperplane to maximize the correct classification of two class data. [2] (fig 1).

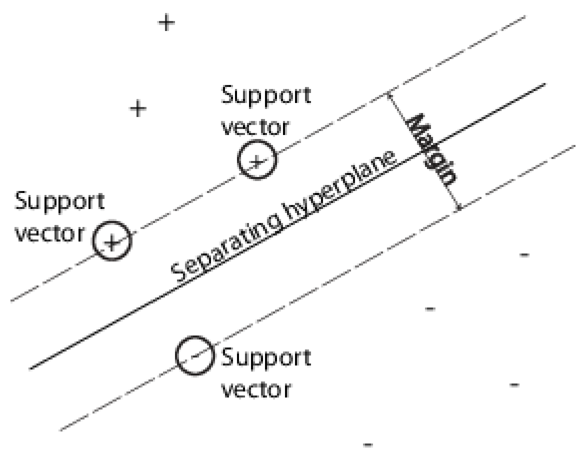

Figure 1 The CNN model,

The data for training, is a set of points (vectors) xj along with their categories yj, where $yj = \pm 1$. The hyperplane equation is

$f(x)=x'\beta+b=0$,

where $\beta \in Rd$ and b are real numbers.

The following problem defines the *best* separating hyperplane (i.e., the decision boundary). Find β and b that minimize $||\beta||$ such that for all data points (xj,yj),

$yjf(xj)\geq 1$.

The support vectors are the xj on the boundary, those for which yjf(xj)=1.

For mathematical convenience, the problem is usually given as the equivalent problem of minimizing ‖β‖.. The optimal solution (ˆβ,ˆb) enables the classification of a vector z as follows:

$class(z)=sign(z'\hat{\beta}+\hat{b})=sign(\hat{f}(z))$.
$\hat{f}(z)$ is the *classification score* and represents the distance z is from the decision boundary.

Some binary classification problems do not have a simple hyperplane as a useful separating criterion. There is a variant of the mathematical approach that retains nearly all the simplicity of an SVM separating hyperplane for those problems.

This approach uses these results from the theory of reproducing kernels:

- There is a class of functions G(x1,x2) with the following property. There is a linear space S and a function φ mapping x to S such that

$G(x_1,x_2) = <\varphi(x_1),\varphi(x_2)>$.

The dot product takes place in space S.

- This class of functions includes:
  - Polynomials: For some positive integer p,

$G(x_1,x_2) = (1 + x_1'x_2)^p$.

  - Radial basis function (Gaussian):

$G(x_1,x_2) = \exp(-\|x_1-x_2)\|^2)$.

3. Results and discussion

The Matlab code provides two types of SVM with polynomial and RBF kernels architectures. On this type was created a one-versus-rest classifier with ten machines for both functions. The cross-validation method was applied to all machines to find optimum C and kernel parameters. The kernel parameters is presented in 'CVSVModelsPol.mat' and 'CVSVModelsRBF.mat'. The best function was chosen by checking the estimation of the out-of-sample misclassification rate, which is presented in Tabel 1.

	Polynomial	RBF
kfoldLoss	0.1489	0.0926

Table 1 misclassification rate.

The RBF SVM has a higher performance than Polynomial and is checked on the test dataset. The correct classification rates on top-left cell and confusion matrices for each of the ten classes are presented in Tabel 2.

“airplane”

rate=0.95	negative	positive
negative	8972	29
positive	504	495

“automobile”

rate=0.95	negative	positive
negative	8977	23
positive	433	567

“bird”

rate=0.94	negative	positive
negative	8997	3
positive	579	421

“cat”

rate=0.93	negative	positive
negative	9000	0
positive	699	301

“deer”

rate=0.92	negative	positive
negative	8996	4
positive	698	302

“dog”

rate=0.95	negative	positive
negative	8983	17
positive	531	469

“frog”

rate=0.94	negative	positive
negative	8997	3
positive	571	429

“horse”

rate=0.95	negative	positive
negative	8984	16
positive	445	555

“ship”

rate=0.96	negative	positive
negative	8938	62
positive	371	629

“truck”

rate=0.95	negative	positive
negative	8949	50
positive	430	571

Table 2 The classes confusion matrices and classification rate.

The overall correct classification rates and confusion matrices for each of the ten classes are presented in Table 3.

rate=0.4739	negative	positive
negative	0	207
posit	5054	4739

Table 3 The overall confusion matrices and classification rate.

The overall correct classification rate is significantly worse than in the CNN method because of the complexity of data and of the case when the image with different classes is close enough that it is impossible to delineate the boundaries. Also, SVM is slower than CNN and needs more calculation, and their need to be more difficult.

4. Conclusion.

The advantages of the SVM are robustness for results, absence of local minima problems, and overfitting. Also, SVM has better performance on high dimensional space on the data with clear class boundaries. Moreover, SVM is more effective with memory use than CNN. However, SVM is more time-consuming on big datasets than CNN. Also, support vector machines are not robust if trained data is noisy because of overlapping. The CNN and SVM are comparable, but SVM has trained longer than CNN on the same data. The best option for solving the image classification problem is to combine two methods. The CNN convolution and filters part of the network could be used to find images' features, and SVM could be used to classify these features. This method could significantly increase the accuracy and time of training.

Reference

1. The official MATLAB site: https://www.mathworks.com/
2. The professor M.R Azimi lecture 17-18