# Fast Classification with Online Support Vector Machines

**Seyda Ertekin**
Computer Science & Engineering
The Pennsylvania State University
University Park, PA, 16802

**Leon Bottou**
NEC Labs America
4 Independence Way
Princeton, NJ, 08540

**C. Lee Giles**
College of Information
Sciences & Technology
The Pennsylvania State University
University Park, PA, 16802

## Abstract

Technological advances in the past several decades have facilitated the generation and storage of digital information. As the amount of digital information increases, there arises the need for more effective tools to better find, filter and manage these resources. Therefore, developing fast and highly accurate algorithms for automatic classification of digital data has become an important part of the machine learning and knowledge discovery research. This poster presents a fast online Support Vector Machine (SVM) classifier algorithm that has an outstanding speed improvement over the classical (batch) SVMs and the other online SVM algorithms, while preserving the classification accuracy rates of the state-of-the-art SVM solvers. The speed improvement and the demand for less memory with the online learning setting enable the SVMs to be applicable to very large data sets.

## Introduction

The amount of digital data has been growing rapidly due to the widespread use of computers and advances in storage systems. The need to access the right and relevant information and the desire to organize and categorize data has increased significantly. This problem has been studied extensively by the machine learning community, as it is a classical example of supervised learning. In supervised learning, computers learn a classifier that is generated using the information from the labeled data instances provided by the supervisor. The classifier is then expected to classify unlabeled instances into one or more predefined categories based on their content. It is known that the learner's approximation of the classifier function improves with the amount of training data supplied to it. Collecting large amount of training data is not difficult due to the abundance of digital data. However, processing such vast amount of information is a challenging task because of the fact that the growth of digital information outpaces the increase of processor speeds. This requires improved machine learning algorithms to make computers to learn faster from labeled training data. SVM is a popular machine learning algorithm for data classification due to its strong theoretical foundation and good generalization performance. However, SVMs have not yet seen widespread adoption in the communities working with very large datasets due to the high computational cost involved in solving quadratic programming (QP) problem in the training phase. This poster presents a fast online SVM classifier algorithm that preserves the highly competitive classification accuracy rates of the state-of-the-art SVM solvers while requiring less computational resources. We also show that better speed performance can be achieved by intelligent selection of the most informative examples from the training set.

### Fast Online SVM Algorithm: LASVM

Support Vector Machines (Cortes & Vapnik, 1995) have become a popular machine learning algorithm in the last decade for data classification tasks due to their strong theoretical foundation and good classification accuracies that has been demonstrated in a wide range of application domains. However, the high computational cost involved in solving QP problem in the training phase and the need for a large main memory to hold the kernel products limit the applicability of the SVMs to very large data sets. In this poster, we present our online SVM algorithm, namely LASVM (Bordes *et al*. 2005), which tolerates much smaller main memory and has a faster training phase. LASVM is an online kernel classifier which relies on the traditional soft margin SVM formulation. It is a reorganization of the Sequential Minimal Optimization (SMO) direction searches (Platt, 1998) to optimize the cost function and it converges to the SVM solution (Bordes *et al*. 2005). The results of our experiments indicate that, when combined with the benefits of online learning, an intelligent way of reorganization of SMO direction searches yield considerable speedups in training of the SVM classifiers. Due to the space constraints, we refer readers to (Bordes et. al. 2005) for a detailed analysis of the ideas discussed here.

Classical SVMs work in a batch setting where the algorithm has a fixed collection of examples in hand and uses them to construct a hypothesis, which is used thereafter for classification without further modification. LASVM, on the other hand, works in an online setting, where the algorithm continually modifies its hypothesis as new training instances become available. It repeatedly accepts a new training instance, predicts its classification according to the model at that point of time, finds out the correct label and updates its hypothesis, if necessary, based on this new observation. Therefore compared to a traditional SVM solver, LASVM brings the computational benefits and flexibility of online learning algorithms.

Part of this work was funded by NSF grant CCR-0325463.

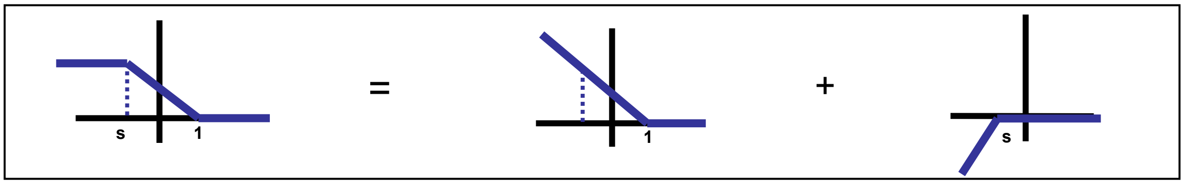


**Figure 1. Ramp Loss function (left) can be decomposed into the sum of the convex Hinge Loss (middle) and a concave loss (right)**

## LASVM with Active Learning and Non-Convex Loss Function

We believe that not all examples are equally informative in the training set. Our experimental results demonstrate that intelligent selection of the informative examples from the training data and exploiting those examples in the process of optimizing the cost function yields faster learning algorithm without the loss of any classification accuracy. In our previous work (Bordes *et al*. 2005), we showed how active example selection can yield faster training, higher accuracies (in noisy collections) and simpler models using only a fraction of the training examples' labels. We also proposed a simple, yet very effective, modification to the classical SVM based Active Learning that enables its applicability to very large datasets. With our contribution, the task of searching the entire dataset at each step to find the most informative example has left its place to searching only the small number of randomly selected data.

We are also investigating another method to further speed up the LASVM algorithm. We believe that the outliers are getting more attention than they should because of the characteristics of the Hinge Loss (Fig. 1, middle). Prior research shows that by using a non-convex loss function instead of the classical convex Hinge Loss, it is possible to achieve an SVM classification algorithm where training errors are no longer support vectors (Collobert, Sinz, Weston & Bottou, 2006). Having fewer number of support vectors is beneficial in terms of less memory requirements and faster SVM algorithms. Therefore we combined the non-convex loss paradigm with the online LASVM algorithm. Although the theoretical foundations of non-convex algorithms are more challenging than the convex algorithms, we show that they can achieve very good results in practice (Ertekin, Bottou & Giles, 2005). Representing the non-convex cost function as the sum of a convex part and concave part like in Figure 1, each iteration of the Concave Convex Procedure (CCCP) approximates the concave part by its tangent and minimizes the resulting convex function. Our initial results show that using a non-convex Ramp Loss (Fig. 1, left) provides significant speed improvement in LASVM, especially on noisy data.

In our experiments, we primarily focused on two values for the Ramp Loss parameter *s* (Figure 1).

$s = -1$ Symmetric loss where outliers can not be SVs.

$s = 0$ Asymmetric loss where misclassified examples can not be SVs.

We observed a correlation in the working principles of one of our proposed Active Learning methods and the Ramp Loss method with *s* parameter adjusted to -1. In both cases, the examples in the margin tend to be selected for training and become support vectors. We believe that this similarity deserves further investigation in theoretical understanding of why SVM based Active Learning method works so well. We plan to focus on this remarkable correlation between Active Learning and Ramp Loss based optimization.

## Visual Poster Layout

We will present a poster first describing the motivation of the research. Then, by using some colored graphs and drawings, we will give some brief background information on SVMs, the support vectors and outliers concept, and the meaning of online and passive learning. We will also include the working principle of SMO technique since it forms the basis of our algorithm. Then, we will show the design and implementation details of LASVM and make comparisons between LASVM and other competitive algorithms. We will demonstrate classification accuracy graphs for text and OCR images collections. Afterwards, we will present the second phase of LASVM and describe the methods that we used to find the most informative examples from training data. We will give background information on Active Learning and loss functions. We will also show by examples what we actually mean by saying informative example. Then we will show how Active Learning method and Ramp Loss function made LASVM faster without any classification accuracy loss. We will finally present graphical experimental results to show that LASVM outperforms other algorithms. We will conclude with our future research goals.

## References


Bordes A., Ertekin, S., Weston J., Bottou L. Fast Kernel Classifiers with Online and Active Learning. *Journal of Machine Learning Research (JMLR), vol. 6, pp. 1579-1619, 2005.*

Collobert R., Sinz F., Weston J., Bottou L. Trading Convexity for Scalability. *ICML, 2006.*

Cortes C., Vapnik V. Support vector networks. *Machine Learning, 20:273-297, 1995.*

Ertekin S., Bottou L., Giles L. Non Convex Online Support Vector Machines & Active Learning. *Working paper, IST, The Pennsylvania State University, 2005.*

Platt J. Fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods – Support Vector Learning. B. Scholkopf, C. Burges and A. Smola Eds., MIT Press, 1998*