

# Multiclass Classification with Pegasos

Gabriele Cerizza

Università degli Studi di Milano  
`gabriele.cerizza@studenti.unimi.it`  
<https://github.com/gabrielecerizza/smml>

## Introduction

In this report we detail the results of our experiments on the Pegasos algorithm [2] over a multiclass classification task, pursuant to the project specifications set out for the Statistical Methods for Machine Learning course of the Università degli Studi di Milano<sup>1</sup>.

In Section 1 we illustrate the dataset used in the experiments. In Section 2 we briefly describe the algorithm and its implementation. In Section 3 we show the results of our experiments and provide comments thereon. Finally, Section 4 contains our concluding remarks.

## 1 Dataset

Our analysis was carried out on the USPS dataset<sup>2</sup>, comprising 9298  $16 \times 16$  grayscale images. These images depict handwritten digits ranging from 0 to 9. A sample of these digits can be seen in Figure 1. The classes do not present severe imbalance issues, as shown in Figure 2.

To gauge the complexity of multiclass classification on this dataset, we performed dimensionality reduction by way of the t-distributed Stochastic Neighbor Embedding (t-SNE) [1] algorithm and projected the instances onto a two-dimensional space. The result is shown in Figure 3. Within this figure the instances are clustered into ten clearly distinguishable groups, thus suggesting that a satisfying classification can be attained.

The USPS dataset is made available with predetermined training and test sets. In our experiments, however, we merged these sets and proceeded to form our own training and test sets, as illustrated in Section 3.

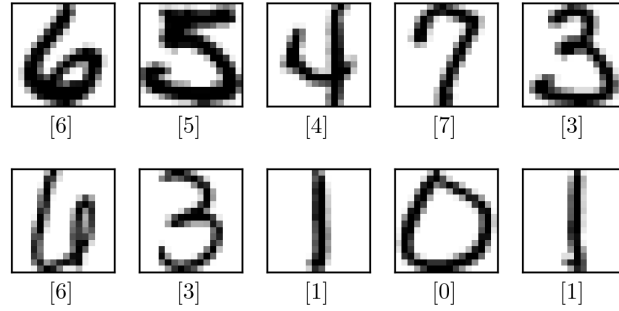
## 2 Pegasos Algorithm

In order to perform the multiclass classification task, we employed the Pegasos algorithm [2]. This algorithm was introduced to solve the Support Vector

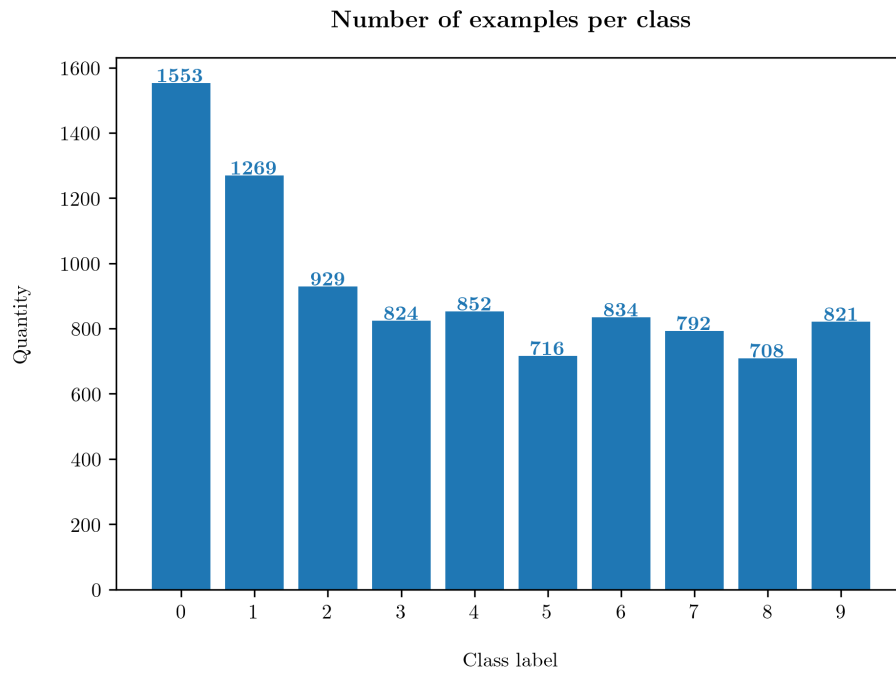
---

<sup>1</sup> <https://cesa-bianchi.di.unimi.it/MSA/index.21-22.html>

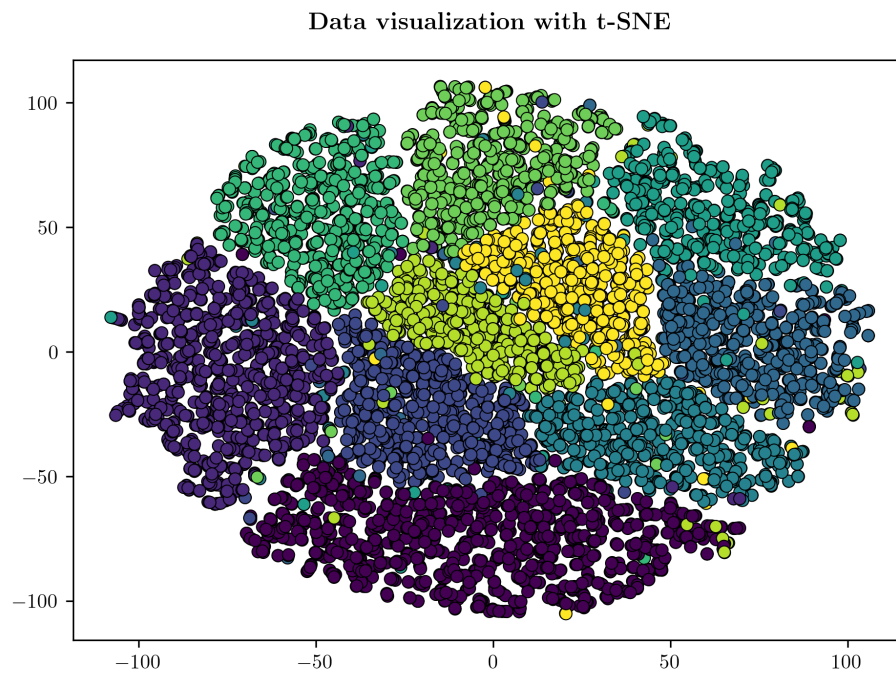
<sup>2</sup> <https://www.kaggle.com/datasets/bistaumanga/usps-dataset>



**Fig. 1.** Sample of images from the USPS dataset. The label is indicated below each image.



**Fig. 2.** Frequency of the classes in the USPS dataset.



**Fig. 3.** Projection of the USPS dataset onto a two-dimensional space using t-SNE.

Machine (SVM) [3] convex optimization problem, which can be formalized as follows.

Let  $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$  be a training set with  $\mathbf{x}_t \in \mathbb{R}^d \ \forall t = 1, \dots, m$ ; and let  $\phi : \mathbb{R}^d \rightarrow \mathcal{H}$  be a function such that  $\dim(\mathcal{H}) \gg d$ . The optimization problem to solve is:

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^d, \xi \in \mathbb{R}^m} \quad & \frac{\lambda}{2} \|\mathbf{w}\|^2 + \frac{1}{m} \sum_{t=1}^m \xi_t, \\ \text{s.t.} \quad & y_t \mathbf{w}^\top \phi(\mathbf{x}_t) \geq 1 - \xi_t \quad t = 1, \dots, m, \\ & \xi_t \geq 0 \quad t = 1, \dots, m, \end{aligned} \tag{1}$$

where

Under scrutiny

### 3 Experiments

The experiments were run on a machine with 16 GB of RAM and a CPU Intel(R) Core(TM) i7-9700K 3.60GHz with 8 cores.

### 4 Conclusions

#### Statement

*I/We declare that this material, which I/We now submit for assessment, is entirely my/our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my/our work. I/We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me/us or any other person for assessment on this or any other course of study.*

#### References

1. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. *Journal of Machine Learning Research* **9**, 2579–2605 (11 2008)
2. Shalev-Shwartz, S., Srebro, N., Cotter, A.: Pegasos: Primal estimated sub-gradient solver for svm. *Math. Program.* **127**, 3–30 (03 2011)
3. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, second edn. (November 1999)