Support Vector Machines for binary classification

D. Sharapov, L.Kottsova, D.Karpukhina (MIPT DCAM)

Abstract: The task of multiclass classification plays a crucial role in machine learning. To solve it, the set of binary classification tasks is often used. This approach implies that binary models are built between classes. Therefore, the necessity to build a fast and precise binary classifier arises. In this work, we have conducted research on SVM model construction in the tasks of binary classification. As a tool to define the support vectors, three different approaches were considered. For these approaches, one common optimization task was solved using Nesterov's accelerated gradient descent.

Key words: SVM, Gauss RBF kernel, Polynomial Kernel, Linear Kernel, Nesterov's Fast Gradient Descent

The task of multi-class classification appears, for instance, in the task of identifying scanned handwritten digits [1]. One of the ways to solve this is the method of support vector machines. In many cases multi-class classification is reduced to binary classification. For the latter, it is convenient to build support vectors, on which the borderline between classes is easy to build.

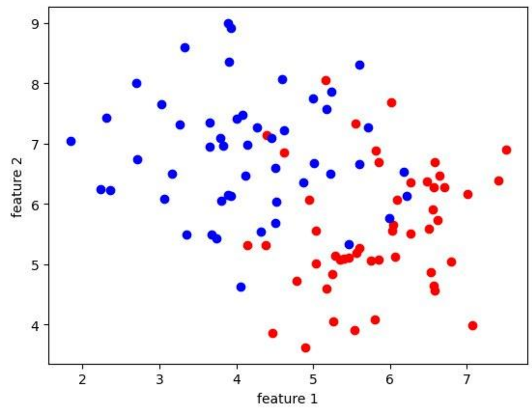
The construction of a hyperplane between two sets of points, called linear separable, can be called the simplest task of constructing support vectors. This task is vital, as it allows to identify classes in non-linear separable data, where with the help of specific mappings, that will be referenced as kernels trick, the input data is converted into a space, where it is linear separable.

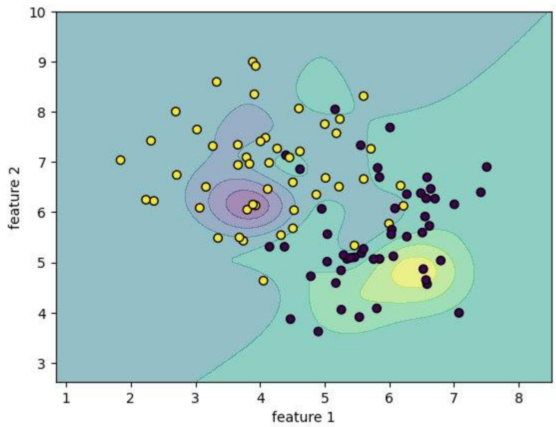
The idea of solving the tasks of binary classification using SVM is based on maximization of margin's width [3], which is in fact the task of the quadratic problem of minimization. In many cases, it suggested to not only expand the margin, but also penalize the model for wrong answers. Therefore, the problem of minimization of the loss function with regularization arises. In addition to l2-regularization, other methods of regularization can be used [4].

As a penalty function, the Hinge loss, Smooth Hinge loss (used in this work), Logistic loss, exponential loss can all be used [5].

After defining the objective function, we should focus on the method of its minimization. In this work, Nesterov's fast gradient descent is used [6,7]. It requires the usage of L-smooth functions, that is why instead of Hinge Loss, which is usually applied in SVM tasks, Smoothed Hinge Loss, requiring an additional hyperparameter, has to be used[8].

Thereby, we have set the task of minimization, that we suggest to solve, completing all of the steps, described above. The accuracy result for the derived class is 0.85, which is rather close to the results for the classic solver from the Python module [9], amounting to 0.90.





Literature

1. https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html
2. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Shujun Huang, Nianguang Cai , Pedro Penzuti Pacheco, Shavira Narrandes, Yang Wang And Wayne Xu, 2018.
3. The Elements of Statistical Learning. Trevor Hastie, Robert Tibshirani, Jerome Friedman. 2017.
4. Regularization approaches for support vector machines with applications to biomedical data. Daniel Lopez-Martinez. 2017.
5. Learning with Smooth Hinge Losses. Junru Luo, Hong Qiao, and Bo Zhang. 2021.
6. Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence O(1/k2). Dokl. Akad. Nauk. USSR 269(3), 543–7 (1983)
7. Nesterov, Y.: Introductory lectures on convex optimization: A basic course. Kluwer (2004). DOI 10.1007/ 978-1-4419-8853-9
8. Epoch Gradient Descent for Smoothed Hinge-loss Linear SVMs. Soomin Lee and Angelia Nedic. 2013.
9. Fast and Scalable Polynomial Kernels via Explicit Feature Maps. Ninh Pham, Rasmus Pagh.