Project
Regularized SVM algorithms for classification problems

Larisa Beilina, larisa@chalmers.se

Regularized SVM algorithms for classification problems

In this project will be studied regularized versions of support vector machines (SVM) algorithms for solution of classification problems. Application of this project can be in classification of skin images from the ISIC project, see link of the project

`https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main`

ML and Tikhonov's regularization techniques will be studied. Tikhonov's regularization technique is presented in the paper [2].

Details about ML algorithms for classification together with machine learning techniques for choosing the regularization parameter can be found in [1, 3, 4].

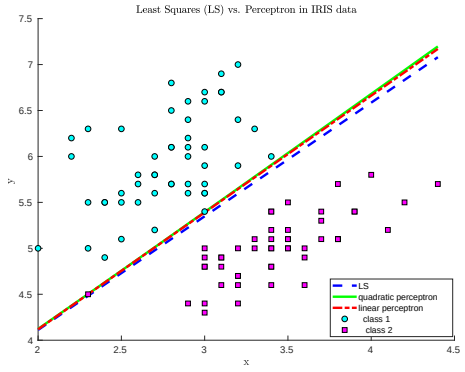

Figure 0.1: Decision lines computed by least squares and the perceptron learning algorithm for separation of two classes using Iris dataset. See details in the project "Regularized least squares and machine learning algorithms for classification" at the course [5].

Hints:

- I) Implement in MATLAB classification algorithms (perceptron, WINNOW, support vector machines (SVM)) and check them on simulated data. Apply developed algorithms on classification of skin images from the ISIC project. You can use the MATLAB code for non-regularized classification at [6].
- II) Try to test different regularization techniques (1-2 techniques is enough to test, choose any Tikhonovs technique of [2] or machine learning technique) for choosing

the regularization parameter. More precisely, test some of Tikhonov's techniques (a priori rule, Morozov's discrepancy, balancing principle) described in Section 5 of [2]. Machine learning techniques for choosing the regularization parameter are presented in Section 7 of [3].

- III) Compute missclassification rate E using the formula (see [4], p. 211-214):

$$\mathrm{E} = \frac{\sum_{i=1}^{K} N_{F,i}}{\sum_{i=1}^{K} (N_{T,i} + N_{F,i})}, \tag{0.1}$$

where K is the number of classes, $N_{T,i}$ is the number of points of the class i which are classified correctly, $N_{F,i}$ is the number of points of the class i which are classified wrong. Precision for class i can be computed as

$$\mathrm{P}(i) = \frac{N_{T,i}}{N_{T,i} + N_{F,j}}. \tag{0.2}$$

- IV) Optional: take some other experimental data for classification from the link

`https://archive.ics.uci.edu/ml/datasets.html`

and classify them.

References

[1] Christopher M. Bishop, *Pattern recognition and machine learning*, Springer, 2009.

[2] L. Beilina, Numerical analysis of least squares and perceptron learning for classification problem, `https://arxiv.org/abs/2004.01138`

[3] Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press, 2016, `http://www.deeplearningbook.org`

[4] Miroslav Kurbat, *An Introduction to Machine Learning*, Springer, 2017.

[5] "Numerical methods and machine learning algorithms for solution of Inverse problems", the course page, `https://canvas.gu.se/courses/122370000000049154`

[6] The Project WavES, `https://waves24.com/download/`