Optmization Methods for Machine Learning - Fall 2016

Assignment # 3
Support Vector Machines

Laura Palagi

Dip. di Ingegneria informatica automatica e gestionale A. Ruberti, Sapienza Università di Roma

Posted on November 27, 2016 - due date December 29, 2016

Instructions

Homework will be done individually: each student must hand in their own answers. It is acceptable for students to collaborate in figuring out answers and helping each other solve the problems. We will be assuming that, as participants in a graduate course, you will be taking the responsibility to make sure you personally understand the solution to any work arising from such collaboration.

Homework must be send by an email both to the teaching assistant Ing. Umberto Dellepiane (umberto.dellepiane@act-OperationsResearch.com) and to laura.palagi@uniroma1.it with subject **[OMML-2016] Project 3**. The mail must contain, as attachment, a file COGNOME_OMML_PROJECT3.zip or .tar.gz containing both a typed report in English and the source code (including instructions to run it).

After you submit, you will receive an acknowledgement email that your project has been received. If you have not received an acknowledgement email within 2 days after you submit then contact the instructor.

Evaluation criteria

As for the **evaluation**, you may reach the max score by correctly answering to all points of Question 1 using the second test set (handwritten numerical digits). Correct answers to all the points of Question 1 using the two dimensional set allow to obtain only up to approximatively 27.
The third homework accounts for 30% of the total vote of the exam.
Homework is due at latest at midnight on the due date. For late homework, the score will be decreased. It is worth 85% for the next 48 hours. It is worth 70% from 48 to 120 hours after the due date. It is worth 50% credit after 120 hours delay.
For the evaluation of the third homework the following criteria will be used:

1. check of the implementation (60% Umberto Dellepiane)
2. Quality of the explanation document and of the overall job (40% - Laura Palagi).

The grade are Italian style namely in the range [0,31], being 18 the minimum degree to pass the exam and 31 = 30 cum laude.

In this assignment you will implement training methods for Support Vector Machines (supervised learning) applied to classification problems.

Data set.

You may choose among two different data set. Different choice allows to get different score. In particular the mark is obtained by multiplying the final mark of Questions 1 and 2 by a coefficient which is ≤ 1. The coefficient of each data set is reported below .

- DATA SET 1 (coefficient 0,9 = max score achievable ≈ 27.9) The two-dimensional input samples reported in the picture **??**. For the supervised learning problem you use as training set the pairs (x^i, y^i) with $x^i \in \mathbb{R}^2$ and $y^i \in \{-1, 1\}$. The coordinates x^i can be estimated from the chart.

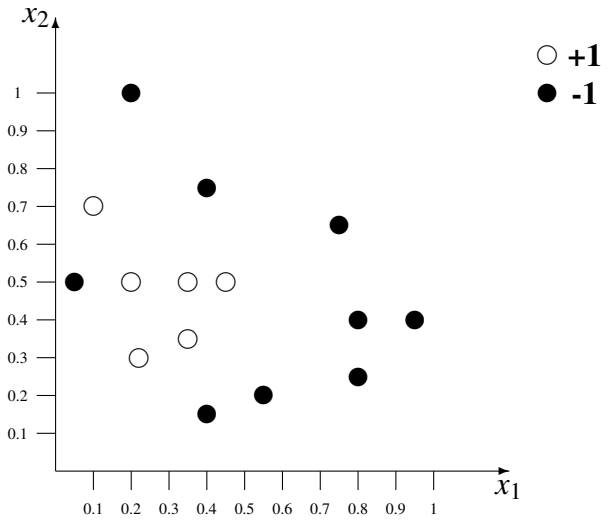

Figure 1: Sample in the two class

- DATA SET 2 (coefficient 1 = max score achievable "30 cum laude") Classification of normalized handwritten digits, automatically scanned from envelopes by the U.S. Postal Service. The images here have been deslanted and size normalized, resulting in 16 x 16 grayscale images (Le Cun et al., 1990).

  The training data are available as separate files per digit (0 1 2 3 4 5 6 7 8 9). Each line consists of the 256 grayscale values. You are receiving only the data set for the numbers 0 1 8 and you need to choose two of these.

  There training and test observations, are distributed as follows:

number	0	1	2	3	4	5	6	7	8	9	Total
Train	1194	1005	731	658	652	556	664	645	542	644	7291
Test	359	264	198	166	200	160	170	147	166	177	2007

The test set can be notoriously "difficult", and a 2.5% error rate is excellent.

In the report you must state:

- which data set do you choose

- If data set 1: report the points in the training set;

- If data set 2: report the two digits you choose and the dimension of the training set that you use.

Question 1. Consider a nonlinear SVM. Choose among a Gaussian kernel $k(x,y) = e^{-\gamma\|x-y\|^2}$ or a polynomial kernel $k(x,y) = (x^T y + 1)^p$ with $p \geq 1$.

Determine the nonlinear decision function

$$y(x) = sign(w^T x + b)$$

where w, b are obtained by means of nonlinear SVMs.

Answer to the following questions.

1. **(max score up to "26")** Write a matlab code which implements the dual quadratic problem and use a standard QP algorithm for its solution.

   In the report you must state:

   - which kernel do you choose;

   - the final setting for the parameter C and of the kernel parameter; how do you choose them and if you can put in evidence over/under fitting;

   - which optimization routine do you use for solving the quadratic minimization problem and the setting of its parameters, if any; you must use the gradient of the objective function $\nabla f(\alpha) = Q\alpha - e$

   - performance from machine learning perspective: report the value of the error on the training and test set;

   - optimization performance: report the number of iterations, the number of gradient evaluations, the computational time

   - If the problem is the two dimensional one, produce a plot of a non-linear decision boundary.

2. **(max score up to "28")** Write a matlab code which implements a decomposition method for the dual quadratic problem with any value $q \geq 2$ (use the same kernel and parameter setting as in Question 1). You must define the selection rule of the working set, construct the subproblem at each iteration and uses a standard QP algorithm for its solutions. A stopping criterion on optimality is needed.

In the report you must state:

- which optimization routine do you use for solving the quadratic minimization subproblem and the setting of its parameters, if any; you must use the gradient of the objective function of the subproblem $\nabla_{\alpha_W} f(\alpha)$

- performance from machine learning perspective: report the value of the error on the training and test set;

- optimization performance: report the number of iterations, the number of gradient evaluations, the computational time

- If the problem is the two dimensional one, produce a plot of a non-linear decision boundary.

3. **(max score up to "30 cum laude")** Fix $q = 2$ and implement the analytic solution of the subproblem. In the report you must state:

   - performance from machine learning perspective in comparison with methods above;

   - optimization performance in comparison with methods above.