Optimization Methods for Machine Learning - Fall 2017

Assignment # 2
Support Vector Machines

Laura Palagi

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

Posted on December 13, 2017 - due date January 10, 2018

Instructions

Homework will be done in groups formed by 1 to 3 people: each group must hand in their own answers. We will be assuming that, as participants in a graduate course, every single student will be taking the responsibility to make sure his/her personal understanding of the solution to any work arising from such collaboration.

Homework must be sent by an email both to the teaching assistant Ing. Tommaso Colombo (tommaso.colombo@uniroma1.it) and to laura.palagi@uniroma1.it with subject **[OMML-2017] Project 2**. 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 instructors.

The mail must contain as attachment a .zip or .tar.gz file with both a typed report in English and the source code following instructions in the text of the project. **The report must be of at most 5 pages excluded figures that must be put at the end.**

Evaluation criteria 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.
The grade are Italian style namely in the range [0,30], being 18 the minimum degree to pass the exam.
The second homework accounts for 35% of the total vote of the exam.
For the evaluation of the second homework the following criteria will be used:

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

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

Data set.

The data set is provided separately. It corresponds to a health application.

Question 1. Consider a nonlinear SVM with kernel k and determine the nonlinear decision function

$$y(x) = sign(\sum_{p=1}^{P} \alpha_p^* y^p k(x^p, x) + b)$$

where α^*, b are obtained by solving a nonlinear SVM problem.

As kernel function you may choose either a RBF kernel

$$k(x,y) = e^{-\gamma \|x-y\|^2}$$

where γ is an hyper parameter or a polynomial kernel

$$k(x,y) = (x^T y + 1)^p$$

with hyper parameter $p \geq 1$.

Answer to the following questions.

1. **(max score up to "24")** Write a program 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 hyper parameter C and the hyper parameter of the kernel; how do you choose them and if you can identify values that highlight 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 a routine which uses the gradient of the objective function $\nabla f(\alpha) = Q\alpha - e$

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

   - optimization performance: report the number of iterations, the number of gradient evaluations, the computational time, the value of the optimal objective function

2. **(max score up to "27")** Write a program 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 accuracy on the training and test set;

- optimization performance: report the number of iterations, the number of gradient evaluations, the computational time, the value of the optimal objective function

3. **(max score up to "30 cum laude")** Fix $q = 2$ and implement a most violating pair (MVP) decomposition method which uses the analytic solution of the subproblems. In the report you must state:

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

   - optimization performance in comparison with methods above.

In the final report (pdf) the comparison among all the implemented methods - in term of accuracy in learning and computational effort in training - must be gathered into a final table (a minimal sample below)

Ex		settings	Training error	test error	optimization its	time
Q1.1	Full QP	$C =?$, $\gamma/p =?$				
Q1.2	Decomposed QP	$C =?$, $\gamma/p =?$				
Q1.3	MVP	$C =?$, $\gamma/p =?$				

MANDATORY Follow the separated instructions for submitting **the code** either Phyton (preferred) or Matlab.