Convex Optimization

Please send your exam and code to `aspremon@ens.fr` with subject `Exam M1` by **Friday Jan. 15, 2020 at 14:00.** Late exams will not be graded.

Exercice 1 (Kernel learning) Given some labeled data points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $x_i \in \mathbf{R}^d$ and $y_i \in \{-1, 1\}$, the soft margin support vector machine training problem is written

$$\begin{array}{ll} \text{minimize} & \|w\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{subject to} & y_i(w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, \ldots, n \end{array}$$

which is a quadratic program in the variables $w \in \mathbf{R}^d$, $\xi \in \mathbf{R}^n$ and $b \in \mathbf{R}$.

- Compute the dual of this SVM training problem.
- Let $K = X^T X$ where X is the matrix whose columns are equal to x_i, and call the optimal value of this problem $\omega(K)$. Show that $\omega(K)$ is a convex function of K.
- Suppose we now directly define the problem in terms of a kernel matrix

  $$K(\lambda) = \sum_{j=1}^m \lambda_j K_j$$

  where K_j are (given) kernel matrices and $\lambda \geq 0$ with $\mathbf{1}^T \lambda = 1$. Show that learning the kernel that minimizes the regularized classification error $\omega(K(\lambda))$ as a function of λ is a convex problem.

Exercice 2 (SOCP) Show that the following program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x^T(A - bb^T)x \leq 0 \\ & b^T x \geq 0 \\ & Dx = g \end{array}$$

in the variable $x \in \mathbf{R}^n$, where $A \in \mathbf{S}_n$ and $A \succeq 0$, is convex.

Hint: if $A \succeq 0$, it can be written $A = A^{1/2}A^{1/2}$ where $A^{1/2}$ is symmetric.

Exercise 3 (Duality.) Derive a dual problem for

$$\text{minimize} \quad \sum_{i=1}^N \|A_i x + b_i\|_2 + (1/2)\|x - x_0\|_2^2.$$

The problem data are $A_i \in \mathbf{R}^{m_i \times n}$, $b_i \in \mathbf{R}^{m_i}$, and $x_0 \in \mathbf{R}^n$. First introduce new variables $y_i \in \mathbf{R}^{m_i}$ and equality constraints $y_i = A_i x + b_i$.

Exercise 4 (Support vector machines solvers.) Given m data points $x_i \in \mathbf{R}^n$ with labels $y_i \in \{-1, 1\}$. Write a function to solve the classification problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|w\|_2^2 + C\mathbf{1}^T z \\ \text{subject to} & y_i(w^T x_i) \geq 1 - z_i, \quad i = 1, \ldots, m \\ & z \geq 0 \end{array}$$

in the variables $w \in \mathbf{R}^n$, $z \in \mathbf{R}^m$, and its dual (warning: this problem is a bit simpler than the one in exercise 1).

- Use the barrier method to solve both primal and dual problems. Test your code on random clouds of points (e.g. generate two classes of data points by picking two bivariate Gaussian samples with different moments). Try various values of $C > 0$ and measure out-of-sample performance. You may set up the data so that finding an initial feasible point is very simple so that you do not need to run a phase one problem.
- Plot primal gap $f(x_k) - f_\star$ versus iteration number for some of the inner Newton iterations, where $f_\star$ is the best value found.
- Plot a simple separation example in 2D (you may add a constant coefficient to the data points x to allow classifiers that do not go through the origin).

NOTE: Please use *graphics and tables* to illustrate your results as much as possible. You can use MATLAB or general purpose languages such as PYTHON or JULIA.

Enjoy :)