Computer Lab Support Vector Machines
SD211

Anas Barakat, Olivier Fercoq, Nidham Gazagnadou, Ekhine Irurozki, Kimia Nadjahi
22 January 2021

The computer lab should be done in pairs or alone. You will write a report and attach the code of functions, and submit it before 4 February at 23:59 on `e-campus`. Each student submits the report of the pair (the report will thus be evaluated twice). You can send either a python notebook or a pdf file.

Then we will allocate two reports for each student to evaluate.

1 Data

We are going to use the data base `breastcancer`. For each observation i, it has a set of features x_i and a lable y_i. To load the data base, download the file `wdbc_M1_B0.data` and use the function `breastcancer_utils.py` in the supplementary file.

2 Sub-gradient method

We wish to solve the following problem

$$\begin{aligned} \min_{v\in\mathbb{R}^m, a\in\mathbb{R}, \xi\in\mathbb{R}^n} \; & \frac{1}{2}\sum_{j=1}^{m} v_j^2 + c\sum_{i=1}^{n}\xi_i \\ & \xi_i \geq 0, \quad \forall i\in\{1,\dots,n\} \\ & \xi_i \geq 1 - y_i(x_i^\top v + a), \quad \forall i\in\{1,\dots,n\} \end{aligned} \tag{1}$$

where $c = 1$.

Question 2.1
Show that Problem (1) is equivalent to

$$\min_{v\in\mathbb{R}^m, a\in\mathbb{R}} \frac{1}{2}\sum_{j=1}^{m} v_j^2 + c\sum_{i=1}^{n}\max(0, 1 - y_i(x_i^\top v + a)) \tag{2}$$

This means that the value of the minimum is the same and that we can reconstruct a minimizer of (1) from a minimizer of (2) and vice versa.

The objective function of (2) is not differentiable but it has subgradients at all points. We will thus try to minimize it by the subgradient method. For a convex function f, the method writes:

$$\begin{aligned} \text{choisir } g_k &\in \partial f(x_k) \\ x_{k+1} &= x_k - \gamma_k g_k \end{aligned}$$

where $(\gamma_k)_{k\geq 0}$ is a step size sequence. Let us denote $\bar{x}_k^\gamma = (\sum_{l=0}^k \gamma_l x_l)/(\sum_{j=0}^k \gamma_j)$. We can show that there exists $C > 0$ such that

$$f(\bar{x}_k^\gamma) - f(x^*) \leq C \frac{\sum_{l=0}^k \gamma_l^2}{\sum_{l=0}^k \gamma_l}$$

Hence, under the condition $\lim_{k\to\infty} \frac{\sum_{l=0}^k \gamma_l^2}{\sum_{l=0}^k \gamma_l} = 0$, the sequence $(\bar{x}_k^\gamma)$ will allow us to get close to a minimizer of the function. We can choose for instance $\gamma_k = 1/(k+1)$.

Question 2.2
Let $h : \mathbb{R} \to \mathbb{R}$ such that $h(z) = \max(0, 1 - z)$. Check graphically that

$$\partial h(z) = \begin{cases} \{-1\} & \text{if } z < 1 \\ [-1, 0] & \text{if } z = 1 \\ \{0\} & \text{if } z > 1 \end{cases}$$

We say that a function H is separable if there exists $h_1, \ldots, h_p$ such that $H(v) = \sum_{i=1}^p h_i(v_i)$. If H is separable, then

$$\partial H(v) = \partial h_1(v_1) \times \ldots \times \partial h_p(v_p).$$

Question 2.3
Let $f(v, a) = \frac{1}{2}\sum_{j=1}^m v_j^2 + c\sum_{i=1}^n \max(0, 1 - y_i(x_i^\top v + a))$. Show that there exists a linear map M and two separable functions N and H such that $f(v, a) = N(v, a) + cH(M(v, a))$. Deduce from it, thanks to Proposition 2.4.2, that

$$\partial f(v, a) = \partial N(v, a) + cM^\top \partial H(M(v, a)) \ .$$

Then calculate ∂N and ∂H.

Question 2.4
Code a function that returns the value of the objective function and one of its subgradient at a given point. You will use the data base `breastcancer` and $c = 1$. It may be convenient to add a column full of ones to the matrix X.

Question 2.5
Code the subgradient method and run it with $(v_0, a_0) = 0$ as an initial condition.

3 Stochastic subgradient method

Let $f_i(v, a) = \frac{1}{2}\sum_{j=1}^m v_j^2 + cn\max(0, 1 - y_i(x_i^\top v + a))$.

Question 3.1
Let I be a random variable following the uniform law on $\{1, \dots, n\}$. Show that

$$f(v, a) = \mathbb{E}[f_I(v, a)]$$

Question 3.2
Give the subdifferential of the function f_i.

Question 3.3
Code the stochastic subgradient method and run it in order to solve Problem (2).

4 Augmented Lagrangian method

Question 4.1
Recall the Lagrangian function associated to Problem (1).

For $\rho > 0$, we define $g : \mathbb{R}^2 \to \mathbb{R}$ such that $g(x, \phi) = -\frac{1}{2\rho}\phi^2 + \frac{\rho}{2}\big(\max(0, x + \rho^{-1}\phi)\big)^2$. We admit that the Augmented Lagrangian for Problem (1) writes as

$$L_\rho(v, a, \xi, \phi, \psi) = \frac{1}{2}\|v\|_2^2 + c\sum_{i=1}^n \xi_i + \sum_{i=1}^n g(-\xi_i, \phi_i) + \sum_{i=1}^n g(-\xi_i + 1 - y_i(x_i^\top v + a), \psi_i)$$

Question 4.2
Show that the derivatives of g with respect to x and ϕ are

$$\nabla_x g(x, \phi) = \rho \max(0, x + \frac{\phi}{\rho})$$
$$\nabla_\phi g(x, \phi) = \max(-\frac{\phi}{\rho}, x)$$

Question 4.3
Show that the function $(x \mapsto g(x, \phi))$ is convex for all ϕ and that $(\phi \mapsto g(x, \phi))$ is concave for all x.

The Augmented Lagrangian method is given by

$$(v_{k+1}, a_{k+1}, \xi_{k+1}) \in \arg\min_{v,a,\xi} L_\rho(v, a, \xi, \phi_k, \psi_k)$$
$$\phi_{k+1} = \phi_k + \rho\nabla_\phi L_\rho(v_{k+1}, a_{k+1}, \xi_{k+1}, \phi_k, \psi_k)$$
$$\psi_{k+1} = \psi_k + \rho\nabla_\psi L_\rho(v_{k+1}, a_{k+1}, \xi_{k+1}, \phi_k, \psi_k)$$

Question 4.4
Given multipliers ϕ_k and ψ_k, code the gradient method with line search in order to solve $\min_{v,a,\xi} L_\rho(v, a, \xi, \phi_k, \psi_k)$. We shall choose $\rho = 2$ and as a stopping test $\|\nabla_{(a,v,\xi)} L_\rho(v, a, \xi, \phi_k, \psi_k)\| \leq \epsilon$ where $\epsilon = 1$.

Question 4.5
Code a function that computes $\nabla_{(\phi,\psi)} L_\rho(v_{k+1}, a_{k+1}, \xi_{k+1}, \phi_k, \psi_k)$.

Question 4.6
Code the Augmented Lagrangian method and run it for the initial condition $(\psi_0, \psi_0) = 0$ during 2000 iterations.

5 Comparison

Question 5.1
Compare the three algorithms.