TP Séparateurs à Vaste Marge
SD211

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

Le TP se fait en binôme ou seul. Vous ferez un rapport accompagné des fonctions associées aux questions à envoyer au plus tard le 4 février à 23h59 sur `e-campus`. Chaque étudiant soumet le rapport du binôme (le rapport sera donc corrigé deux fois). Vous pouvez rendre le rapport sous forme d'un notebook python ou d'un fichier pdf.

Nous allouerons ensuite à chaque étudiant deux rapports à évaluer.

1 Données

Nous utiliserons la base de données `breastcancer`, qui pour chaque observation i comporte un ensemble de caractéristiques x_i et une étiquette y_i. Pour charger la base de données, téléchargez le fichier `wdbc_M1_B0.data` et utilisez la fonction fournie dans le fichier annexe `breastcancer_utils.py`.

2 Méthode du sous-gradient

On veut résoudre le problème suivant

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

où $c = 1$.

Question 2.1
Montrer que le problème (1) est équivalent au problème

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

c'est à dire que la valeur du minimum est la même et que l'on peut construire un optimum de (1) à partir d'un optimum de (2) et vice versa.

La fonction objectif du problème (2) n'est pas dérivable mais elle possède des sous-gradient en tout point. Nous nous proposons donc d'essayer de la minimiser par la méthode du sous-gradient. Pour une fonction f convexe, la méthode du sous gradient s'écrit :

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

où $(\gamma_k)_{k\geq 0}$ est une suite de pas. Dénotons $\bar{x}_k^\gamma = (\sum_{l=0}^k \gamma_l x_l)/(\sum_{j=0}^k \gamma_j)$. On peut montrer qu'il existe $C > 0$ tel que

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

Ainsi, sous la condition que $\lim_{k\to\infty} \frac{\sum_{l=0}^k \gamma_l^2}{\sum_{l=0}^k \gamma_l} = 0$, la suite $(\bar{x}_k^\gamma)$ va permettre d'approcher un minimiseur de la fonction. On peut par exemple choisir $\gamma_k = 1/(k+1)$.

Question 2.2
Soit $h : \mathbb{R} \to \mathbb{R}$ telle que $h(z) = \max(0, 1-z)$. Vérifier graphiquement que

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

On dit qu'une fonction H est séparable si il existe $h_1, \ldots, h_p$ tels que $H(v) = \sum_{i=1}^p h_i(v_i)$. Si H est séparable, alors

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

Question 2.3
Posons $f(v, a) = \frac{1}{2}\sum_{j=1}^m v_j^2 + c\sum_{i=1}^n \max(0, 1 - y_i(x_i^\top v + a))$. Montrer qu'il existe une application linéaire M et deux fonctions séparables N et H telles que $f(v, a) = N(v, a) + cH(M(v, a))$. En déduire, grâce à la proposition 2.4.2, que

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

puis calculer ∂N et ∂H.

Question 2.4
Coder une fonction qui retourne la valeur de la fonction et un de ses sous-gradient en tout point. Vous utiliserez la base de données `breastcancer` fournie et $c = 1$. Il pourra être pratique de rajouter une colonne de uns à la matrice X.

Question 2.5
Coder la méthode du sous-gradient et la lancer avec comme condition initiale $(v_0, a_0) = 0$.

3 Méthode du sous-gradient stochastique

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

Question 3.1
Soit I une variable aléatoire qui suit la loi uniforme sur $\{1,\dots,n\}$. Montrer que

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

Question 3.2
Donner la sous-différentielle de la fonction f_i.

Question 3.3
Coder la méthode du sous-gradient stochastique et la lancer pour résoudre le problème (2).

4 Méthode du lagrangien augmenté

Question 4.1
Rappeler le Lagrangien associé au problème (1).

Pour $\rho > 0$, on définit $g : \mathbb{R}^2 \to \mathbb{R}$ telle que $g(x,\phi) = -\frac{1}{2\rho}\phi^2 + \frac{\rho}{2}\big(\max(0, x + \rho^{-1}\phi)\big)^2$. On admet que le lagrangien augmenté associé au problème (1) s'écrit

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

Question 4.2
Montrer que les dérivées de g par rapport à x et par rapport à ϕ sont

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

Question 4.3
Montrer que la fonction $(x \mapsto g(x,\phi))$ est convexe pour tout ϕ et que $(\phi \mapsto g(x,\phi))$ est concave pour tout x

La méthode du lagrangien augmenté s'écrit

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

Question 4.4
Étant donnés des multiplicateurs ϕ_k et ψ_k, coder la méthode du gradient avec recherche linéaire pour résoudre $\min_{v,a,\xi} L_\rho(v, a, \xi, \phi_k, \psi_k)$. On choisira $\rho = 2$ et comme test d'arrêt $\|\nabla_{(a,v,\xi)} L_\rho(v, a, \xi, \phi_k, \psi_k)\| \leq \epsilon$ où $\epsilon = 1$.

Question 4.5
Coder une fonction qui calcule $\nabla_{(\phi,\psi)} L_\rho(v_{k+1}, a_{k+1}, \xi_{k+1}, \phi_k, \psi_k)$.

Question 4.6
Coder la méthode du lagrangien augmenté et la lancer pour la condition initiale $(\psi_0, \psi_0) = 0$ pour 2000 itérations.

5 Comparaison

Question 5.1
Comparer les trois algorithmes.