TD5

Friday, December 4, 2020 3:11 PM

1

UNIVERSITE DE GENEVE — Section de Mathématiques

Faculté des sciences

Optimization with Application I

Exercise Sheet 5 - Discussed on 04.12.2020

Exercise 1. For $y \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times d}$ and $\lambda > 0$, derive the formula to implement a relaxation algorithm to solve:

— Ridge Regression

$$\arg\min_\alpha \frac{1}{2}\|y - X\alpha\|_2^2 + \lambda\|\alpha\|_2^2$$

— The Lasso

$$\arg\min_\alpha \frac{1}{2}\|y - X\alpha\|_2^2 + \lambda\|\alpha\|_1$$

$\boxed{\text{Relaxation:}}$

initialize α somehow ← warm start

for i in range(n):

$$\alpha_i = \arg\min_{\alpha_i} \|y - X\alpha\|_2^2 + \lambda \underbrace{\|\alpha\|_1}_{\text{or } \|\alpha\|_2^2}$$

explicit formula

$\boxed{\text{Optimal descent}}$

initialize α somehow

$$\max\left|\begin{pmatrix} 1 \\ 2 \\ [-3,-1] \\ \vdots \\ \vdots \end{pmatrix}\right| = \begin{pmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ \vdots \end{pmatrix}$$

while not done:

subdifferential = subdiff($f(\alpha)$)

i = max index(|subdifferential|)

$\alpha_i = \arg\min_{\alpha_i} \|y - X\alpha\|_2^2 + \lambda\|\alpha\|_1$

Fix i → calculate optimal i-th coordinate given all the others

$$\arg\min_{\alpha_i} \frac{1}{2}\|y - \underbrace{X_{-i}}_{X \text{ without the } i\text{-th column}}\alpha_{-i} - \alpha_i \underbrace{x_i}_{i\text{-th column of } X}\|_2^2 + \lambda\|\alpha\|_2^2$$

$$\arg\min_{\alpha_i} \frac{1}{2}\|\underbrace{y - X_{-i}\alpha_{-i}}_{D \in \mathbb{R}^{N \times 1}} - \alpha_i x_i\|_2^2 + \lambda\alpha_i^2$$

$$= (x_i^T x_i + \lambda)^{-1} x_i^T D$$

$$= \frac{x_i^T D}{\|x_i\|_2^2 + \lambda}$$

[Reminder: $\arg\min_\alpha \frac{1}{2}\|X\alpha - y\|_2^2 + \lambda\|\alpha\|_2^2 = (X^T X + \lambda I)^{-1} X^T y$]

[Reminder: $\arg\min_{\alpha \in \mathbb{R}} \frac{1}{2}\|x\alpha - y\|_2^2 + \lambda|\alpha| = \text{sign}(\langle x, y\rangle)\left(\frac{|\langle x, y\rangle| - \lambda}{\|x\|_2^2}\right)_+$]

$$\arg\min_{\alpha_i} \frac{1}{2}\|X\alpha - y\|_2^2 + \lambda\|\alpha\|_1$$

$$= \arg\min_{\alpha_i} \frac{1}{2}\|\underbrace{y - X_{-i}\alpha_{-i}}_{D} - \alpha_i x_i\|_2^2 + \lambda|\alpha_i|$$

$$= \text{sign}(\langle x_i, D\rangle)\left(\frac{|\langle x_i, D\rangle| - \lambda}{\|x_i\|_2^2}\right)_+$$

Relaxation: − more steps required

\+ easier to implement (no subdifferentials → error prone)

CPU time + ?