Optimization Theory and Algorithm **Homework 4 - 06/01/2021**

Homework 4

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

HW 1 *Prove the following theorems.*

Theorem 1 *f is α-strongly convex if and only if $f(\mathbf{x}) - \frac{\alpha}{2}\|\mathbf{x}\|^2$ is convex.*

Theorem 2 *Suppose $f \in \mathcal{C}^1$. Then the following arguments are equivalent:*

1. *f is α-strongly convex.*
2. *$\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \alpha \|\mathbf{x} - \mathbf{y}\|^2$.*
3. *Additionally, if $f \in C^2$, then $\nabla^2 f \succeq \alpha I$ everywhere ($\nabla^2 f$ is positive definite).*

HW 2 *Please compute subgradients and subdifferentials of the functions.*

(1) $f(x) = \max(x, 0)$ *is called ReLU which is widely used in Deep Learning models.*

(2) $f(\mathbf{x}) = \max_{i=1,\ldots,m}\{\mathbf{a}_i^\top \mathbf{x} + b_i\}$.

HW 3 *Suppose that:*

- *Data: $\{\mathbf{a}_i, b_i\}_{i=1}^m$, where $\mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}$.*
- *$b_i = \mathbf{a}_i^\top \mathbf{x} + \epsilon_i$, where $\mathbf{x} = (x_1, \ldots, x_n)^\top$ is denoted as regression coefficient and $\epsilon_i \sim \mathcal{N}(0, 1)$.*
- *Prior distribution: $\mathbf{x} \sim \mathcal{L}(0, \frac{1}{\lambda} I_n)$, where*

$$\mathbb{P}(\mathbf{x}) = \frac{1}{g(\lambda)} \exp\left\{-\frac{\lambda\|\mathbf{x}\|_1}{2}\right\}.$$

- *Posterior distribution:*

$$\mathbb{P}(\mathbf{x}|A, \mathbf{b}) = \frac{\mathbb{P}(A, \mathbf{b}|\mathbf{x})\mathbb{P}(\mathbf{x})}{\mathbb{P}(A, \mathbf{b})}.$$

Using Maximal Posterior (MAP) Method to derive the optimization formulation of LASSO:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1.$$

HW 4 *The following optimization problem is called elastic net which is proposed by [Zou and Hastie, 2005]:*

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 + (1-\lambda)\|\mathbf{x}\|_2^2. \tag{1}$$

(1) Let $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1 + (1-\lambda)\|\mathbf{x}\|_2^2$. Please compute its proximal operator $prox_{\gamma g}(\mathbf{z})$.

(2) Give us the proximal gradient descent algorithm for solving the elastic net.

HW 5 *See Page 242 of the text book and the readme file.*

(1) Reproduce the results of LASSO problem via the subgradient descent algorithm.

(2) Using proximal gradient descent algorithm to solve LASSO.

(3) Compared the LASSO solution with Ridge regression for different λs.

HW 6 *(Additional Question)*

Suppose that f is a β-smooth function, then please provide an iterative algorithm to solve the following constrained optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x})$$
$$s.t.\ \ A\mathbf{x} = \mathbf{b}.$$

References

[Zou and Hastie, 2005] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.