Comp6211e Homework 2

Assignment date: March 18
Due date: April 1

Theoretical Problems (9 points)

1. (3 points) Consider the regularized logistic regression:

$$f(w) = \frac{1}{n}\sum_{i=1} \ln(1 + \exp(-w^\top x_i y_i)) + \frac{\lambda}{2}\|w\|_2^2,$$

where $x_i \in \mathbb{R}^d$ and $y_i \in \{\pm 1\}$. If $\|x_i\|_2 \leq 1$ for all i, find the best upper bound of the smoothness parameter L of $f(w)$. Find the best Lipschitz constant G in the region $\{w : f(w) \leq f(0)\}$.

2. (3 points) Use the method in Proposition 1 of Lecture 08 to find an (L, ϵ)-smooth approximation $\phi_\epsilon(z)$ of the hinge loss

$$f(z) = (1 - z)_+$$

as

$$\phi_\epsilon(x) = \min_z \left[(1 - z)_+ + \frac{1}{2\epsilon}(z - x)^2\right].$$

Given ϵ, find the corresponding smoothness parameter L of $\phi_\epsilon(z)$. Given ϵ, use this smoothed hinge loss to write down an (L, ϵ)-smooth approximation of the SVM method

$$\frac{1}{n}\sum_{i=1}^n (1 - w^\top x_i y_i)_+ + \frac{\lambda}{2}\|w\|_2^2,$$

and find an estimate of the corresponding L using the eigenvalues of $\sum_i x_i x_i^\top$. Here $x_i \in \mathbb{R}^d$ and $y_i \in \{\pm 1\}$.

3. (3 points) In Nesterov's General Accleration Method (Algorithm 3 of Lecture 07), if we take $\eta_t = \eta$, $\gamma_0 = 1/\eta$ and $\lambda = 0$, show that there exists $c > 0$ such that for all $t \geq 1$,

$$\theta_t \leq c/t.$$

Find the smallest c which makes this equation holds. Find the best upper bound of the convergence rate λ_t using this estimate $\theta_t \leq c/t$.

[hint: please note the relationship that $\gamma_t = \theta_t^2/\eta$ and $\lambda_t = \theta_t^2$]

Programming Problem (6 points)

We consider optimization with the smoothed hinge loss $\phi_\gamma(x)$ of Problem 2, and randomly generated data. The goal is to implement gradient descent and experiment with different strong-convexity settings and different learning rates.

- Use the python template "prog_template.py", and implement functions marked with '# implement'.
- Submit your code and outputs. Compare to the theoretical convergence rates in class, and discuss your experimental results.