Weiqi Wang 20563565 wwangbw@connect.ust.hk
COMP5212: Machine Learning
October 6, 2022

COMP5212 Written Homework 1 Submission

Question 1.

(a) Prove $\langle \nabla^2 f(x)v, v\rangle \leq L||v||_2^2, \forall x, v \in R^d$

To prove this, we let $x_t = x + t \cdot v$, and let $g(t) = f(x_t)$. Then we have $g'(t) = \nabla f(x_t)^\top v$ for some $v \in R^d$. Since f is twice differentiable at x, we have that g is twice differentiable at 0 and therefore that for all $\epsilon > 0$ there exist some $\delta > 0$ such that

$$\left|\frac{g'(\delta) - g'(0)}{\delta} - g''(0)\right| \leq \epsilon$$

Note that $g'(t) = \nabla f(x_t)^\top v$, and $g''(t) = v^\top \nabla^2 f(x_t)^\top v$, thus we take the equations back, reverse the inequality, and take off the absolute value to get:

$$\frac{1}{\delta}[(\nabla f(x_\delta) - \nabla f(x_0))^\top v] \geq v^\top \nabla^2 f(x)^\top v - \epsilon$$

Recall that we have the $||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$. We can extract v out from the left part, and the left part has the inequality:

$$(\nabla f(x_\delta) - \nabla f(x_0)) \cdot v^\top \leq \delta \cdot [L \cdot ||v||_2^2 - \epsilon]$$

Which can be further connected as

$$[L \cdot ||v||_2^2 - \epsilon] \geq \frac{1}{\delta}[(\nabla f(x_\delta) - \nabla f(x_0))^\top v] \geq v^\top \nabla^2 f(x)^\top v - \epsilon$$

When the ϵ is approximated to 0, we can get the final inequality to prove:

$$L||v||_2^2 \geq v^\top \nabla^2 f(x)^\top v = \langle \nabla^2 f(x)v, v\rangle$$

(b) Prove $f(y) < f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2} \parallel y - x \parallel_2^2$

For all $x, y \in R^n$, we have:

$$f(y) = f(x) + \int_0^1 \langle f'(x + \epsilon(y - x)), y - x\rangle d\epsilon$$

$$= f(x) + \langle f'(x), y - x\rangle + \int_0^1 \langle f'(x + \epsilon(y - x)) - f'(x), y - x\rangle d\epsilon$$

Therefore, we can get:

$$|f(y) - f(x) - \langle f'(x), y - x\rangle = |\int_0^1 \langle f'(x + \epsilon(y - x)) - f'(x), y - x\rangle d\epsilon| \leq \int_0^1 |\langle f'(x + \epsilon(y - x)) - f'(x), y - x\rangle d\epsilon|$$

By the Cauchy–Schwarz inequality, we know that $|\langle u, v\rangle| \leq ||u|| \cdot ||v||$. Thus, we can get:

$$\int_0^1 |\langle f'(x + \epsilon(y - x)) - f'(x), y - x\rangle d\epsilon| \leq \int_0^1 ||f'(x + \epsilon(y - x)) - f'(x)|| \cdot ||y - x|| d\epsilon \leq \int_0^1 \epsilon L||y - x||^2 d\epsilon$$

by the definition of L-smooth.

Since

$$\int_0^1 \epsilon L||y - x||^2 d\epsilon = \frac{L}{2}||y - x||^2$$

We finally have $f(y) - f(x) - \langle f'(x), y - x\rangle \leq \frac{L}{2}||y - x||^2$, which is equivalent to $f(y) < f(x) + \langle \bigtriangledown f(x), y - x\rangle + \frac{L}{2} \parallel y - x \parallel_2^2$. Thus the inequality is proved.

Question 2.

(a) Prove $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\ell}{2}||y - x||^2$

From the question, we know that $\nabla^2 f(x) \geq \ell I$ with constant $\ell > 0$. By multiplying $z, z^\top$ at both ends, we can get $z^\top \nabla^2 f(x) z \geq \ell ||z||_2^2$. Then, we integrate the left part twice to make the second differential to first order.

$$f(y) - f(x) - f(x)^\top (y - x) = \int_0^1 \int_0^t (y - x)^\top \nabla^2 f(x_\alpha)(y - x) d\alpha dt$$

For the right hand side, by integrating the right hand side of $z^\top \nabla^2 f(x) z \geq \ell ||z||_2^2$, we can have:

$$\int_0^1 \int_0^t (y - x)^\top \nabla^2 f(x_\alpha)(y - x) d\alpha dt \geq \frac{\ell}{2}||y - x||_2^2$$

Thus, connecting them can get the final ineuqality to be proved:

$$f(y) - f(x) - \nabla f(x)^\top (y - x) \geq \frac{\ell}{2}||y - x||_2^2$$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\ell}{2}||y - x||_2^2$$

(b) Prove $f(y) \geq f(x) - \frac{1}{2\ell}||\nabla f(x)||^2$

From the inequality we proved in (a), we have $f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\ell}{2}||y - x||_2^2$. The right hand side $f(x) + \nabla f(x)^\top (y - x) + \frac{\ell}{2}||y - x||_2^2$ is convex quadratic in y and minimized at $y' = x - \frac{1}{\ell}\nabla f(x)$. Therefore, we can apply the above inequality to show that:

$$f(y) \geq f(x) + \nabla f(x)^\top (y' - x) + \frac{\ell}{2}||y' - x||_2^2$$

$$= f(x) + \nabla f(x)^\top (x - \frac{1}{\ell}\nabla f(x) - x) + \frac{\ell}{2}||x - \frac{1}{\ell}\nabla f(x) - x||_2^2$$

$$= f(x) - \frac{1}{\ell}\nabla f(x)^\top \nabla f(x) + \frac{\ell}{2} \cdot \frac{1}{\ell^2}||\nabla f(x)||_2^2$$

$$= f(x) - \frac{1}{\ell}||\nabla f(x)||_2^2 + \frac{1}{2\ell}||\nabla f(x)||_2^2$$

$$f(x) - \frac{1}{2\ell}||\nabla f(x)||_2^2$$

Thus we can finally get

$$f(y) \geq f(\arg\min_y f(y)) \geq f(x) - \frac{1}{2\ell}||\nabla f(x)||_2^2$$

(c) Prove $f(x^{t+1}) - f^* \leq (1 - \frac{\ell}{L})(f(x^t) - f^*)$

By the property of convex function in question 1, we know that

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}||y - x||_2^2$$

For $y = x^k - t\nabla f(x^k)$ and $x = x^k$, we have:

$$f(x^k - t\nabla f(x^k)) \leq f(x^k) + \nabla f(x^k)^\top (-t\nabla f(x^k)) + \frac{L}{2}|| - t\nabla f(x^k)||_2^2$$

$$= f(x^k) - t \cdot ||\nabla f(x^k)||_2^2 + \frac{L}{2}t^2||\nabla f(x^k)||_2^2$$

Particularly, with a step size of $t = \frac{1}{L}$, we have:

$$\begin{aligned} f(x^k - t\nabla f(x^k)) &\leq f(x^k) - \frac{1}{L}||\nabla f(x^k)||_2^2 + \frac{L}{2}\frac{1}{L^2}||\nabla f(x^k)||_2^2 \\ &= f(x^k) - \frac{1}{L}||\nabla f(x^k)||_2^2 + \frac{1}{2L}||\nabla f(x^k)||_2^2 \\ &= f(x^k) - \frac{1}{2L}||\nabla f(x^k)||_2^2 \end{aligned}$$

Notice that $x^{k+1} = x^k - t \cdot \nabla f(x^k)$, thus:

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L}||\nabla f(x^k)||_2^2$$

Subtracting $f^* = min_x f(x)$ from both sides can get us:

$$f(x^{k+1}) - f^* \leq f(x^k) - f^* - \frac{1}{2L}||\nabla f(x^k)||_2^2$$

Applying the inequality we derived in (b) $||\nabla f(x^k)||_2^2 \geq 2\ell(f(x^k) - f^*)$ can get us:

$$f(x^{k+1}) - f^* \leq f(x^k) - f^* - \frac{\ell}{L}(f(x^k) - f^*) = (1 - \frac{\ell}{L})(f(x^k) - f^*)$$

Thus, we replace k by t and successfully get the inequality:

$$f(x^{t+1}) - f^* \leq (1 - \frac{\ell}{L})(f(x^t) - f^*)$$

Question 3.

(a) Discuss why the objective function is non-differentiable.

When we are using the $L1$ norm, the loss function is not differentiable because of the regularization. Since we are using $\lambda||w||_1 := \sum_i |w_i|$ as the regularization, which is simply the sum of absolute values, it is not differentiable at the origin as absolute function has a peak at origin and the derivative at the left isn't equal to the right. Thus the loss is not differentiable because the regularization is not differentiable at the origin, otherwise it is differentiable.

(b) Derive the close form solution of w_{t+1} given w_t, $\nabla g(w_t)$, η, λ. What's the time complexity for one proximal gradient descent iteration?

The objective function is in the form of $min_w g(w) + h(w)$, and $g(w)$ is differentiable. We can rewrite it as

$$f(w) = g(w_t) + \nabla g(w_t)^\top (w - w_t) + \frac{\eta}{2}||w - w_t||^2 + \lambda||w||_1$$

Take the derivative with respect to w gets us:

$$\nabla f(w) = \nabla g(w_t) + \eta(w - w_t) + \lambda upd(w)$$

The upd function is related to w_i, if $w_i > 0$, then $upd(w_i) = 1$, if $w_i = 0$, then $upd(w_i) = 0$, for $w_i < 0$, $upd(w_i) = -1$.

Since our $f(w)$ function is a convex function, and the minimization is achieved at $\nabla f(x) = 0$, which can be further derived as

$$\eta w + \lambda upd(w) = \eta w_t - \nabla g(w_t)$$

Taking upd function back to this equation, and divide η at both sides, we get:

$$[w_{t+1}]_i = \begin{cases} [w_t - \frac{\nabla g(w_t) + \lambda}{\eta}]_i & \text{if } [\eta w_t - \nabla g(w_t)]_i > \lambda \\ 0 & \text{if } [\eta w_t - \nabla g(w_t)]_i \in [-\lambda, \lambda] \\ [w_t - \frac{\nabla g(w_t) - \lambda}{\eta}]_i & \text{if } [\eta w_t - \nabla g(w_t)]_i < \lambda \end{cases} \tag{1}$$

For the time complexity, we need to compute $\nabla g(w_t) = 2X^\top(Xw_t - y)$, the Xw_t takes $n \times d$ multiplications. Thus, the time complexity is bounded by $O(nd)$.