Optimization Methods in Machine Learning HW2

Mmesomachi Nwachukwu

November 2024

1 Theory Tasks

2 Consider Lemma 12 from the lecture notes: Let f be a differentiable, L-smooth function. Then

$$f(x) \leq \underbrace{f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2}_{g(x):=}$$

for all $x, y \in \mathbb{R}^d$. Find the optimal x that minimizes the upper bound $g(x)$.

Solution

$$\begin{aligned}
g(x) &= f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2 \\
&= \langle \nabla f(y), x \rangle + \frac{L}{2}\|x\|^2 - L\langle x, y \rangle + f(y) - \langle \nabla f(y), y \rangle + \frac{L}{2}\|y\|^2 \\
&= \langle \nabla f(y), x \rangle + \frac{L}{2}\langle x, x \rangle - L\langle x, y \rangle + f(y) - \langle \nabla f(y), y \rangle + \frac{L}{2}\|y\|^2 \\
\Rightarrow g(x) &\textbf{ is quadratic. This implies the function is convex} \\
\Rightarrow dg &= \langle \nabla f(y), dx \rangle + \frac{L}{2}\langle (I_n + I_n^T)x, dx \rangle - L\langle y, dx \rangle \\
&= \langle \nabla f(y), dx \rangle + \frac{L}{2}\langle 2x, dx \rangle - L\langle y, dx \rangle \\
&= \langle \nabla f(y), dx \rangle + L\langle x, dx \rangle - L\langle y, dx \rangle \\
&= \langle \nabla f(y), dx \rangle + L\langle x - y, dx \rangle \\
dg &= \langle \nabla f(y) + L(x - y), dx \rangle \\
\Rightarrow \nabla g(x) &= \nabla f(y) + L(x - y) \\
\Rightarrow x^* &= \frac{yL - \nabla f(y)}{L}
\end{aligned}$$

Hence, the optimal value of x that minimizes the upper bound of $g(x)$ is $\frac{yL - \nabla f(y)}{L}$

For convexity,

$$
\begin{aligned}
d^2g &= d(\langle \nabla f(y) + L(x-y), dx_1\rangle) = \langle d(\nabla f(y) + L(x-y)), dx_1\rangle \\
&= \langle LI_n dx, dx_1\rangle \\
\Rightarrow \nabla^2 g &= LI_n \succ 0
\end{aligned}
$$

optimal value,

$$x^* = \frac{yL - \nabla f(y)}{L}$$

3 Consider Theorem 22 and the corresponding proof from the lecture notes. How would the result change if, instead of L–smoothness (implies Lemma 12), the function f satisfies the inequality

$$f(x) \leq f(y) + \langle \nabla f(y), x-y\rangle + \frac{L}{2}\|x-y\|^2 + \delta$$

for all $x, y \in \mathbb{R}^d$ and some $\delta \geq 0$? (Theorem 22 is true when $\delta = 0$. What if $\delta > 0$?)

Solution

From the given inequality and GD update rule $x^{k+1} = x^k - \gamma \nabla f(x^k)$, we have

$$
\begin{aligned}
& f(x^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k\rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 + \delta \\
\Rightarrow\ & f(x^{k+1}) \leq f(x^k) - \gamma\|\nabla f(x^k)\|^2 + \frac{L\gamma^2}{2}\|\nabla f(x^k)\|^2 + \delta \\
\Rightarrow\ & f(x^{k+1}) \leq f(x^k) - (\gamma - \frac{L\gamma^2}{2})\|\nabla f(x^k)\|^2 + \delta
\end{aligned}
$$

For $0 < \gamma \leq \frac{1}{L}$, we have that $\gamma - \frac{L\gamma^2}{2} \leq \frac{\gamma}{2}$, hence,

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

Summing the inequality over $k = 0, 1, \dots, T-1$, we get,

$$f(x^T) \leq f(x^0) - \frac{\gamma}{2}\sum_{k=0}^{T-1}\|\nabla f(x^k)\|^2 + T\delta$$

. Let f^* denote the lower bound of f. Then,

$$f^* \leq f(x^T) \leq f(x^0) - \frac{\gamma}{2}\sum_{k=0}^{T-1}\|\nabla f(x^k)\|^2 + T\delta$$

This gives,

$$
\begin{aligned}
f^* &\leq f(x^0) - \frac{\gamma}{2} \sum_{k=0}^{T-1} \|\nabla f(x^k)\|^2 + T\delta \\
\Rightarrow \frac{\gamma}{2} \sum_{k=0}^{T-1} \|\nabla f(x^k)\|^2 &\leq f(x^0) - f^* + T\delta \\
\Rightarrow \frac{1}{T} \sum_{k=0}^{T-1} \|\nabla f(x^k)\|^2 &\leq \frac{2(f(x^0) - f^*)}{\gamma T} + \frac{2\delta}{\gamma}
\end{aligned}
$$