Review of Homework-2 (Group-4)

Group 2

1- Problem_1

a.

Therefore, for $\|\boldsymbol{w}\|_2 \leq D$, we have

$$\|\nabla f_i(\boldsymbol{w})\|_2 = \frac{|y_i| \|\boldsymbol{x}_i\|_2}{1+\exp\{y_i \boldsymbol{w}^T \boldsymbol{x}_i\}} \leq \frac{|y_i| \|\boldsymbol{x}_i\|_2}{1+\exp\{-D|y_i| \|\boldsymbol{x}_i\|_2\}} \triangleq B_{f_i}$$

For $r(\boldsymbol{w}) = \lambda \|\boldsymbol{w}\|_2^2$ and $\|\boldsymbol{w}\|_2 \leq D$, we have

$$\|\nabla r(\boldsymbol{w})\|_2 = 2\lambda \|\boldsymbol{w}\|_2 \leq 2\lambda D \triangleq B_r$$

This inequality is correct but It's better to prove the second term from the first equation that is grater that the norm of gradient. You can use Cauchy-Schwartz inequality.

b.

$$\begin{aligned}
\nabla^2 f(\boldsymbol{w}) = J(\nabla f(\boldsymbol{x})) &= \frac{1}{N} \sum_{i\in[N]} \boldsymbol{x}_i \boldsymbol{x}_i^T \frac{y_i^2 \exp\{y_i \boldsymbol{w}^T \boldsymbol{x}_i\}}{(1+\exp\{y_i \boldsymbol{w}^T \boldsymbol{x}_i\})^2} + 2\lambda I && \text{(5a)}\\
&\preceq \frac{1}{4N} \sum_{i\in[N]} \boldsymbol{x}_i \boldsymbol{x}_i^T y_i^2 + 2\lambda I && \text{(5b)}\\
&= \frac{1}{4N} AA^T + 2\lambda I && \text{(5c)}\\
&\preceq \sigma_{\max}\left(\frac{1}{4N} AA^T + 2\lambda I\right) I && \text{(5d)}\\
&= \frac{1}{4N}(\sigma_{\max}(A^T A) + 2\lambda) I, && \text{(5e)}
\end{aligned}$$

where $A \triangleq [y_1 \boldsymbol{x}_1, y_2 \boldsymbol{x}_2, \ldots, y_N \boldsymbol{x}_N]$. In (5e), we used the facts that $\operatorname{eig}(X+\lambda I) = \operatorname{eig}(X) + \lambda$ and $\operatorname{eig}(AB) = \operatorname{eig}(BA)$. Hence, f is L-smooth with

$$L \leq \frac{1}{4N} \sigma_{\max}(A^T A) + 2\lambda.$$

If the objective function is convex and L-smooth, the hessian of the function is less than LI, so the last equation in your solution is not correct. Therefor it's better to mention that find a lower bound for L.

c.

Correct and precise.

2- Problem_2

a.

It's correct and same our solution.

3- Problem_3

a.

Correct.