HW 2(a):

1- Lipschitz continuity: $\|w\|_2 \leq D \Rightarrow \|\nabla f(w)\|_2 \leq B$

$$f(w) = \frac{1}{N} \sum_{i \in [N]} f_i(w) + \lambda \|w\|_2^2 \longrightarrow \nabla f(w) = \frac{1}{N} \sum \nabla f_i(w) + 2\lambda w$$

$$= \frac{1}{N} \sum_i \frac{e^{-y_i w^T x_i}}{1 + e^{-y_i w^T x_i}} (-y_i x_i) + 2\lambda w$$

$$\Rightarrow \|\nabla f(w)\|_2 \leq \frac{1}{N} \sum_i \underbrace{\left| \frac{e^{-y_i w^T x_i}}{1 + e^{-y_i w^T x_i}} \right|}_{\leq 1} |y_i| \|x_i\|_2 + 2\lambda \|w\|_2$$

$$\Rightarrow \|\nabla f(w)\|_2 \leq \frac{1}{N} \sum_i |y_i| \|x_i\|_2 + 2\lambda \|w\|_2 \leq \underline{\frac{1}{N} \sum |y_i| \|x_i\|_2 + 2\lambda D < \infty}$$

By using Human Activity Recognition Using smartphones dataset, we have: $\underline{B = 66.7}$

2-

$$\nabla^2 f_i(w) = \underbrace{\frac{+ e^{-y_i w^T x_i} y_i^2}{(1 + e^{-y_i w^T x_i})^2}}_{\leq 1/4} x_i x_i^T \Rightarrow \nabla^2 f_i(w) \leq \underbrace{\frac{1}{4} |y_i|^2 x_i x_i^T}_{L \text{ for } f_i(w)}$$

$$\nabla^2 f(w) = \frac{1}{N} \sum_i \underbrace{\frac{e^{-y_i w^T x_i} y_i^2}{(1 + e^{-y_i w^T x_i})^2}}_{\leq 1/4} x_i x_i^T + 2\lambda I$$

$$\Rightarrow \nabla^2 f(w) \leq \underbrace{\frac{1}{4N} \sum_i |y_i|^2 x_i x_i^T + 2\lambda I}_{L \text{ for } f(w)}$$

Again, by using the dataset:

$\underline{L = 5948.37}$

HW 2 (a)

3 – $\nabla_w^2 f_i(w) = y_i^2 \dfrac{e^{-y_i w^T x_i}}{(1+e^{-y_i w^T x_i})^2} x_i x_i^T$

$\nabla^2 f_i(w)$ is always positive. By defining $w = k y_i x_i$, if $k \to \infty$, then $\nabla^2 f_i \to 0$. so $f_i(w)$ is convex.

$\nabla^2 f(w) = \frac{1}{N} \sum \nabla^2 f_i(w) + 2\lambda I \xrightarrow[\text{\& in } \infty \text{ we have}]{\text{again with the same reasoning,}} \nabla^2 f(w) \succcurlyeq 2\lambda I$

$\Rightarrow$ $f(w)$ is strongly convex with $\mu = 2\lambda$

HW2

b)

$$F(w_2) \le f(w_1) + \nabla f(w_1)^T (w_2 - w_1) + \frac{1}{2} L \|w_2 - w_1\|_2^2$$

$$\nabla f(w_k)^T E_{\xi_k}[g(w_k;\xi_k)] \geq c\|\nabla f(w_k)\|_2^2 \quad , \quad \|E_{\xi_k}[g(w_k;\xi_k)]\|_2 \le c_0 \|\nabla f(w_k)\|$$

$$Var_{\xi_k}[g(w_k;\xi_k)] := \underbrace{E_{\xi_k}\|g(w_k;\xi_k)\|} - \underbrace{\|E_{\xi_k}[g(w_k,\xi_k)]\|_2^2}$$

$$V_{\xi_k}[g(w_k;\xi_k)] \le M + M_G \|\nabla f(w_k)\|_2^2$$

$$E_{\xi_k}\|g(w_k;\xi_k)\|_2^2 - \|E_{\xi_k}[g(w_k;\xi_k)]\|_2^2 \le M + M_G \|\nabla f(w_k)\|^2$$

$$\|E_{\xi_k}[g(w_k;\xi_k)]\|^2 \le c_0 \|\nabla f(w_k)\|_2$$

$$\Rightarrow E_{\xi_k}\|g(w_k;\xi_k)\|_2^2 \le M + M_G\|\nabla f(w_k)\|_2^2 + c_0^2 \|\nabla f(w_k)\|_2^2 \quad \Rightarrow \alpha = M, \; \beta = M_G + c_0^2 \geq c^2 > 0$$

c) $\sum \alpha_k \to \infty$ define $A_K = \sum_K \alpha_K$

$\sum \alpha_k^2 < \infty \Rightarrow \{\alpha_k\} \to 0$

$$E_{\xi_k}[F(w_{k+1}) - f(w_k)] \le -\mu\alpha_k \|\nabla f(w_k)\|_2^2 + \frac{1}{2}\alpha_k^2 L E_{\xi_k}\|g(w_k;\xi_k)\|_2^2]$$

$$\le -(\mu - \frac{1}{2}\alpha_k L M_G)\alpha_k \|\nabla f(w_k)\|_2^2 + \frac{1}{2}\alpha_k^2 LM$$

$$\le -\frac{1}{2}\mu\alpha_k E[\|\nabla f(w_k)\|_2^2] + \frac{1}{2}\alpha_k^2 LM$$

sum over $k \in [K]$

$$f_{inf} - E[f(w_1)] \le E[F(w_{k+1}) - f(w_1)] \le -\frac{1}{2}\mu \sum_k \alpha_k E\|\nabla f(w_k)\|_2^2] + \frac{1}{2}\sum_k \alpha_k^2 LM$$

Dividing by $\frac{\mu}{2}$: C

$$\sum_{k=1}^{K} \alpha_k E[\|\nabla f(w_k)\|_2^2] \le \frac{2E[f(w_1) - f_{inf}]}{\mu} + \frac{LM}{\mu}\sum_{k=1}^{K} \alpha_k^2$$

since $\sum \alpha_k^2 < \infty \Rightarrow C < \infty$

sin $\sum \alpha_k \to \infty$ $A_k \to \infty$