EP3260: Machine Learning Over Networks
Homework Assignment 2
Instructors: Hossein S. Ghadikolaei, José Mairton B. da Silva Jr., Carlo Fischione
Due Date: February 26, 2020

Problem 2.1

Consider Human Activity Recognition Using Smartphones dataset $\{(\mathbf{x}_i, y_i)\}_{i\in[N]}$, with inputs defined as the accelerometer and gyroscope sensors, and outputs defined as moving (e.g., walking , running, dancing) or not (sitting or standing). Consider the logistic ridge regression loss function

$$\underset{\mathbf{w}}{\text{minimize}} \quad f(\mathbf{w}) = \frac{1}{N}\sum_{i\in[N]} f_i(\mathbf{w}) + \lambda\|\mathbf{w}\|_2^2,$$

where $f_i(\mathbf{w}) = \log\left(1 + \exp\{-y_i \mathbf{w}^T \mathbf{x}_i\}\right)$.

Then, address the following questions:

(a) Is f Lipschitz continuous? If so, find a small B?

(b) Is f_i smooth? If so, find a small L for f_i? What about f?

(c) Is f strongly convex? If so, find a high μ?

Problem 2.2

Let us assume that there exist scalars $c_0 \geq c > 0$ such that for all $k \in \mathbb{N}$

$$\nabla f(\mathbf{w}_k)^T \mathbb{E}_{\zeta_k}\left[g(\mathbf{w}_k;\zeta_k)\right] \geq c\|\nabla f(\mathbf{w}_k)\|_2^2, \tag{1a}$$

$$\|\mathbb{E}_{\zeta_k}\left[g(\mathbf{w}_k;\zeta_k)\right]\|_2 \leq c_0\|\nabla f(\mathbf{w}_k)\|_2. \tag{1b}$$

Furthermore, let us assume that there exist scalars $M \geq 0$ and $M_V \geq 0$ such that for all $k \in \mathbb{N}$

$$\mathrm{Var}_{\zeta_k}\left[g(\mathbf{w}_k;\zeta_k)\right] \leq M + M_V\|\nabla f(\mathbf{w}_k)\|_2^2. \tag{2}$$

For the convergence proof of SGD with an L-smooth convex objective function (see slides), prove that

$$\mathbb{E}_{\zeta_k}\left[\|g(\mathbf{w}_k;\zeta_k)\|_2^2\right] \leq \alpha + \beta\|\nabla f(\mathbf{w}_k)\|_2^2.$$

Problem 2.3

For the SGD with non-convex objective function, prove that with square summable but not summable step-size, we have for any $K \in \mathbb{N}$

$$\mathbb{E}\left[\sum_{k\in[K]} \alpha_k \|\nabla f(\mathbf{w}_k)\|_2^2\right] < \infty \tag{3}$$

and therefore

$$\mathbb{E}\left[\frac{1}{\sum_{k\in[K]} a_k} \sum_{k\in[K]} \alpha_k \|\nabla f(\mathbf{w}_k)\|_2^2\right] \xrightarrow{K\to\infty} 0 \tag{4}$$