This homework is due at **2pm, October 18**.

Problem 1

Recall that the logistic regression corresponds to the optimization problem

$$w^\star \in \operatorname*{argmin}_w \left\{ f(w) \,\middle|\, w \in \mathbb{R}^p \right\},$$

where

$$f(w) := \frac{1}{n} \sum_{i=1}^n \log\left(1 + \mathrm{e}^{-y_i \langle x_i, w \rangle}\right),$$

for given data $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \{-1, +1\}$.

1. (10 points) Suppose that the data satisfies

$$\mathsf{P}\left(y_i = 1\right) = 1 - \mathsf{P}\left(y_i = -1\right) = \frac{1}{1 + \mathrm{e}^{-\langle x_i, w^\natural \rangle}}, \quad \forall 1 \le i \le n,$$

for some $w^\natural \in \mathbb{R}^p$, and that $y_1, \ldots, y_n$ are independent random variables. **Show that $w^\star$ is a maximum-likelihood estimator of $w^\natural$.**

2. (10 points) Let $X \in \mathbb{R}^{n \times p}$, whose i-th row is given by x_i. Denote by $\mathsf{var}(y_i)$ the variance of y_i for every i. **Show that**

$$\nabla^2 f(w) = \frac{1}{4n} X^{\mathrm{T}} D X,$$

where $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix given by

$$D := \begin{bmatrix} \mathsf{var}(y_1) & 0 & \cdots & 0 \\ 0 & \mathsf{var}(y_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathsf{var}(y_n) \end{bmatrix}.$$

3. (10 points) **Show that $\mathsf{var}(y_i) \le 1$ for all $1 \le i \le n$.**

4. (10 points) **Show that the function f is L-smooth for some $L > 0$. Specify the corresponding parameter L.**

   (HINT: The parameter L may depend on, for example, the matrix X.)
   (HINT: You do not need to find the best value of L.)

5. (10 points) **Give an algorithm that, for any given $\varepsilon > 0$, computes some w_ε such that**

$$f(w_\varepsilon) \le f(w^\star) + \varepsilon$$

with $O(\varepsilon^{-1/2})$ calls to the first-order oracle.

Problem 2

Let $\mathscr{X} \subseteq \mathbb{R}^p$ be a closed convex set. Let $f : \mathscr{X} \to \mathbb{R}$ be differentiable, L-Lipschitz, and μ-strongly convex with respect to the ℓ_2-norm, for some strictly positive real numbers L and μ. Consider solving the optimization problem

$$f^\star = \min_x \left\{ f(x) \,\middle|\, x \in \mathscr{X} \right\}.$$

Problem 2 continued on the next page…

via the following algorithm.

Set $x_1 \in \mathcal{X}$ and $T \in \mathbb{N}$.
for $t = 2, \dots, T$ **do**
  $x_t \leftarrow \operatorname{proj}_{\mathcal{X}}\left(x_{t-1} - \eta_t \nabla f(x_{t-1})\right)$
end for

Of course, the step sizes η_t should be properly chosen. The projection is defined with respect to the ℓ_2-norm, i.e.,

$$\operatorname{proj}_{\mathcal{X}}(y) := \operatorname*{arg\,min}_x \left\{ \|y - x\|_2 \;\middle|\; x \in \mathcal{X} \right\}.$$

1. (10 points) **Show that**

$$2\left(f(x_t) - f(x^\star)\right) \le 2\left\langle \nabla f(x_t), x_t - x^\star \right\rangle - \mu \|x_t - x^\star\|_2^2, \quad \forall t \in \mathbb{N}.$$

2. (10 points) **Show that projection onto $\mathcal{X}$ is a non-expansive mapping.** That is, show that

$$\|\operatorname{proj}_{\mathcal{X}}(y) - \operatorname{proj}_{\mathcal{X}}(x)\|_2 \le \|y - x\|_2, \quad \forall x, y \in \mathbb{R}^p.$$

3. (10 points) **Show that**

$$2\left\langle \nabla f(x_t), x_t - x^\star \right\rangle \le \frac{\|x_t - x^\star\|_2^2 - \|x_{t+1} - x^\star\|_2^2}{\eta_{t+1}} + \eta_{t+1} L^2, \quad \forall t \in \mathbb{N}.$$

4. (20 points) Set

$$\eta_{t+1} := \frac{1}{\mu t}, \quad \forall t \in \mathbb{N}.$$

Show that

$$\min\left\{ f(x_1), \dots, f(x_t) \right\} - f^\star \le \frac{L^2 (1 + \log t)}{2\mu t}, \quad \forall t \in \mathbb{N}.$$