LAB 7: LOGISTIC REGRESSION AND STOCHASTIC METHODS

Exercise 1: Logistic regression. We study logistic regression from a probabilistic point of view. Define the sigmoid function $\sigma\colon \mathbb{R} \to \mathbb{R}$ by

$$\sigma\colon x \mapsto \frac{1}{1+e^{-x}}\,.$$

We consider a pair of random variables $(X, Y) \in \mathbb{R}^d \times \{-1, +1\}$, assume their conditional distribution is given by

$$\mathbb{P}_\theta(Y = +1|X) = \sigma\big(\theta^T X\big)\,,$$

for some parameters $\theta \in \mathbb{R}^d$.

(1) Sigmoid properties: show that σ is always in $[0, 1]$, $\sigma' = \sigma(1-\sigma)$ and make a drawing (approximate).

(2) Log-Likelihood function: given a sample of size n $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$ independent and identically distributed (i.i.d.) with conditional laws given by

$$\mathbb{P}_\theta(Y_i = +1|X_i) = \sigma\big(\theta^T X_i\big) \quad \text{for } i = 1, \dots, n\,,$$

the likelihood of θ is defined by

$$L_n(\theta) = \mathbb{P}_\theta(Y_1, \dots, Y_n|X_1, \dots, X_n)\,.$$

Show that

$$\log L_n(\theta) = -\sum_{i=1}^{n} \log(1 + e^{-Y_i X_i^T \theta})\,.$$

(3) Decision boundary: show that if $\theta^T x > 0$, then

$$\mathbb{P}_\theta(Y = +1|X = x) > \mathbb{P}_\theta(Y = -1|X = x)$$

Exercise 2: Properties of the logistic regression problem. Consider the logistic regression problem

$$\min_{\theta\in\mathbb{R}^d} f(\theta) = \frac{1}{n}\sum_{i=1}^{n} f_i(\theta) \quad \text{where} \quad f_i(\theta) = \log\Big(1 + e^{-y_i x_i^T \theta}\Big)\,,$$

with $x_1, \dots, x_n \in \mathbb{R}^d$, $y_1, \dots, y_n \in \{-1, +1\}$.

(1) Show that $t \mapsto \log(1 + e^t)$ is convex and that, as a consequence, f is convex.

(2) Define

$$L = \frac{1}{4n}\sum_{i=1}^{n} \|x_i\|_2^2\,.$$

Show that f is L-smooth, i.e., that $\lambda_{\max}(\nabla^2 f)$ is always bounded by L.
Hint: start by noticing that $t/(1+t)^2 \le 1/4$ for any $t \in \mathbb{R}$ and begin by bounding $\lambda_{\max}(\nabla^2 f_i)$ for $i = 1, \dots, n$.

(3) Show that f is not strongly convex.
Hint: study the behaviour of $x_1^T \nabla^2 f(tx_1)x_1$ as $t \to +\infty$.

Exercise 3: Stochastic Gradient Method. Consider the minimization problem

$$\min_{x \in \mathbb{R}^d} f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x)\,,$$

where $f_1, \ldots, f_n$ are differentiable convex functions with gradients bounded by

$$\|\nabla f_i\| \leq G\,.$$

Let x^* be a minimizer.

The stochastic gradient method, started at x_1, is defined by the recursion

$$x_{t+1} = x_t - \gamma \nabla f_{i_{t+1}}(x_t) \quad t = 1, 2, \ldots\,,$$

where i_{t+1} is sampled uniformly from $\{1, \ldots, n\}$ independently from i_s for $s \leq t$.

(1) Show that, for $t \geq 2$,

$$\begin{aligned} &\mathbb{E}[\|x_t - x^*\|_2^2] \\ &= \mathbb{E}[\|x_{t-1} - x^*\|_2^2] - 2\gamma\, \mathbb{E}[\nabla f_{i_t}(x_{t-1})^T (x_{t-1} - x^*)] + \gamma^2\, \mathbb{E}[\|\nabla f_{i_t}(x_{t-1})\|_2^2]\,. \end{aligned}$$

(2) Prove that

$$\mathbb{E}[\nabla f_{i_t}(x_{t-1})^T (x_{t-1} - x^*)] = \mathbb{E}[\nabla f(x_{t-1})^T (x_{t-1} - x^*)]$$

(3) Using the convexity of f, get that

$$\gamma\, \mathbb{E}[f(x_{t-1}) - f(x^*)] \leq \frac{1}{2}\big(\mathbb{E}[\|x_{t-1} - x^*\|_2^2] - \mathbb{E}[\|x_t - x^*\|_2^2] + \gamma^2 G^2\,.\big)$$

(4) By summing the equation above along iterations and invoking the convexity of f again, show that we have

$$\mathbb{E}\left[f\left(\frac{1}{T}\sum_{t=1}^{T} x_t\right)\right] - f(x^*) \leq \frac{\|x_1 - x^*\|_2^2}{2\gamma T} + \frac{\gamma G^2}{2}\,,$$

and find the optimal $\gamma > 0$ that minimizes this bound.