COMS 4995-004: Optimization for Machine Learning
Homework 1.

HW1 is due Thursday, Sept 26 by 1:00 pm. No late assignments will be accepted[1]. Please refer to `https://www.satyenkale.com/optml-f19/` for instructions on how to submit homework assignments.

Question 1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable convex function. In this question, we will prove that $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^d$. We will prove this by showing that for all vectors $u \in \mathbb{R}^d$, we have $u^\top \nabla^2 f(x) u \geq 0$.

(a) **(2 points)** Let $x, y \in \mathbb{R}^d$. Prove that

$$\int_{t=0}^{1} (1-t) \frac{\partial^2 f(x+ty)}{\partial^2 t} dt = f(x+y) - f(x) - \nabla f(x)^\top y.$$

(Hint: think about integration by parts.)

(b) **(2 points)** Prove that

$$\frac{\partial^2 f(x+ty)}{\partial^2 t} = y^\top \nabla^2 f(x+ty) y.$$

(c) **(3 points)** Set $y = \alpha u$, for some $\alpha \in \mathbb{R}$. Using the convexity of f and parts (a) and (b), show that there exists a $t' \in [0,1]$ such that $u^\top \nabla^2 f(x + t'\alpha u) u \geq 0$. *(Hint: use the mean-value theorem on the integral in part (a).)*

(d) **(2 points)** Show that part (c) implies that $u^\top \nabla^2 f(x) u \geq 0$.

Question 2. Consider the following training set: $S = \{(x_i, y_i) \in \mathbb{R}^3 \times \mathbb{R} \mid i = 1, 2, 3\}$, where

$$\begin{aligned} (x_1, y_1) &= ((2,0,0), 1) \\ (x_2, y_2) &= ((0,1,0), -1) \\ (x_3, y_3) &= ((0,0,0.5), 1). \end{aligned}$$

Suppose we want to train a linear predictor $f_w = \langle w, x \rangle$ for some weight vector $w \in \mathbb{R}^3$. Consider training the predictor using the following three loss functions and regularization functions:

(i) (Square loss with no regularization) loss function $\ell(\hat{y}, y) = (\hat{y} - y)^2$, no regularization.

(ii) (Square loss with ℓ_1 regularization) loss function $\ell(\hat{y}, y) = (\hat{y} - y)^2$, regularization $R(w) = \|w\|_1$, regularization constant $\lambda = 1$.

[1] Unless you have an emergency; in that case please write to Satyen as soon as possible.

(iii) (Logistic loss with no regularization) loss function $\ell(\hat{y}, y) = \log(1 + \exp(-\hat{y}y))$, no regularization.

(iii) (Logistic loss with ℓ_2 regularization) loss function $\ell(\hat{y}, y) = \log(1 + \exp(-\hat{y}y))$, regularization $R(w) = \frac{1}{2}\|w\|_2^2$, regularization constant $\lambda = 1$.

For training loss function in each of the above cases, answer the following questions:

1. **(2 points per function)** Give formulas for the gradient (or subgradient, if the function is not differentiable) and Hessian (if it exists) as a function of w.

2. **(1 point per function)** Is the function strongly convex? If yes, compute a lower bound on the strong convexity constant μ. Try to make it as tight as possible.

3. **(1 point per function)** Is the function smooth? If yes, compute an upper bound on the smoothness constant β. Try to make it as tight as possible. *(Note: smoothness wasn't defined as of the class of Sept 12; we will cover it on the class of Sept 17.)*