New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6923: (Mini) Written Homework 3.
Due Wednesday, March 12, 2025, 11:59pm.
NO SLIP DAY FOR THIS HOMEWORK.

Collaboration is allowed on this problem set, but solutions must be written-up individually.

Problem 1: Convexity (10pts)

(a) Prove that if functions $f(\boldsymbol{\beta})$ and $g(\boldsymbol{\beta})$ are both convex, then $f(\boldsymbol{\beta}) + g(\boldsymbol{\beta})$ is also convex.

(b) Prove that, for $\lambda \geq 0$, the functions $\lambda\|\boldsymbol{\beta}\|_2^2$ and $\lambda\|\boldsymbol{\beta}\|_1$ are both convex. Combined with the claim from class that $\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_2^2$ is convex, conclude that the ℓ_2 and ℓ_1 regularized regression objectives are both convex, and thus we can find a global minimum of these loss functions using gradient descent.

Problem 2. Steepest Descent (10pts)

Recall from Lecture 6 that gradient descent is often considered a "steepest descent" method because the search direction, $\frac{\nabla L(\boldsymbol{\beta})}{\|\nabla L(\boldsymbol{\beta})\|_2^2}$, solves the maximization problem:

$$\frac{\nabla L(\boldsymbol{\beta})}{\|\nabla L(\boldsymbol{\beta})\|_2^2} = \operatorname*{arg\,max}_{\mathbf{v}:\|\mathbf{v}\|_2=1} \langle \nabla L(\boldsymbol{\beta}), \mathbf{v} \rangle.$$

In other words, if we had to choose a vector $\mathbf{v}$ of fixed Euclidean norm to maximize the decrease in objective values from $L(\boldsymbol{\beta})$ to $L(\boldsymbol{\beta} - \eta\mathbf{v})$ as the step-size $\eta \to 0$, we should choose $\mathbf{v} = \frac{\nabla L(\boldsymbol{\beta})}{\|\nabla L(\boldsymbol{\beta})\|_2^2}$.

If we replace the Euclidean norm with a different norm, we obtain different variants of "steepest descent".

(a) What steepest search direction should we choose to solve $\max_{\mathbf{v}:\|\mathbf{v}\|_1=1} \langle \nabla L(\boldsymbol{\beta}), \mathbf{v} \rangle$? Justify your answer.

(b) Suppose we implement steepest descent with update rule $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta} - \eta\mathbf{v}^*$, where $\mathbf{v}^*$ is the alternative steepest descent directions derived in Part(a). Prove that $\lim_{\eta \to 0} \frac{L(\boldsymbol{\beta} - \eta\mathbf{v}^*) - L(\boldsymbol{\beta})}{\eta}$ is negative. In other words, like gradient descent, for small learning rates, this choice always decreases the objective value.

A consequence of Part (b) is that any of this steepest descent methods will provably converge to a stationary point for sufficiently small learning rate, η.