Optimization and Computational Linear Algebra for Data Science
Homework 10: Regression

Due on November 26, 2019

- Unless otherwise stated, all answers must be mathematically justified.
- Partial answers will be graded.
- You can work in groups but each student must write his/her own solution based on his/her own understanding of the problem. Please list on your submission the students you work with for the homework (this will not affect your grade).
- Problems with a ($\star$) are extra credit, they will not (directly) contribute to your score of this homework. However, for every 4 extra credit questions successfully answered your lowest homework score get replaced by a perfect score.
- If you have any questions, feel free to contact me (`lm4271@nyu.edu`) or to stop at the office hours.

DRAFT

Problem 10.1 (2 points)**.** *Let $A \in \mathbb{R}^{n\times m}$. Show that if A has linearly independent columns, then $A^{\dagger} = (A^{\mathsf{T}} A)^{-1} A^{\mathsf{T}}$.*

Problem 10.2 (2 points)**.** *Let $A \in \mathbb{R}^{n\times d}$ and $y \in \mathbb{R}^n$. The Ridge regression adds a ℓ_2 penalty to the least square problem in order to "select" a solution of smaller norm. The Ridge regression problem is then:*

$$\text{minimize} \qquad \|Ax - y\|^2 + \lambda \|x\|^2 \qquad \text{with respect to } x \in \mathbb{R}^d, \tag{1}$$

for some penalization parameter $\lambda > 0$. Show that (1) admits a unique solution given by

$$x^{\text{Ridge}} = (A^{\mathsf{T}} A + \lambda \mathrm{Id})^{-1} A^{\mathsf{T}} y.$$

Problem 10.3 (3 points).

Problem 10.4 (3 points).

Problem 10.5 ($\star$).