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.

Problem 10.1 (2 points). *Let $A \in \mathbb{R}^{n \times m}$ and $y \in \mathbb{R}^n$. We consider the least square problem:*

$$\text{minimize} \quad \|Ax - y\|^2 \quad \text{with respect to } x \in \mathbb{R}^m. \tag{1}$$

We know from the lecture that $x^{\mathrm{LS}} \stackrel{\mathrm{def}}{=} A^{\dagger} y$ is a solution of (1).

(a) *Show that $x^{\mathrm{LS}} \perp \mathrm{Ker}(A)$.*

(b) *Deduce that x^{LS} is the solution of (1) that has the smallest (Euclidean) norm.*

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:*

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

for some penalization parameter $\lambda > 0$. Show that (2) admits a unique solution given by

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

Problem 10.3 (3 points). *Recall that $\|M\|_{\mathrm{Sp}}$ denotes the spectral norm of a matrix M.*

(a) *Let $A \in \mathbb{R}^{n \times m}$. Show that for all $x \in \mathbb{R}^m$,*

$$\|Ax\| \leq \|A\|_{\mathrm{Sp}} \|x\|.$$

(b) *Show that for all $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times k}$:*

$$\|AB\|_{\mathrm{Sp}} \leq \|A\|_{\mathrm{Sp}} \|B\|_{\mathrm{Sp}}.$$

(c) *Is it true that for all $n, m, k \geq 1$, all $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times k}$:*

$$\|AB\|_F \leq \|A\|_F \|B\|_F \ ?$$

Give a proof or a counter-example.

Problem 10.4 (3 points). *Consider the 5×4 matrix A and $y \in \mathbb{R}^5$ given by:*

$$A = \begin{pmatrix} 1.1 & -2.3 & 1.7 & 4.5 \\ 1.7 & 1.6 & 3.8 & 0.3 \\ 1 & 0.1 & 1.3 & 0.2 \\ -0.5 & -0.4 & 0 & -1.3 \\ -0.5 & 2.9 & -0.3 & 2 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} -13.8 \\ -2.7 \\ 9.6 \\ -2.4 \\ 3.9 \end{pmatrix}.$$

In each of the following questions, it is intended that you solve the problem using the programming language of your choice and only report the numerical answer to 3 decimal places, without including your code files in your submission.

(**a**) *Compute the minimizer $x^* \in \mathbb{R}^4$ of*

$$\|Ax - y\|.$$

(**b**) *Find a vector $v \in \mathbb{R}^5$ with $v_1 > 0$ and $\|v\| = 1$ such that the minimizer of*

$$\|Ax - (y + v)\|$$

is also x^*.

(**c**) *Find a vector $w \in \mathbb{R}^5$ with $\|w\| = 1$ such that the minimizer x' of*

$$\|Ax - (y + w)\|$$

maximizes the error $\|x^* - x'\|$ and also give the resulting error. That is, we are trying to corrupt the vector y with a fixed amount of noise w that maximally modifies the least squares solution.

Problem 10.5 ($\star$). *Let $A \in \mathbb{R}^{n \times n}$ be an orthogonal matrix, and $y \in \mathbb{R}^n$. We fix $\alpha, \lambda > 0$ and consider the so-called "elastic net" problem:*

$$\text{minimize} \quad \frac{1}{2}\|Ax - y\|^2 + \frac{\alpha}{2}\|x\|^2 + \lambda\|x\|_1 \quad \text{with respect to } x \in \mathbb{R}^n. \tag{3}$$

Give the expression of the solution x^* of (3) in term of A, y, λ and α.