Rules:

- Unless otherwise stated, all answers must be mathematically justified.
- Partial answers will be graded.
- Your submission has to be uploaded to Gradescope. In Gradescope, indicate the page on which each problem is written.
- You can work in groups but each student must write his/her/their own solution based on his/her/their 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 (`mgabrie@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 $x^{\mathrm{LS}} \stackrel{\text{def}}{=} A^\dagger y$ 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 $\lambda > 0$.

(a) *Without solving (2), show that (2) admits a unique solution. You can use HW9 results but you must justify everything precisely.*

(b) *Show that this solution is given by*

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

Justify your answer precisely, including why $(A^\mathsf{T} A + \lambda \mathrm{Id}_d)^{-1}$ exists.

Problem 10.3 (2 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 $B \in \mathbb{R}^{m\times k}$:*

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

Problem 10.4 (3 points)**.** *In this problem we will investigate the role of ridge regularization in polynomial regression. This problem can be solved using our linear regression tools using a trick described in the notebook* `polyreg.ipynb`*, which also contains instructions and questions. Please merge a pdf version of the notebook completed to your submission.*

Problem 10.5 ($\star$)**.** *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.