Optimization and Computational Linear Algebra for Data Science
Homework 8: SVD, linear algebra and graphs

Due on November 5, 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 8.1 (2 points). *For any two matrices $A, B \in \mathbb{R}^{n\times m}$ we define*

$$\langle A, B\rangle_F = \mathrm{Tr}(A^\mathsf{T} B).$$

(a) *Show that $\langle \cdot,\cdot\rangle_F$ is an inner-product on $\mathbb{R}^{n\times m}$, i.e. that it verifies the points of the definition 2.1 of Lecture 4. $\langle \cdot,\cdot\rangle_F$ is called the* Frobenius inner-product.

(b) *The induced norm $\|A\|_F = \sqrt{\mathrm{Tr}(A^\mathsf{T} A)}$ is called the* Frobenius norm. *Show that*

$$\|A\|_F = \sqrt{\sum_{i=1}^{\min(n,m)} \sigma_i^2},$$

where $\sigma_1, \dots, \sigma_{\min(n,m)}$ are the singular values of A.

Problem 8.2 (2 points). *Let A be a $n \times n$ matrix.*

(a) *Show that A is invertible if and only if all the singular values of A are non-zero.*

(b) *We assume here that A is invertible. Show that*

$$\sigma_1(A)\sigma_1(A^{-1}) \geq 1,$$

where $\sigma_1(A)$ and $\sigma_1(A^{-1})$ denote the largest singular value of respectively A and A^{-1}.

Problem 8.3 (2 points). *Let $A \in \mathbb{R}^{n\times n}$ be the adjacency matrix of a graph G. We define a path from a node i_1 to a node i_k as a succession of nodes $i_1, i_2, \dots, i_k$ such that*

$$i_1 \sim i_2 \sim \cdots \sim i_{k-1} \sim i_k, \quad i.e. \quad A_{i_1,i_2} = A_{i_2,i_3} = \cdots = A_{i_{k-1},i_k} = 1.$$

The nodes i_j of the path do not need to be distinct. We say that the path $i_1, \dots, i_k$ has length $k-1$ which is the number of edges in this path. The goal of this exercise is to prove that for all $k \geq 1$

$$\mathcal{H}(k) : \text{« For all } i, j \in \{1, \dots, n\}, \text{ the number of paths of length } k \text{ from } i \text{ to } j \text{ is } (A^k)_{i,j} \text{ ».}$$

We will prove that $\mathcal{H}(k)$ holds for all k by induction, that is, we will first prove that $\mathcal{H}(1)$ is true. Then we will prove that if $\mathcal{H}(k)$ is true for some k, then $\mathcal{H}(k+1)$ is true. Combining these two things, we get that $\mathcal{H}(2)$ holds, hence $\mathcal{H}(3)$ holds, hence $\mathcal{H}(4)$ holds... and therefore $\mathcal{H}(k)$ will be true for all $k \geq 1$.

(**a**) *Show that $\mathcal{H}(1)$ is true.*

(**b**) *Show that if $\mathcal{H}(k)$ is true for some k, then $\mathcal{H}(k+1)$ is also true.*

Problem 8.4 (4 points)**.** *The goal of this problem is to use spectral clustering techniques on real data. The file* `adjacency.tex` *contains the adjacency matrix of a graph taken from a social network. This graphs has $n = 328$ nodes (that corresponds to users). An edge between user i and user j means that i and j are "friends" in the social network. The notebook* `friends.ipynb` *contains functions to read the adjacency matrix as well as instructions/questions.*

While we focused in the lectures (and in the notes) on the graph Laplacian

$$L = D - A,$$

where A is the adjacency matrix of the graph, and $D = \mathrm{Diag}(\deg(1), \dots, \deg(n))$ is the degree matrix, we will use here the "normalized Laplacian" (instead of L)

$$L_{\mathrm{norm}} = D^{-1/2} L D^{-1/2} = \mathrm{Id}_n - D^{-1/2} A D^{-1/2},$$

where $D^{-1/2} = \mathrm{Diag}(\deg(1)^{-1/2}, \dots, \deg(n)^{-1/2})$. The reason for using a different Laplacian is that then "unnormalized Laplacian" L does not perform well when the degrees in the graph are very broadly distributed, i.e. very heterogeneous. In such situations, the normalized Laplacian L_{norm} is supposed to lead to a more consistent clustering.

It is intended that you code in Python and use the provided Jupyter Notebook. Please only submit a pdf version of your notebook (right-click → 'print' → 'Save as pdf').

Problem 8.5 ($\star$)**.** *Define, for $M \in \mathbb{R}^{n \times m}$*

$$\|M\|_\star = \sum_{i=1}^{\min(n,m)} \sigma_i,$$

where the σ_i's are the singular values of M. Show that $\|\cdot\|_\star$ is a norm (i.e. verifies the definition 1.1 of Lecture 4) on $\mathbb{R}^{n \times m}$.