DS-GA 1014: Homework Problem Set 8

Optimization and Computational Linear Algebra for Data Science (Fall 2018)

Afonso S. Bandeira
bandeira@cims.nyu.edu
http://www.cims.nyu.edu/~bandeira

Brett D. Bernstein
brettb@
~brettb

Due on Friday November 16, 2018

This homework problem set is due on November 16 on NYU Classes.

If you have questions about the homework feel free to contact Brett Bernstein (brett.bernstein@nyu.edu**) or myself, or stop by our office hours.**

Unless otherwise stated all answers must be mathematically justified, even questions that ask for a computation.

Try not to look up the answers. You'll learn much more if you try to think about the problems without looking up the solutions. If you need hints, feel free to ask on Piazza.

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).

Late submissions will be graded with a penalty of 10% per day late. Weekend days do not count — from Friday to Monday counts as only 1 day.

If you need to impose extra conditions on a problem to make it easier (or consider specific cases of the question, like taking n to be 2, e.g.), state explicitly that you have done so. Solutions where extra conditions were assumed, or where only special cases where treated, will also be graded (probably scored as a partial answer).

Problems with a $(*)$ are extra credit — they will not (directly) contribute to your score for this homework. However, for every four extra credit questions successfully answered you get a homework "bye": your lowest homework score (or one you did not hand in) gets replaced by a perfect score.

Problem 8.1 *Let $A, B \in \mathbb{R}^{n\times n}$ with A invertible. Below all matrix norms are spectral.*

1. *Compute $1/\|A^{-1}\|$ in terms of the singular values of A, and justify your answer.*
2. *Prove that $\|Bx\| \geq \|Ax\| - \|(A-B)x\|$ for all $x \in \mathbb{R}^n$.*

3. *Prove that B is invertible if $\|A - B\| < 1/\|A^{-1}\|$. In words, a matrix that is close to an invertible matrix is also invertible.*

Problem 8.2 *We say that $S \subseteq \mathbb{R}^n$ is an affine subspace of $\mathbb{R}^n$ if it has the form $S = V + u$ for some (linear) subspace $V \subseteq \mathbb{R}^n$ and some $u \in \mathbb{R}^n$ where $V + u$ is defined to mean*

$$V + u := \{v + u : v \in V\}.$$

In other words, affine subspaces are shifted linear subspaces.

1. *Suppose $S = V + u$ is an affine subspace of $\mathbb{R}^n$ for some (linear) subspace $V \subseteq \mathbb{R}^n$ and $u \in \mathbb{R}^n$. Determine the set of all $x \in \mathbb{R}^n$ that satisfy $V + u = V + x$, and justify your answer.*

2. *For $v_1, \ldots, v_k \in \mathbb{R}^n$ an affine combination has the form*

$$\alpha_1 v_1 + \cdots + \alpha_k v_k$$

where $\alpha \in \mathbb{R}^k$ and $\alpha_1 + \cdots + \alpha_k = 1$. Prove that if $v_1, \ldots, v_k$ are elements of an affine subspace $S \subseteq \mathbb{R}^n$, then any affine combination of them are also in S.

Problem 8.3 *Let $a_1, \ldots, a_n \in \mathbb{R}^p$ be n datapoints, and let $A \in \mathbb{R}^{p \times n}$ denote the matrix with a_i as its ith column. Let $\mathbf{1} \in \mathbb{R}^n$ denote the vector of all ones, let $\mu \in \mathbb{R}^p$ be defined by*

$$\mu = \frac{1}{n} \sum_{i=1}^{n} a_i,$$

and let $\tilde{A} = A - \mu \mathbf{1}^T$.

1. *Define the total variance of the data by*

$$\Gamma = \frac{1}{n-1} \sum_{i=1}^{n} \|a_i - \mu\|^2.$$

   (a) *Give a formula for Γ in terms of the singular values of a matrix, and justify your answer.*

   (b) *Suppose we use PCA to replace each $a_i \in \mathbb{R}^p$ with a lower-dimensional approximation $\hat{a}_i \in \mathbb{R}^d$ where $d < p$. What is the total variance of the new approximate datapoints $\hat{a}_1, \ldots, \hat{a}_n$ (express your answer in terms of singular values)? Justify your answer.*

2. *Define the variance of the data in the direction $v \in \mathbb{R}^p$ (where $\|v\| = 1$) by*

$$\Gamma_v = \frac{1}{n-1} \sum_{i=1}^{n} (v^T a_i - \mu_v)^2,$$

where μ_v is given by

$$\mu_v = \frac{1}{n} \sum_{i=1}^{n} v^T a_i.$$

(a) *Compute Γ_v in terms of $\tilde{A}$, and justify your calculation.*

(b) *Determine which v with $\|v\| = 1$ maximizes Γ_v, and justify your answer.*

Problem 8.4 *In each of the following questions, it is intended that you solve the problem using the numpy/scipy libraries in Python, but you can use any programming language. Please include just the answer in your writeup, and include the code files in your submission (i.e., do not include the code in your writeup). A good idea is to test your code on some small simple instances you create by hand to make sure it is correct. Note that calculations on the computer do not use exact arithmetic, so you should allow for some margin of error in your code.*

The file `HW8matrix1.txt` contains a 5×4 matrix A, and the file `HW8matrix2.txt` contains a 5×1 vector y. Each line of each file represents a row of the corresponding matrix, and the values on each line are space-delimited. Report all answer to two decimal places.

1. *Compute the minimizer $x^* \in \mathbb{R}^4$ of*

$$\min_{x \in \mathbb{R}^4} \|Ax - y\|.$$

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

$$\min_{x \in \mathbb{R}^4} \|Ax - (y + v)\|$$

   is also x^*.

3. *Find a vector $w \in \mathbb{R}^5$ with $w_1 > 0$ and $\|w\| = 1$ such that the minimizer $\tilde{x}$ of*

$$\min_{x \in \mathbb{R}^4} \|Ax - (y + w)\|$$

   maximizes the error $\|x^* - \tilde{x}\|$, 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 8.5 (For Extra Credit)** *Prove that if $M \in \mathbb{R}^{n \times n}$ is a symmetric matrix and $d \leq n$ then*

$$\max_{\substack{U \in \mathbb{R}^{n \times d} \\ U^T U = \mathrm{I}_{d \times d}}} \mathrm{Tr}\left(U^T M U\right) = \sum_{k=1}^{d} \lambda_k^{(+)}(M),$$

where $\lambda_k^{(+)}$ is the kth largest eigenvalue of M (i.e., $\lambda_1^{(+)} \geq \lambda_2^{(+)} \geq \cdots$).