CSE 392: Matrix and Tensor Algorithms for Data Spring 2025

Homework 1

Due Date: 02-07-2025

Assignments are to be submitted through Canvas, and should be individual work. You can discuss the problems, but should submit individually. Preferably typewritten.

Problem 1. Norm inequalities

(i) For any $\boldsymbol{v} \in \mathbb{R}^n$, prove:

$$\|\boldsymbol{v}\|_2^2 \leq \|\boldsymbol{v}\|_1 \|\boldsymbol{v}\|_\infty \leq \sqrt{n}\|\boldsymbol{v}\|_2^2$$

Hint: Try the special case $n = 2$ such that $\boldsymbol{v} = [v_1, v_2]$, and assume without loss of generality that $v_1 \geq v_2 \geq 0$. Why does this cover the general case?

(ii) For any vector $\boldsymbol{v}$ and matrix $\boldsymbol{A}$, show that:

$$\|\boldsymbol{A}\boldsymbol{v}\|_2 \leq \|\boldsymbol{A}\|_F \|\boldsymbol{v}\|_2.$$

(iii) Use the above to show that the Frobenius norm is consistent (sub-multiplicative):

$$\|\boldsymbol{A}\boldsymbol{B}\|_F \leq \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F.$$

Hint: Start by writing that $\|\boldsymbol{A}\boldsymbol{B}\|_F^2 \leq \sum_{j=1}^n \|\boldsymbol{A}\boldsymbol{b}_j\|_2^2$, where $\boldsymbol{b}_j$ is the j-th column of $\boldsymbol{B}$. Then apply part (ii) to $\|\boldsymbol{A}\boldsymbol{b}_j\|_2^2$ for each j.

(iv) Show the same using Cauchy-Schwartz inequality (exercise in Page 31 of Lecture 1).

Problem 2. Exploring Concentration Bounds

Let x be a random variable uniformly distributed in the interval $[0, 1]$. Since we know x's distribution exactly, we can easily check that $\Pr[\mathrm{x} \geq 7/8] = 1/8$. But let us take a look at what various concentration inequalities would predict about this probability using less information about x.

(i) Given an upper bound on $\Pr[\mathrm{x} \geq 7/8]$ using Markov's inequality.

(ii) Give an upper bound on $\Pr[\mathrm{x} \geq 7/8]$ using Chebyshev's inequality.

(iii) Given an upper bound on $\Pr[\mathrm{x} \geq 7/8]$ by applying Markov's inequality to the random variables x^2 (the "raw" second moment). Note that this is slightly different than using Chebyshev's inequality, which applies Markov to "central" second moment $(\mathrm{x} - \mathbb{E}[\mathrm{x}])^2$.

(iv) What happens for higher moments? Applying Markov's to x^q for $q = 3, 4, \ldots, 10$. Describe what you observe in a sentence. Which value of q gives the tightest bound?

(v). One take-away here is that, depending on the random variable being studied, it is not always optimal to use the variance as a deviation measure to show concentration. Markov's can be used with any monotonic function g, and as we see above, different choices might give better bounds. Exhibit a monotonic function g so that applying Markov's to $g(\mathrm{x})$ gives as tight an upper bound on $\Pr[\mathrm{x} \geq 7/8]$ as you can. Maximum points if you can get $\Pr[\mathrm{x} \geq 7/8] \leq 1/8$, which would be the best possible.

Problem 3. Outer products and matrices

Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

(i) Assume $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{v}^\top$ is non-singular. The inverse $\boldsymbol{A}^{-1}$ has the form $\boldsymbol{I} - \alpha\boldsymbol{u}\boldsymbol{v}^\top$ for some scalar α. Derive a formula for this α in terms of $\boldsymbol{u}, \boldsymbol{v}$.
Hint: Multiply out $(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{v}^\top)(\boldsymbol{I} - \alpha\boldsymbol{u}\boldsymbol{v}^\top)$ and find value for α to reduce the product to $\boldsymbol{I}$.

(ii) What is $\det(\boldsymbol{u}\boldsymbol{v}^\top)$? What is $\mathrm{Tr}(\boldsymbol{u}\boldsymbol{v}^\top)$?

(iii) What is $\det(\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top)$? What is $\mathrm{Tr}(\boldsymbol{I} + \boldsymbol{u}\boldsymbol{v}^\top)$?

(iv) What is the determinant of $\boldsymbol{B} + \boldsymbol{u}\boldsymbol{v}^\top$ when $\boldsymbol{B}$ is nonsingular.

Problem 4. Prove the Eckart-Young-Mirsky Theorem

For any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times d}$ with rank r, let $k \le r$ and $\boldsymbol{A}_k = \sum_{i=1}^k \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ then

$$\min_{\boldsymbol{B}:rank(\boldsymbol{B})=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2 = \|\boldsymbol{A} - \boldsymbol{A}_k\|_2 = \sigma_{k+1}.$$

Hints:
(a) First, show that $\|\boldsymbol{A} - \boldsymbol{B}\|_2 \ge \sigma_{k+1}$ for any rank k matrix $\boldsymbol{B}$.
(b)For this, consider a particular subspace $\mathcal{X}$ such that $Null(\boldsymbol{B}) \cap \mathcal{X} \ne \{0\}$. Think in terms of dimensions, what should $\dim(\mathcal{X})$ be such that this is true.
(c) Let $\boldsymbol{x}_0 \in Null(\boldsymbol{B}) \cap \mathcal{X}, \boldsymbol{x}_0 \ne 0$. Show that $\|(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x}_0\|_2 \ge \sigma_{k+1}\|\boldsymbol{x}_0\|_2$.
(d) Next, show when $\boldsymbol{B} = \boldsymbol{A}_k$, we achieve the min.

Problem 5. Regression

In this problem, we will explore the performances of different regression methods, which we studied in the class, on a classical machine learning dataset. We will consider the Arrhythmia dataset, where the goal is to distinguish between the presence and absence of cardiac arrhythmia amongst patients. The dataset has 257 features (after removing missing and categorical features) and 452 instances. Load the dataset:
`load arrhythmia-clean.mat` in matlab or
`scipy.io.loadmat(arrhythmia-clean.mat)` in Python.

(i) Compute the rank of the feature matrix X. What can you say about using least squares for the dataset based on this info?

(ii) Split the dataset into training and test sets (80%-20%). You can use the `randompartition` function which we used in the class. Try the following regression methods on the dataset:

- *Least squares regression.* Can you improve the results using truncated SVD?
- *Ridge regression* with different regularization parameter λ (ranging from 0.1 to 100).
- *Lasso regression* with different regularization parameter λ.
- *Kernel ridge regression* with different kernels (try linear, polynomial and Gaussian kernels).

Compute the coefficients on the training set, and report results in terms of the mean squared error (MSE) obtained on the test set. Submit your scripts along with the assignment.