CSE 392: Matrix and Tensor Algorithms for Data Spring 2024

Homework 1

Due Date: 02-14-2024

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) Show that for any $\boldsymbol{x} \in \mathbb{R}^n$:

$$\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_2 \leq \sqrt{n}\|\boldsymbol{x}\|_\infty$$

(ii) Then use this to prove that for $n \times n$ matrices :

$$\frac{1}{\sqrt{n}}\|\boldsymbol{A}\|_\infty \leq \|\boldsymbol{A}\|_2 \leq \sqrt{n}\|\boldsymbol{A}\|_\infty$$

(iii) Show that the Frobenius norm and the 2-norm of $n \times n$ matrices are related by:

$$\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_F \leq \sqrt{n}\|\boldsymbol{A}\|_2.$$

(iv) Show that $\|\boldsymbol{AB}\|_\xi \leq \|\boldsymbol{A}\|_2 * \|\boldsymbol{B}\|_\xi$, for $\xi \in \{2, F\}$.

Problem 2. Hoeffding's Lemma and Inequality

(i) For a random variable x, with $\mathbb{E}[\mathrm{x}] = 0$ and $a \leq \mathrm{x} \leq b$, prove that

$$\mathbb{E}[e^{t\mathrm{x}}] \leq e^{\frac{t^2(b-a)^2}{8}}$$

(ii) Use the above to show that, if $\mathrm{x}_1, \ldots, \mathrm{x}_n$ are mean-zero independent random variables with each $a_i \leq \mathrm{x}_i \leq b_i$, then

$$\mathbb{E}\left[\exp\left(t\sum_{i=1}^{n}\mathrm{x}\right)\right] \leq \exp\left(\frac{t^2\sum_{i=1}^{n}(b_i-a_i)^2}{8}\right)$$

(iii) Use the above inequality to derive the *Hoeffding's inequality*: under the same setting as in the Case (ii), we have, for any $t > 0$,

$$\Pr[|\bar{\mathrm{x}}_n| \geq t] \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

where $\bar{\mathrm{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathrm{x}$ is the sample mean.

(iv) Next, show that

$$\mathrm{Var}[\bar{\mathrm{x}}_n] \leq \frac{\sum_{i=1}^{n}(b_i-a_i)^2}{n^2}.$$

(v) Using the above, estimate the probability that the sample average is more than t standard deviations from its expectation. Recall in the class, we saw that we can also estimate this probability using Chebyshev's inequality. How do the two estimates compare?

Problem 3. Projection

(i) For orthonormal $\boldsymbol{U}$, show that $\boldsymbol{U}^\top \boldsymbol{b} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{U}\boldsymbol{x} - \boldsymbol{b}\|_2$. (You might use the normal equations, or the Pythagorean theorem as stated for projections.)

(ii) For orthonormal $\boldsymbol{U}$, show that $\boldsymbol{P}_{\boldsymbol{U}}\boldsymbol{b}$ is the closest vector to $\boldsymbol{b}$ in $span(\boldsymbol{U})$.

(iii) Show that $\boldsymbol{A}^\dagger = (\boldsymbol{A}^\top \boldsymbol{A})^\dagger \boldsymbol{A}^\top$.

(iv) Show for symmetric $\boldsymbol{P}$ that $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P} \implies \boldsymbol{P} = \boldsymbol{U}\boldsymbol{U}^\top$ for some orthonormal $\boldsymbol{U}$.

Problem 4. Prove the Eckart-Young-Mirsky Theorem

For any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times d}$ with rank r, let $k \leq r$ 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 $\|\boldsymbol{A} - \boldsymbol{B}\|_2 \geq \sigma_{k+1}$ for any rank k matrix $\boldsymbol{B}$.
(b)For this, consider a particular subspace $\mathcal{X}$ such that $Null(\boldsymbol{B}) \cap \mathcal{X} \neq \{0\}$. Think in terms of dimensions, what should $\dim(\mathcal{X})$ be such that this is true.
(c) Let $\boldsymbol{x}_0 \in Null(\boldsymbol{B}) \cap \mathcal{X}, \boldsymbol{x}_0 \neq 0$. Show that $\|(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x}_0\|_2 \geq \sigma_{k+1}\|\boldsymbol{x}_0\|_2$.
(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.