Indian Institute of Technology Kharagpur
Centre of Excellence in Artificial Intelligence

AI61003 Linear Algebra for AI and ML
Assignment 2, Due on: October 20, 2021

ANSWER ALL THE QUESTIONS

1. Let $A, B \in \mathbb{R}^{n\times n}$. Prove that $\|AB\|_2 \leqslant \|A\|_2\|B\|_2$. This property of 2-norm is called as sub-multiplicativity property. Does this property hold true for Frobenius norm?

2. Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix. Define $\max \operatorname{mag}(A)$ and $\min \operatorname{mag}(A)$ and $\operatorname{cond}(A)$. Show that

   (a) $\max \operatorname{mag}(A) = \dfrac{1}{\min \operatorname{mag}(A^{-1})}$

   (b) $\operatorname{cond}(A) = \dfrac{\max \operatorname{mag}(A)}{\min \operatorname{mag}(A)}$

3. In each of the following cases, consider the matrix $A \in \mathbb{R}^{m\times n}$ as a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Plot the unit sphere in $\mathbb{R}^n$. Plot the ellipsoid obtained in $\mathbb{R}^m$ as image of the unit sphere in $\mathbb{R}^n$. Compute the condition number of A (using inbuilt command). Further, if $m = n$, check whether the matrix is invertible. Compute the determinant of A as well. Is there any relationship between determinant and condition number?

   (a) $A = \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} \\ -1 & 1 \end{pmatrix}$

   (b) $A = \begin{pmatrix} -2 & 1 & 2 \\ 0 & 2 & 0 \end{pmatrix}$

   (c) $A = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 0.8 \end{pmatrix}$

   (d) $A = \begin{pmatrix} 1 & 0 \\ 0 & -10 \end{pmatrix}$

   (e) $A = \begin{pmatrix} 1 & 1 \\ 1 & \varepsilon \end{pmatrix}$, where $\varepsilon = 10, 5, 1, 10^{-1}, 10^{-2}, 10^{-4}, 0$.

4. For a matrix A with the property that the columns of A are linearly independent, give the geometrical interpretation of the least squares solution to the problem $Ax = b$ and justify the name *normal equations.* In case, the matrix A does not have linearly independent columns, comment on the nature of the least squares solution.

5. Consider the system of linear equations $Ax = b$ where $A \in \mathbb{R}^{n \times n}$ is an invertible matrix and $b \in \mathbb{R}^n$ is a given vector. Discuss the advantages in the case when A is orthogonal.

6. *Bi-linear interpolation*: We are given scalar value at each of the MN grid points of a grid in $\mathbb{R}^2$ with a typical grid point represented as $P_{ij} = (x_i, y_j)$ where $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$ and $x_1 < x_2 < \cdots < x_M$ and $y_1 < y_2 < \cdots < y_N$. Let the scalar value at the grid point P_{ij} be referred to as F_{ij} for $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$. A *bi-linear interpolation* is a function of the form

$$f(u, v) = \theta_1 + \theta_2 u + \theta_3 v + \theta_4 uv$$

where $\theta_1, \theta_2, \theta_3, \theta_4$ are the coefficients. This function further satisfies $f(P_{ij}) = F_{ij}$ for $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$.

   (a) Express these interpolation conditions as a system linear equations of the form $A\theta = b$ where b is an MN vector consisting of F_{ij} values. Write clearly all the entries of A, θ and b and their sizes.

   (b) What are the minimum values of M and N so that you may expect a unique solution to the system of equations $A\theta = b$?

7. *Iterative LS*: Let $A \in \mathbb{R}^{m \times n}$ have linearly independent columns and let $b \in \mathbb{R}^m$ be a given vector. Further, let $\widehat{x}$ denote the LS solution to the problem $Ax = b$. Define $x^{(1)} = 0$ and for $k = 0, 1, 2, \dots$

$$x^{(k+1)} = x^{(k)} - \frac{1}{\|A\|^2} A^\top \left(Ax^{(k)} - b \right)$$

   (a) Show that the sequence $\{x^{(k)}\}$ converges to $\widehat{x}$ as $k \to \infty$.

   (b) Discuss the computational complexity of computing $\{x^{(k)}\}$ for any $k \geqslant 1$.

   (c) Generate a 30×10 random matrix A and a 30×1 random vector b. Check that the matrix is full column rank! Run the algorithm for 100 steps. Verify numerically that the algorithm converges to $\widehat{x}$.

   (d) Do you think this iterative method may be computationally beneficial over the direct methods of computing the LS solution?

8. Suppose that $z_1, z_2, \dots, z_{100}$ is observed time series data. An autoregressive model for this data has the following form.

$$\widehat{z}_{t+1} = \theta_1 z_t + \cdots + \theta_M z_{t-M+1}, \quad t = M, M+1, \dots, 100$$

where M is the memory or the lag of the model. This model can be used to predict the next observation in the time series.

   (a) Set up a least squares problem to estimate the parameters in the model.

   (b) Clearly write down the matrices A and b in the least squares formulation.

   (c) What is the special structure that one can observe in A?

(d) Is there any relation of rank of A with M?

9. *Polynomial Classifier:* Generate 500 random vectors $x^{(i)} \in \mathbb{R}^2$ for $i = 1, 2, \ldots, 500$ from a standard normal distribution. Define, for $i = 1, 2, \ldots, 500$,

$$y^{(i)} = \begin{cases} +1 & x_1^{(i)} x_2^{(i)} \geqslant 0 \\ -1 & \text{otherwise} \end{cases}$$

Fit a polynomial least squares classifier of degree 2 to the data set using the polynomial

$$\widetilde{f}(x) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 x_1^2 + \theta_6 x_2^2$$

(a) Give the error rate of the classifier using the confusion matrix.

(b) Show the regions in the $\mathbb{R}^2$ plane where the classifier model $\widehat{f}(x) = 1$ and $\widehat{f}(x) = -1$.

(c) Does the second degree polynomial $g = x_1 x_2$ classify the generated points with zero error? Compare the parameters estimated polynomial model from the data with those of g.

10. *MNIST dataset*: For each of the digit $0, 1, \ldots, 9$ randomly select 1000 images to generate a training data set of size 10000 images. Similarly generate a test data set of 1000 images as a test data set. Fit a linear least squares classifier to classify the data set into 10 classes and test prediction accuracy of the model using the 10×10 confusion matrix. Do not use any inbuit functions for fitting the model.

***************** THE END *****************