Indian Institute of Science Banglore
Department of Computational and Data Sciences (CDS)

DS284: Numerical Linear Algebra

Mid-semester Exam 2024

Faculty Instructor: Dr. Phani Motamarri

TAs: Gourab Panigrahi, Srinibas Nandi, Nihar Shah, Rushikesh Pawar, Surya Neelakandan

Duration: 8:30 AM to 10:00 AM, Oct 22 2024 Max Points: 50

Notations: (i) Vectors and matrices are denoted by bold faced lower case and upper case alphabets respectively. (ii) Set of all real numbers is denoted by $\mathbb{R}$ (iii) Set of all n dimensional vectors is denoted by $\mathbb{R}^n$ and set of all $m \times n$ matrices is denoted by $\mathbb{R}^{m\times n}$. (iv) $\mathbf{I}_n$ denotes the identity matrix of order n.

Problem 1 [6×3.5=21]

Assert if the following statements are True or False. Give a detailed reasoning for your assertion. Marks will be awarded only for your general reasoning and not just for counter examples

(a) If $\hat{\mathbf{x}} \in \mathbb{R}^n$ satisfies the system of equations $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a full rank matrix with $m > n$ and $\mathbf{b} \in \mathbb{R}^m$, then $\hat{\mathbf{x}}$ should always satisfy the system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$

(b) Let $\mathbf{u} \in \mathbb{R}^m$ be such that $\|\mathbf{u}\|_2 = 1$ with all entries of $\mathbf{u}$ are positive. If we define $\mathbf{A} \in \mathbb{R}^{1\times m}$ as $\mathbf{A} = \mathbf{u}^T$, then induced norm of $\mathbf{A}$ in the 2-norm sense is the max element of $\mathbf{u}$ i.e., $\|\mathbf{A}\|_2 = \max\{u_i\}$.

(c) For the matrix $\mathbf{G} = \mathbf{I}_m - 3\mathbf{q}\mathbf{q}^T$ where $\mathbf{q} \in \mathbb{R}^m$ is a unit vector, the geometric multiplicity of the eigenvalue 1 is $m-1$.

(d) Let $\mathbf{S} \in \mathbb{R}^{m\times m}$ be a symmetric positive definite matrix. There can exist a maximum element in $\mathbf{S}$ that does not lie on the diagonal. (**Hint:** If S_{ij} is the largest element for some $i \neq j$, then $\mathrm{S}_{ii} + \mathrm{S}_{jj} \leqslant 2\mathrm{S}_{ij}$ should always be true.)

(e) Let $\mathbf{v}_1 \in \mathbb{R}^m$ be a unit-vector and the matrix $\mathbf{F} = \mathbf{I}_m - 2\mathbf{v}_1\mathbf{v}_1^T$ denote a Householder matrix. Furthermore, $(\epsilon_1,\mathbf{v}_1)$, $(\epsilon_2,\mathbf{v}_2)$, $(\epsilon_3,\mathbf{v}_3)$ $\cdots$ $(\epsilon_n,\mathbf{v}_n)$ represent the first $n(< m)$ largest eigenvalue/eigenvector pairs of a real symmetric matrix $\mathbf{S} \in \mathbb{R}^{m\times m}$. Now, assert the following statement: The n-dimensional space spanned by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\cdots$ $\mathbf{v}_n$ is same as the eigenspace spanned by the eigenvectors of a matrix $\mathbf{T} = \mathbf{F}^{-1}\mathbf{S}\mathbf{F}$ corresponding to first n largest eigenvalues of $\mathbf{T}$.

(f) If $\mathbf{P} \in \mathbb{R}^{m\times m}$ is an orthogonal projector matrix, then one can choose the same set of orthogonal eigenvectors to diagonalize $\mathbf{P}$ and $\mathbf{I_m} - \mathbf{P}$.

Problem 2 [5+2+5+4=16]

Consider four vectors

$$\mathbf{a}_1 = \begin{bmatrix}1\\1\\4\end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix}1\\2\\0\end{bmatrix} \quad \mathbf{v} = \begin{bmatrix}2\\2\\1\end{bmatrix} \quad \mathbf{w} = \begin{bmatrix}0.4\\0.7\\0.4\end{bmatrix}$$

(a) Construct the matrix $\mathbf{A}$ whose columns are the two vectors $\mathbf{a}_1, \mathbf{a}_2$. Compute the successive orthogonal transformations $\mathbf{Q}_1, \mathbf{Q}_2$ that transforms $\mathbf{A}$ into an upper triangle matrix $\mathbf{R}$ as shown below:

$$\mathbf{Q}_2\mathbf{Q}_1\mathbf{A} = \mathbf{R}$$

From this exercise, write down the orthonormal vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ that span the successive column subspaces of the matrix $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2]$

(b) Do the vectors $\mathbf{v}$ and $\mathbf{w}$ lie in the space spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$? Explain.

(c) If the answer for (b) is "yes" for one of the vectors $\mathbf{v}$ and $\mathbf{w}$, write down the component(s) of this vector in the basis of $\mathbf{a}_1$ and $\mathbf{a}_2$ in a column vector format. If the answer for (b) is "no" for one of the vectors $\mathbf{v}$ and $\mathbf{w}$, then compute the solution $\mathbf{x}$ which minimizes the difference between $\mathbf{Ax}$ and this vector in the 2-norm sense.

(d) If the matrix $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2]$ denotes a data matrix related to 3 experiments and 2 features, construct a zero mean feature data matrix $\mathbf{A}_o$. Using this data matrix $\mathbf{A}_o$, now compute the unit vector in the direction of the second maximum variance.

Problem 3 [5+8=13]

Regression is one of the key aspects of machine learning. In most linear regression type problems, the aim is to find a linear map between input feature vector $\mathbf{x}$ and an output target scalar y based on a given dataset. One intrinsic assumption in linear regression is that all the data samples are given equal weightage. However, the data is usually collected from multiple sources with varying levels of fidelity involved in the measurement of target scalar y. This necessitates attributing different weights to different data samples. In such a scenario, we define the loss function (or the objective function that one has to minimize to find the optimal parameters in our regression) corresponding to our linear regression problem as follows:

$$\mathbf{L} = \sum_{i=1}^{n} w_i(y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \theta_3 x_{i3} - \ldots - \theta_m x_{im})^2 \tag{1}$$

where n is the number of data samples in the given data set $(\mathbf{x}_i, y_i)_{i=1}^{n}$. Here $\mathbf{x}_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}$, $w_i > 0$ are given and $n >> m$. The unknown quantities in linear regression are the parameter values $\theta_0, \theta_1, \ldots, \theta_m$, which need to be estimated through minimization of the loss function. Define a feature matrix $\mathbf{X} \in \mathbb{R}^{n\times(m+1)}$ such that the first entry of each row is 1, and the remaining m entries in each row correspond to the m entries of the feature vector $\mathbf{x}_i$. Based on the information above, answer the following questions:

(a) Show that if $\mathbf{X}$ is a full rank matrix, the optimal parameters obtained through minimization of $\mathbf{L}$ is unique. (*Hint:* Find the first derivatives of $\mathbf{L}$)

(b) To improve the generalizability of the regression model, it is typical to define a modified loss function $\mathbf{J} = \mathbf{L} + \eta \sum_{j=0}^{m} \theta_i^2$, where $\mathbf{L}$ is defined as in equation (1) and $\eta > 0$ is a user-controllable parameter. The set of optimal parameters can be obtained by minimizing this modified loss function $\mathbf{J}$. Show that the set of optimal parameters, i.e., $\theta_0, \theta_1, \ldots, \theta_m$ can always be made unique by varying the value of η irrespective of the rank of $\mathbf{X}$. Given a dataset $(\mathbf{x}_i, y_i)_{i=1}^{n}$, for what value of (or range of values of) η will the set of optimal parameters be unique?