CIS4930 - Math for Machine Learning

Homework 2

Fernando Sckaff

February 6^{th}, 2023

Question 1: Matrix Properties

For any $n \times n$ non-singular matrix A, show that the following 4 properties are equivalent.
(i) A has an inverse,
(ii) $\det(A) \neq 0$,
(iii) $\operatorname{rank}(A) = n$ and
(iv) for any vector $z \neq O_n, Az \neq O_n$ (i.e., A annihilates no nontrivial vector).

Solution:

Note:-

(a) A non-singular matrix is a square one whose determinant is not zero.
(b) For a $2x2$ square matrix A, has a determinant $det(A) = |ad - bc| \neq 0$
(c) The inverse of a $2x2$ matrix A, A^{-1} can be given by:

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{|ad-bc|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

As seen, A^{-1} can only exist if $det(A) \neq 0$
(d) The determinant of a matrix is zero if and only if the column vectors are linearly dependent
Given a constant x and a matrix $\begin{bmatrix} a & x \cdot a \\ c & x \cdot c \end{bmatrix}$, its determinant is $|a \cdot (x \cdot c) - c \cdot (x \cdot a)|$, which is equivalent to $|x \cdot a \cdot c - x \cdot a \cdot c| = 0$

Given note (c), properties (i) and (ii) hold true since an inverse of a matrix can only exist if $det(A) \neq 0$. Also, (iii) is equivalent to (i) and (ii) since the column vectors of the $n \times n$ matrix are linearly independent and the determinant is not zero given (d). Therefore, the column vectors of A span the whole $\mathbb{R}^n$ and $rank(A) = n$.

Lastly, since matrix A spans $\mathbb{R}^n$, then there is no vector $z \neq O_n$ (which is a vector in $\mathbb{R}^n$) orthogonal to A $\Rightarrow Az \neq O_n$.

All properties are equivalent of one another. ☺

Question 2: The Groups of StrangeBrew

There are n citizens living in StrangeBrew. Their main occupation was forming various groups, which at some point started threatening the very survival of the city.
In order to limit the number of groups, the city council decreed the following innocent-looking rules:
- Each group has to have an odd number of members.
- Every two groups must have an even number of members in common.

Prove that under these rules, it is impossible to form more groups than n, the number of citizens. You must use matrix properties to prove this theorem.

Hint: Consider defining an $m \times n$ matrix A (for n citizens and m groups $G_1, G_2, \ldots, G_m$) by $A_{ij} = 1$ if $j \in G_i$ and 0 otherwise.

Solution: Let's evaluate the worst-possible scenario.
Given:
(1) The size of a group G_m is $(2k+1)|k \in \mathbb{Z}^+$
(2) The size of the set $G_a \cap G_b$ is $(2u)|u \in \mathbb{Z}^+$, and $0 \leqslant a < b \leqslant m$.

Note:-

Since the number of citizens in each group is odd, and their intersection must be even, then we know there will be at least one citizen that will not be in both groups/sets. **Therefore, subsets are not allowed.**

The number of groups/rows m cannot exceed the number of citizens/columns n. If $m > n$ then all the rows within $(r|r \in n < r \leqslant m)$ are merely linear transformations from the first n rows. Consider identity matrix T:

$$T = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

In other words, for any new row $r = [a\, b\, \dots\, n]$ where $r \neq 0$ or any previous row, there will be a combination of at most n rows that could create r. Any one person will be a subset of your new group - and (2) says the number of members in common of two groups should be an even number. The $\mathbb{R}^n$ identity matrix the easiest way to visualize a correct answer for the upper bound. Every group consisting of an odd number of members (one person), and an even number of people in common with other groups (zero).

In case you think the worst-case scenario would consist of citizens trying to get into groups with as many people as possible, that could not be true since, conceptually, any group consisting of 3 people are erasing the possibility of having 2 more groups (in case everyone stood by themselves).
Also, theoretically, if people wanted to join other people's groups:
(I) For an odd number of citizens:

$$M = \begin{bmatrix} \underbrace{1 \quad 1 \quad 1 \quad \cdots \quad 1}_{n\,[odd]} \end{bmatrix}$$

All members will eventually become one single odd group.
(II) For an even number of citizens:

$$M = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 \end{bmatrix}$$
$$\underbrace{\qquad\qquad\qquad}_{n-1\,[odd]}$$

All members but one will become one single odd group, and the odd one out will be their own group. Citizens willing to join bigger groups are actually discouraged, and a centralized union tends to form. Want more groups in this society? In the end, you might be by yourself, or with everyone else. 😄

Corollary 0.0.1

By the result of the proof, we can then show some basic identity-like patterns that give us the largest amount of groups. Be mindful - column pattern may change as it spans the same column space:
(1) For odd number of citizens:

$$ODD = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{bmatrix}$$

(2) For even number of citizens:

$$EVEN = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 1 & 1 \\ 0 & 1 & 0 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 1 & \dots & 0 & 1 \\ 0 & 1 & 1 & \dots & 1 & 0 \end{bmatrix}$$

Question 3: Linear Least-Squares

Assume that the $m \times n$ matrix A has rank n where $m > n$. Let $P = A\left(A^T A\right)^{-1} A^T$
1. Show that $P^T = P$ and $P^2 = P$.
2. Show that $\|P\boldsymbol{b}\| \leqslant \|\boldsymbol{b}\|$ for any vector p-norm. What does this mean in the context of linear least squares?
3. Show that the solution to linear least-squares satisfies the condition $A\boldsymbol{x} = P\boldsymbol{b}$. Give a geometric explanation of this condition.

Solution:
1.
Proof of $P^T = P$

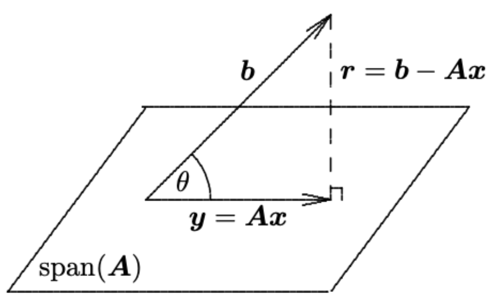
$$\begin{aligned} P^T &= (A(A^TA)^{-1}A^T)^T \\ &= (A^T)^T((A^TA)^{-1})^TA^T \\ &= A(A^TA)^T)^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ P^T &= P \quad 😄 \end{aligned}$$

Proof of $P^2 = P$:

$$\begin{aligned} P^2 &= (A(A^TA)^{-1}A^T)(A(A^TA)^{-1}A^T) \\ &= A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ P^2 &= P \quad 😄 \end{aligned}$$

2.
Conceptually speaking, taking the norm of the vector after projecting it onto a subspace of $\mathbb{R}^n$ will be lower than the norm of the vector before projecting - the difference is the orthogonal vector between itself and the plane. In case the vector is not already in the subspace of the projection, the vector will essentially reduce its norm to fit the subspace.
Since $\boldsymbol{b} = \boldsymbol{Pb} + \boldsymbol{P}_\perp\boldsymbol{b}$, when you project to Pb you are essentially losing $P_\perp b$, thus reducing your norm. 😄

3.
Consider image:



The Projection Matrix P is a special idempotent ($P^2 = P$), and symmetrical ($P^T = P$) matrix. It simply projects any vector onto the subspace by subtracting the vector orthogonal between the plane and the end of the vector. This orthogonal subtraction is equivalent to the residual of b in comparison to the subspace A that P is built upon.
The image above graphically shows b, the plane in which the matrix A spans and y, the vector that is the result from subtracting r from b.

$$Ax \approx b \Rightarrow Ax = Pb$$

😄

Question 4: Matrix Norms

Write a program to roughly carry out the optimization problem inherent in $\|A\|_\infty$, namely $\max_{x \neq O_n} \frac{\|Ax\|_\infty}{\|x\|_\infty}$. The program should take an arbitrary $m \times n$ matrix A and sample over 100000 random vectors $x \in \mathbb{R}^n$ (and use Gaussian random numbers) to approximately pick the maximum. Compute and empirically show that the analytic solution is an upper bound for the search. Argue if you have really shown anything and/or if the empirical search is merely suggestive (that is, it is not a proof of anything). For demonstration purposes, use $m = 9$ and $n = 6$. The matrix entries can also be randomly chosen.

Solution: on Google Colab, Link:

$https://colab.research.google.com/drive/1H7OmKjdwwLVBVQn6BJlD2mnxsCFRSgW0?usp=sharing$

$\boxed{\text{F.S.}}$

Note:-

- Question 1: 20/30 You cannot conclude anything from an argument centered on just 2x2 matrices.
- Question 2: 5/30 The argument does not work. Just because the ”identity matrix the easiest way to visualize...upper bound” does not imply that there are no other means of achieving this upper bound if indeed it is one since that has not been shown.
- Question 3: Correct
- Question 4: 10/20 Code looks incorrect. Need

  ```
  matrix_Linfnorm_obj = np.max(np.abs(y),0)/np.max(np.abs(x),0)
  ```

  and not what you have.