Math 118 Homework 1

Jiaping Zeng

10/7/2020

1. Suppose that $A \in \mathbb{R}^{m \times n}$, $\mathbf{v} \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n \times r}$. Determine the number of floating point operations required to compute $A\mathbf{v}$ and AB.

   Answer:

   $A\mathbf{v}$: Multiplying each row of A to v requires n multiplications, then combining the products requires $n-1$ additions. Since there are m rows, the total number of floating point operations is $m(n + (n-1)) = 2mn - m$.

   AB: For each column in B we would need $2mn - m$ operations as shown above, repeating it for r column would require $r(2mn - m) = 2mnr - mr$ operations.

2. Let $\mathbf{v} = [1, 0, 5, -3]^\top$ and $\mathbf{w} = [-2, 4, 5, 1]^\top$.

   (a) Compute $\|\mathbf{v}\|_1$, $\|\mathbf{w}\|_3$ and $\|\mathbf{v}\|_\infty$.

   Answer:

   $\|\mathbf{v}\|_1 = 1 + 0 + 5 + -3 = 3$

   $\|\mathbf{w}\|_3 = \sqrt[3]{(-2)^3 + 4^3 + 5^3 + 1^3} = \sqrt[3]{182} \approx 5.667051$

   $\|\mathbf{v}\|_\infty = \max_i |v_i| = |5| = 5$

   (b) Verify that $\langle \mathbf{v}, \mathbf{w} \rangle \le \|\mathbf{v}\|_2 \|\mathbf{w}\|_2$ by computing both sides.

   Answer: We have $\langle v, w \rangle = -2 + 0 + 25 - 3 = 20$ and $\|\mathbf{v}\|_2 \|\mathbf{w}\|_2 = \sqrt{1^2 + 6^2 + (-3)^2} \cdot \sqrt{(-2)^2 + 4^2 + 5^2 + 1^2} = \sqrt{46} \cdot \sqrt{46} = 46$, therefore $\langle \mathbf{v}, \mathbf{w} \rangle \le \|\mathbf{v}\|_2 \|\mathbf{w}\|_2$.

3. This question will introduce you to three different ways of thinking about matrix multiplication. For all three parts, let $U \in \mathbb{R}^{m \times n}$, $V \in \mathbb{R}^{m \times r}$ and $W \in \mathbb{R}^{r \times n}$. By $\mathbf{u}_i$ (respectively $\mathbf{v}_i$ and $\mathbf{w}_i$) I mean the i-th column of U (respectively V and W).

   (a) Show that $U^\top V = \begin{bmatrix} \mathbf{u}_1^\top \mathbf{v}_1 & \mathbf{u}_1^\top \mathbf{v}_2 & \ldots & \mathbf{u}_1^\top \mathbf{v}_n \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{u}_n^\top \mathbf{v}_1 & \mathbf{u}_n^\top \mathbf{v}_2 & \ldots & \mathbf{u}_n^\top \mathbf{v}_n \end{bmatrix}$

   Answer: $U^\top V = [\mathbf{u}_1, \ldots, \mathbf{u}_n]^\top [\mathbf{v}_1, \ldots, \mathbf{v}_n]$

   (b) Show that $UW^\top = \sum_{i=1}^{n} \mathbf{u}_i \mathbf{w}_i^\top$.

   Answer:

   (c) Show that $U^\top V = [U^\top \mathbf{v}_1, U^\top \mathbf{v}_2, \ldots, U^\top \mathbf{v}_r]$.

   Answer:

4. We say that a function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is a metric if:

    1. $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. *(symmetry)*

    2. $d(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$. *(definiteness)*

    3. $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *(triangle inequality)*.

Show that if $\| \cdot \|$ is a norm on $\mathbb{R}^n$, then $d(\cdot, \cdot)$ defined by $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$ is a metric on $\mathbb{R}^n$.

Answer: We can verify each of the properties of metric functions as follows:

    (a) $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sum_{k=1}^{n} |u_k - v_k| = \sum_{k=1}^{n} |v_k - u_k| = \|\mathbf{v} - \mathbf{u}\| = d(\mathbf{v}, \mathbf{u})$

    (b) $\Rightarrow$: We have $d(\mathbf{u}, \mathbf{v}) = 0 \implies \|u - v\| = 0 \implies \sum_{k=1}^{n} |u_k - v_k| = 0$. Since $|u_k - v_k| \geq 0$ by definition of absolute value, we must have $u_k = v_k$ for $1 \leq k \leq n$. Therefore $\mathbf{u} = \mathbf{v}$ since each corresponding entry is identical.

    $\Leftarrow$: Since $\mathbf{u} = \mathbf{v}$, $|u_k - v_k| = 0$ for $1 \leq k \leq n$. Then $\sum_{k=1}^{n} |u_k - v_k| = 0 \implies \|\mathbf{u} - \mathbf{v}\| = 0 \implies d(\mathbf{u}, \mathbf{v}) = 0$.

    (c) On the left hand side, we have $d(\mathbf{u}, \mathbf{w}) = \|\mathbf{u} - \mathbf{w}\| = \sum_{k=1}^{n} |u_k - w_k|$. On the right hand side, we have $d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}) = \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v} - \mathbf{w}\| = \sum_{k=1}^{n} |u_k - v_k| + \sum_{k=1}^{n} |v_k - w_k| = \sum_{k=1}^{n} (|u_k - v_k| + |v_k - w_k|)$. By triangle inequality of absolute values, $|u_k - v_k| + |v_k - w_k| \geq |u_k - v_k + v_k - w_k| = |u_k - w_k|$. So we have $|u_k - v_k| + |v_k - w_k| \geq |u_k - w_k|$ for $1 \leq k \leq n$, therefore $\sum_{k=1}^{n} |u_k - v_k| + \sum_{k=1}^{n} |v_k - w_k| \geq \sum_{k=1}^{n} |u_k - w_k| \implies \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{u} - \mathbf{w}\| \implies d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}) \geq d(\mathbf{u}, \mathbf{w})$.

5. Prove that $\lim_{p \to \infty} \|\mathbf{v}\|_p = \|\mathbf{v}\|_\infty$. Recall that $\| \cdot \|_\infty$ is the max-norm defined as $\|\mathbf{v}\|_\infty := \max_i |v_i|$.
This is a hard question, so let me start you off:

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p} \quad \text{Let } |v_{i^*}| = \max_i |v_i|$$

$$= \left(|v_{i^*}|^p \left(1 + \sum_{i \neq i^*} \frac{|v_i|^p}{|v_{i^*}|^p} \right) \right)^{1/p}$$

6. Suppose that:

$$A = LU \text{ where } L = \begin{bmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

Find an easy formula for the determinant $\det(A)$, in terms of the components of L and U. *(Hint: you'll need the formula, valid for any matrices B_1 and B_2, $\det(B_1 B_2) = \det(B_1)\det(B_2)$.)*

Answer: Since both L and U are triangular matrices, their respective determinant is simply the product of the diagonal entries. Then $\det(A) = \det(LU) = \det(L)\det(U) = u_{11}u_{22}u_{33}$.

7. In class we defined the condition number of an invertible square matrix as $\kappa(A) = \|A\|_2 \|A^{-1}\|_2$. Assume in addition that A is symmetric, so that A is unitarily diagonalizable. In this question you'll

2

show that $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$, where $\lambda_{\max}(\cdot)$ (resp. $\lambda_{\min}(\cdot)$) denotes the largest-in-magnitude (resp. smallest-in-magnitude) eigenvalue.

(a) Show that $\|A\|_2 = \lambda_{\max}(A)$. *(Don't overthink this, it's essentially "Special Case 1" from slide 8 of Lecture 2).*

(b) Show that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A then $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ are the eigenvalues of A^{-1}.

(c) Conclude that $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$

8. Consider the problem $A\mathbf{x} = \mathbf{b}$ where:

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & \epsilon \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

1. Compute the condition number of A, $\kappa(A)$.

2. Use Python or matlab to solve this linear system for various small value of ϵ. How does the solution found by the computer change as ϵ goes to zero?

9. Consider the least squares problem $\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^4} \|A\mathbf{x} - \mathbf{b}\|_2$ where A and $\mathbf{b}$ are given below.

$$A = \begin{bmatrix} 0.7922 & 0.6787 & 0.7060 & 0.6948 \\ 0.9595 & 0.7577 & 0.0318 & 0.3171 \\ 0.6557 & 0.7431 & 0.2769 & 0.9502 \\ 0.0357 & 0.3922 & 0.0462 & 0.0344 \\ 0.8491 & 0.6555 & 0.0971 & 0.4387 \\ 0.9340 & 0.1712 & 0.8235 & 0.3816 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0.7655 \\ 0.7952 \\ 0.1869 \\ 0.4898 \\ 0.4456 \\ 0.6463 \end{bmatrix}$$

$$Q = \begin{bmatrix} -0.4191 & -0.1861 & -0.5979 & -0.1311 & -0.2584 & -0.5901 \\ -0.5076 & -0.1321 & 0.5495 & -0.3017 & -0.5548 & 0.1553 \\ -0.3469 & -0.4172 & -0.1468 & 0.7373 & -0.0250 & 0.3739 \\ -0.0189 & -0.5331 & -0.2732 & -0.5822 & 0.3160 & 0.4494 \\ -0.4492 & -0.0950 & 0.3924 & 0.0215 & 0.7089 & -0.3637 \\ -0.4941 & 0.6933 & -0.3005 & -0.0938 & 0.1500 & 0.3919 \end{bmatrix}$$

$$R_1 = \begin{bmatrix} -1.8902 & -1.3134 & -0.8595 & -1.1681 \\ 0 & -0.6892 & 0.2859 & -0.3632 \\ 0 & 0 & -0.6673 & -0.3327 \\ 0 & 0 & 0 & 0.4674 \end{bmatrix}$$

1. Find $\mathbf{x}^*$ using the QR decomposition method (Use the Q and R given above, which are such that $A = QR$ where $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$.). Show your work.

2. Check your work by using matlab or python to solve the least squares problem.

10. Let $A \in \mathbb{R}^{m \times n}$. Show that $\|A\|_F = \sqrt{\sum_{i=1}^{n} \sigma_i(A)^2}$. You will need to use the fact that $\|\cdot\|_F$ is unitarily invariant: $\|UA\|_F = \|A\|_F$ and $\|AV\|_F = \|A\|_F$ for all unitary matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$.

11. Recall that $A^{(k)} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ is the best rank k approximation to A. Express the following quantities in terms of the singular values of A:

    1. $\|A_k\|_F$.

    2. $\|A_k\|_2$.

    3. $\|A - A_k\|_F$.

    4. $\|A - A_k\|_2$.

12. Compute the SVD of $A = \begin{bmatrix} 1 & 0 \\ 2 & 7 \\ 3 & 5 \end{bmatrix}$ by computing the eigenvalues and eigenvectors of $A^\top A$. Do this by hand and show your work.

13. In class we used PCA to analyze the "iris" data set. In this question you will repeat the analysis for another classic machine learning data set, "wine". Instead of

    ```
    iris = load_iris()
    ```

    use the command:

    ```
    wine = load_wine()
    ```

    in the Jupyter notebook from class. Make appropriate adjustments to the code from class, and attach a plot of the loadings on to the first two singular vectors of this data set. Color the data points to indicate which of the three classes each data point belongs to.