A Not-for-Profit University

Department of Electrical Engineering
School of Science and Engineering

EE212 Mathematical Foundations for Machine Learning and Data Science

ASSIGNMENT 2 – SOLUTIONS

Due Date: 23:55, Saturday. July 25, 2020 (Submit online on LMS)
Format: 12 problems, for a total of 100 marks
Instructions:

- You are not allowed to submit a group assignment. Each student must submit his/her own hand-written assignment, scanned in a single PDF document.
- You are allowed to collaborate with your peers but copying your colleague's solution is strictly prohibited. Anybody found guilty would be subjected to disciplinary action in accordance with the university rules and regulations.

Problem 1 (10 marks)

Potential customers are divided into m market segments, which are groups of customers with similar demographics, e.g., college educated women aged 25–29. A company markets its products by purchasing advertising in a set of n channels, i.e., specific TV or radio shows, magazines, web sites, blogs, direct mail, and so on. The ability of each channel to deliver impressions or views by potential customers is characterized by the $reachmatrix$, the $m \times n$ matrix R, where R_{ij} is the number of views of customers in segment i for each dollar spent on channel j. (We assume that the total number of views in each market segment is the sum of the views from each channel, and that the views from each channel scale linearly with spending.) The n-vector c will denote the company's purchases of advertising, in dollars, in the n channels. The m-vector v gives the total number of impressions in the m market segments due to the advertising in all channels. Finally, we introduce the m-vector a, where a_i gives the profit in dollars per impression in market segment i. The entries of R, c, v, and a are all nonnegative.

(a) [**2 marks**] Express the total amount of money the company spends on advertising using vector/matrix notation.

(b) [**2 marks**] Express v using vector/matrix notation, in terms of the other vectors and matrices.

(c) [**2 marks**] Express the total profit from all market segments using vector/matrix notation.

(d) [**2 marks**] How would you find the single channel most effective at reaching market segment 3, in terms of impressions per dollar spent?

(e) [**2 marks**] What does it mean if R_{35} is very small (compared to other entries of R)?

Solution:

(a) $\mathbf{1}^T c$. The company's purchases in advertising for the n channels is given by c, so summing up the entries of c gives the total amount of money the company spends across all channels.

(b) $v = Rc$. To find the total number of impressions v_i in the ith market segment, we need to sum up the impressions from each channel, which is given by an inner product of the ith row of R and the amounts c spent on advertising per channel.

(c) $a^T v = a^T Rc$. The total profit is the sum of the profits from each market segment, which is the product of the number of impressions for that segment v_i and the profit per impression for that segment a_i. The total profit is the sum $\sum_i a_i v_i = a^T v$. Substituting $v = Rc$ gives $a^T v = a^T Rc$

(d) $\text{argmax}_j\ R_{3j}$, i.e., the column index of the largest entry in the third row of R. The number of impressions made on the third market segment, for each dollar spent, is given by the third row of R. The index of the greatest element in this row is then the channel that gives the highest number of impressions per dollar spent.

(e) The fifth channel makes relatively few impressions on the third market segment per dollar spent, compared to the other channels.

Problem 2 (15 marks)

An $n \times n$ matrix A is called skew-symmetric if $A^T = -A$, i.e., its transpose is its negative. (A symmetric matrix satisfies $A^T = A$.)

(a) [**3 marks**] Find all 2×2 skew-symmetric matrices.

(b) [**3 marks**] Explain why the diagonal entries of a skew-symmetric matrix must be zero.

(c) [**5 marks**] Show that for a skew-symmetric matrix A, and any n-vector x, $(Ax) \perp x$. This means that Ax and x are orthogonal. *Hint.* First show that for any n x n matrix A and n-vector x, $x^T(Ax) = \sum_{i,j=1}^n A_{ij} x_i x_j$.

(d) [**4 marks**] Now suppose A is any matrix for which $(Ax) \perp x$ for any n-vector x. Show that A must be skew-symmetric. *Hint.* You might find the formula

$$(e_i + e_j)^T (A(e_i + e_j)) = A_{ii} + A_{jj} + A_{ij} + A_{ji},$$

valid for any $n \times n$ matrix A, useful. For $i = j$, this reduces to $e_i^T(Ae_i) = A_{ii}$.

Solution:

(a) The equality $A^T = -A$ for a 2×2 matrix means

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^T = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix} = \begin{bmatrix} -A_{11} & -A_{12} \\ -A_{21} & -A_{22} \end{bmatrix}$$

From $A_{11} = -A_{11}$, we find that $A_{11} = 0$; similarly we get $A_{22} = 0$. We also have $A_{21} = -A_{12}$ and $A_{12} = -A_{21}$. These are the same equation; we can have any value of A_{12}, say α, and then we have

$$A = \alpha \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

as the general form of a skew-symmetric 2×2 matrix.

(b) Since $A^T = -A$, we have $(A^T)_{ii} = (-A)_{ii}$. But the left-hand side is A_{ii}, so we have $A_{ii} = -A_{ii}$. This implies that $A_{ii} = 0$.

(c) First we establish the formula given in the link:

$$x^T(Ax) = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} A_{ij} x_j = \sum_{i,j=1}^{n} A_{ij} x_i x_j.$$

Now suppose that A is skew-symmetric. This means that (by part (b)) $A_{ii} = 0$. Also, $A_{ij} = -A_{ji}$. In the sum above, we lump the term $A_{ij} x_i x_j$ and $A_{ji} x_j x_i$ together, and we see that they cancel each other, i.e., add to zero. So when we sum over all i and j, we get zero, and it follows that $x^T(Ax) = 0$, which means that $(Ax) \perp x$.

(d) For $x = e_i$, the property $(Ax) \perp x$ gives $A_{ii} = e_i^T A e_i = 0$. All the diagonal entries are zero. Applying the same property with $x = e_i + e_j$ and using the expression in the hint then gives

$$0 = (e_i + e_j)^T A (e_i + e_j) = A_{ij} + A_{ji}.$$

We conclude that if $(Ax) \perp x$ for all x, then $A_{ij} = -Aji$ for all i and j. In other words, A is skew-symmetric.

Problem 3 (5 marks)

Suppose A is an $m \times n$ matrix and x is an n-vector. A famous inequality relates $\|x\|$, $\|A\|$, and $\|Ax\|$:

$$\|Ax\| \leq \|A\| \|x\|$$

The left-hand side is the (vector) norm of the matrix-vector product; the right-hand side is the (scalar) product of the matrix and vector norms. Show this inequality. *Hints.* Let a_i^T be the ith row of A. Use the Cauchy–Schwarz inequality to get $(a_i^T x)^2 \leq \|a_i\|^2 \|x\|^2$. Then add the resulting m inequalities.

Solution: Following the hint we get

$$\|Ax\|^2 = \sum_{i=1}^{m} (a_i^T x)^2 \leq \sum_{i=1}^{m} \|a_i\|^2 \|x\|^2 = \|A\|^2 \|x\|^2$$

using the fact that the sum of the squared norms of the rows of a matrix is the squared norm of the matrix. Taking the squareroot gives the inequality.

Problem 4 (5 marks)

We consider a set of n basic foods (such as rice, beans, apples) and a set of m nutrients or components (such as protein, fat, sugar, vitamin C). Food j has a cost given by c_j (say, in dollars per gram), and contains an amount N_{ij} of nutrient i (per gram). (The nutrients are given in some appropriate units, which can depend on the particular nutrient.) A daily diet is represented by an n-vector d, with d_i the daily intake (in grams) of food i. Express the condition that a diet d contains the total nutrient amounts given by the m-vector n^{des}, and has a total cost B (the budget) as a set of linear equations in the variables $d_1, \ldots, d_n$. (The entries of d must be non-negative, but we ignore this issue here.)

Solution:

$$\begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1n} \\ N_{21} & N_{22} & \cdots & N_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ N_{m1} & N_{m2} & \cdots & N_{mn} \\ c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} = \begin{bmatrix} n_1^{des} \\ n_2^{des} \\ \vdots \\ n_m^{des} \\ B \end{bmatrix}$$

Problem 5 (4 marks)

Suppose that A is an $m \times n$ matrix, D is a diagonal matrix, and $B = DA$. Describe B in terms of A and the entries of D. You can refer to the rows or columns or entries of A.

Solution: The ith row of B is the ith row of A, scaled by D_{ii}. To see, consider the i, j entry of $B = DA$:

$$B_{ij} = \sum_{k=1}^{m} D_{ik} A_{kj} = D_{ii} Aij$$

The first equality is the definition of a general matrix-matrix product. The second equality follows because $D_{ik} = 0$ for $k \neq i$.

Problem 6 (5 marks)

The sum of the diagonal entries of a square matrix is called the trace of the matrix, denoted $\mathbf{tr}(A)$.Suppose A and B are $m \times n$ matrices. Show that

$$\mathbf{tr}(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}$$

Solution: The diagonal entries of $A^T B$ are

$$(A^T B)_{jj} = \sum_{i=1}^{m} (A^T)_{ji} B_{ij} = \sum_{i=1}^{m} A_{ij} B_{ij}$$

And hence, the trace is given by the diagonal terms summed up over n since $A^T B$ is an $n \times n$ matrix.

$$\mathbf{tr}(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}$$

Problem 7 (6 marks)

We consider m students, n classes, and p majors. Each student can be in any number of the classes (although we'd expect the number to range from 3 to 6), and can have any number of the majors (although the common values would be 0, 1, or 2). The data about the students' classes and majors are given by an m x n matrix C and an $m \times p$ matrix M, where

$$C_{ij} = \begin{cases} 1 & \text{student } i \text{ is in class } j \\ 0 & \text{student } i \text{ is not in class } j, \end{cases}$$

and

$$M_{ij} = \begin{cases} 1 & \text{student } i \text{ is in major } j \\ 0 & \text{student } i \text{ is not in major } j, \end{cases}$$

(a) **[3 marks]** Let E be the n-vector with E_i being the enrollment in class i. Express E using matrix notation, in terms of the matrices C and M.

(b) **[3 marks]** Define the $n \times p$ matrix S where S_{ij} is the total number of students in class i with major j. Express S using matrix notation, in terms of the matrices C and M.

Solution:

(a) $E = C^T \mathbf{1}$. This follows from

$$E_i = \sum_{k=1}^{n} C_{ki} = \sum_{k=1}^{n} (C^T)_{ik} = (C^T \mathbf{1})_i.$$

(b) $S = C^T M$. This follows from

$$S_{ij} = \sum_{k=1}^{n} C_{ki} M_{kj} = \sum_{k=1}^{n} (C^T)_{ik} M_{kj} = (C^T M)_{ij}.$$

Problem 8 (10 marks)

Let $G \in \mathbf{R}^{m \times n}$ represent a contingency matrix of m students who are members of n groups:

$$G_{ij} = \begin{cases} 1 & \text{student } i \text{ is in group j} \\ 0 & \text{student } i \text{ is not in group j} \end{cases}$$

(A student can be in any number of the groups.)

(a) [**2 marks**] What is the meaning of the 3rd column of G?

(b) [**2 marks**] What is the meaning of the 15th row of G?

(c) [**2 marks**] Give a simple formula (using matrices, vectors, etc.) for the n-vector M, where M_i is the total membership (i.e., number of students) in group i.

(d) [**2 marks**] Interpret $(GG^T)_{ij}$ in simple English.

(e) [**2 marks**] Interpret $(G^T G)_{ij}$ in simple English.

Solution:

(a) The roster for the 3rd group.

(b) The list of groups student 15 is in.

(c) $M = G^T \mathbf{1}$.

(d) $(GG^T)_{ij}$ is the number of groups student i and student j have in common. (Sanity check: $GG^T \in \mathbf{R}^{m \times m}$)

(e) $(G^T G)_{ij}$ is the number of students group i and group j have in common. (Sanity check: $G^T G \in \mathbf{R}^{n \times n}$)

Problem 9 (5 marks)

Let Z be a tall $m \times n$ matrix with linearly independent columns, and let X and Y be left inverses of Z. Show that for any scalars α and β satisfying $\alpha + \beta = 1$, $\alpha X + \beta Y$ is also a left inverse of Z. It follows that if a matrix has two different left inverses, it has an infinite number of different left inverses.

Solution: We are asked to show that $(\alpha X + \beta Y)Z = I$.

$$(\alpha X + \beta Y)Z = \alpha XZ + \beta YZ = \alpha I + \beta I = (\alpha + \beta)I = I.$$

Problem 10 (10 marks)

Find the inverse of the $n \times n$ running sum matrix,

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

Does your answer make sense?

Solution: The inverse is

$$S^-1 = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

The matrix S is triangular. To find thte inverse, we solve $SX = I$, column by column, by forward substitution. The equations for column j are

$$\begin{aligned} X_{1j} &= 0 \\ X_{1j} + X_{2j} &= 0 \\ &\vdots \\ X_{1j} + X_{2j} + \cdots + X_{j-1,j} &= 0 \\ X_{1j} + X_{2j} + \cdots + X_{jj} &= 1 \\ X_{1j} + X_{2j} + \cdots + X_{j+1,j} &= 0 \\ &\vdots \\ X_{1j} + X_{2j} + \cdots + X_{nj} &= 0 \end{aligned}$$

IF $j < n$, then forward substitution gives

$$X_{1j} = \cdots = Xj - 1, j = 0, \qquad X_{jj} = 1, \qquad X_{j+1,j} = -1 \qquad X_{j+2,j} = \cdots = Nnj = 0.$$

If $j = n$, the solution is $X_{1n} = \cdots = X_{n-1,n} = 0$ and $X_{nn} = 1$.
The matrix S^{-1} is the difference matrix, with an additional top row that is e_1^T. We have

$$S^{-1}y = (y_1, y_2 - y_1, \dots, y_n - y_{n-1}).$$

This makes sense: to undo a running sum (which is like an integral), you should differentiate, which is what S^{-1} (sort of) does.

Problem 11 (5 marks)

Suppose A is an $n \times p$ matrix and B is a $p \times n$ matrix, so $C = AB$ makes sense. Explain why C cannot be invertible if A is tall and B is wide, i.e., if $p < n$. *Hint.* First argue that the columns of B must be linearly dependent.

Solution: By the independence-dimension inequality, the columns of B are linearly dependent, since they have dimension p and there are n of them, and $n > p$. So there is a nonzero n-vector x that satisfies $Bx = 0$. This implies that $(AB)x = A(Bx) = A0 = 0$. This shows that the columns of $C = AB$ are linearly dependent, and so C is not invertible.

Problem 12 (20 marks)

Let A be $n \times n$ symmetric matrix which can be written as $A = S\Lambda S^T$ using EVD or as $A = U\Sigma V^T$ using SVD.

(a) [**4 marks**] Assuming all eigenvalues of A are positive i.e. $\lambda_i \geq 0$, show that $S = V = U$.

(b) [**16 marks**] For an $m \times n$ matrix A, following information is known:

- $A^T A$ has 9 and 4 as its two non-zero eigenvalues and one of the two normalised eigenvectors is $\frac{1}{\sqrt{5}}\begin{bmatrix}1 & 2\end{bmatrix}^T$
- It has SVD as $U\Sigma V^T$ with matrix U having its third column equal to $u_3 = \begin{bmatrix}2/3 & -1/3 & -2/3\end{bmatrix}^T$
- It is also given $Av_1 = \frac{1}{\sqrt{5}}\begin{bmatrix}5 & 2 & 4\end{bmatrix}^T$ and $Av_2 = \frac{2}{\sqrt{5}}\begin{bmatrix}0 & 2 & -1\end{bmatrix}^T$ where v_1 and v_2 are first two columns of matrix V

i. What is the size of A?

ii. Write down U matrix.

iii. Write down V matrix.

iv. Write down Σ matrix.

Solution: A is symmetric so $A = A^T$, which means:

$$AA^T = A^T A = A^2 \tag{1}$$

(a) We begin by taking SVD of AA^T, we know that $A = U\Sigma V^T$ and $A^T = V\Sigma U^T$. We also know that $A = S\Lambda S^T$, using (1) we get:

$$AA^T = U(\Sigma)^2 U^T = A^T A = V(\Sigma)^2 V^T = A^2 = S(\Lambda)^2 S^T$$

Since $(\Sigma)^2 = (\Lambda)^2$, we can safely say that $U = V = S$ by comparison.

(b) i. We know that A is an $m \times n$. From the vectors given we can deduce: $u_3 \in \mathbf{R}^3$ so $U : 3 \times m$, similarly $v_1 \in \mathbf{R}^2$ and so $V : 2 \times n$ (and $V^T : n \times 2$). We can now write the SVD of A as follows:

$$A = U_{3\times m}\Sigma_{m\times n}V^T_{n\times 2}$$

so A is of size 3×2

ii. From SVD we know $u_1 = \frac{1}{\sqrt{\lambda_1}} Av_1$

$$\lambda_1 = 9 : \; u_1 = \frac{1}{3} \times \frac{1}{\sqrt{5}}\begin{bmatrix}5 & 2 & 4\end{bmatrix}^T$$

$$u_1 = \frac{1}{3\sqrt{5}}\begin{bmatrix}5 & 2 & 4\end{bmatrix}^T$$

$$\lambda_2 = 4 : \; u_2 = \frac{1}{2} \times \frac{2}{\sqrt{5}}\begin{bmatrix}0 & 2 & -1\end{bmatrix}^T$$

$$u_1 = \frac{1}{\sqrt{5}}\begin{bmatrix}0 & 2 & -1\end{bmatrix}^T$$

Now that we have all columns of U, we can merge them together to give the final matrix:

$$U = \begin{bmatrix} \frac{5}{3\sqrt{5}} & 0 & \frac{2}{3} \\ \frac{2}{3\sqrt{5}} & \frac{2}{\sqrt{5}} & -\frac{1}{3} \\ \frac{4}{3\sqrt{5}} & -\frac{1}{\sqrt{5}} & -\frac{2}{3} \end{bmatrix}$$

iii. We are given $v_1 = \frac{1}{\sqrt{5}}\begin{bmatrix}1 & 2\end{bmatrix}^T$. Since V is an orthogonal matrix $v_2^T \cdot v_1 = 0$ applies. Lets take an arbitrary $v_2 = \begin{bmatrix}a & b\end{bmatrix}^T$. This gives:

$$v_2^T \cdot v_1 = \frac{1}{\sqrt{5}}(a+2b) = 0$$
$$a + 2b = 0$$
$$a = -2b$$

Since each vector in V is orthonormal, we take norm over v_2:

$$a^2 + b^2 = 1$$
$$4b^2 + b^2 = 1$$
$$b = \tfrac{1}{\sqrt{5}} \text{ and } a = -\tfrac{2}{\sqrt{5}}$$
$$v_2 = \tfrac{1}{\sqrt{5}} \begin{bmatrix} -2 & 1 \end{bmatrix}^T$$

Final matrix is given by

$$V = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

iv. Σ is of the same size as A, i.e. 3×2, plus it has singular values along its diagonal (rest are zero)

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \\ 0 & 0 \end{bmatrix}$$
$$\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 4 \\ 0 & 0 \end{bmatrix}$$

— End of Assignment —