LUMS
A Not-for-Profit University

Department of Electrical Engineering
School of Science and Engineering

EE212 Mathematical Foundations for Machine Learning and Data Science

ASSIGNMENT 1

Due Date: 23:55, Saturday. July 18, 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 (4 marks)

Let $u = \begin{bmatrix} 5 \\ 3 \\ -4 \end{bmatrix}$, $v = \begin{bmatrix} -1 \\ 5 \\ 2 \end{bmatrix}$. $w = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}$. Find:

(a) [**2 marks**] $5u - 2v$

(b) [**2 marks**] $-2u + 4v - 3w$

Solution.

(a) $\begin{bmatrix} 27 \\ 5 \\ -24 \end{bmatrix}$.

(b) $\begin{bmatrix} -23 \\ 17 \\ 22 \end{bmatrix}$.

Problem 2 (5 marks)

Find x, y, z such that $(x - y, x + y, z - 1) = (4, 2, 3)$.

Solution.

By definition of equality of vectors, corresponding entries must be equal. Thus,

$$x - y = 4; \quad x + y = 2; \quad z - 1 = 3$$

Solving the above system of equations yields $x = 3, y = -1, z = 4$.

Problem 3 (6 marks)

Let $u = (1, -2, 3)$, $v = (4, 5, -1)$ and $w = (2, 7, 4)$. Then, find:

(a) **[2 marks]** $u^T v$

(b) **[2 marks]** $u^T w$

(c) **[2 marks]** $v^T w$

Solution.

(a) -9

(b) 0

(c) 39

Problem 4 (10 marks)

Normalize the following (i.e. find the divided-by-norm vectors for the following):

(a) **[3 marks]** $u = (3, -4)$

(b) **[3 marks]** $v = (4, -2, -3, 8)$

(c) **[4 marks]** $w = (\frac{1}{2}, \frac{2}{3}, -\frac{1}{4})$

Solution.

(a) $(\frac{3}{5}, -\frac{4}{5})$

(b) $(\frac{4}{\sqrt{93}}, -\frac{2}{\sqrt{93}}, -\frac{3}{\sqrt{93}}, \frac{8}{\sqrt{93}})$

(c) $(\frac{6}{\sqrt{109}}, \frac{8}{\sqrt{109}}, -\frac{3}{\sqrt{109}})$

Problem 5 (5 marks)

Suppose $u = (1, -2, 3)$ and $v = (2, 4, 5)$. Find the angle between u and v.

Solution.

$cos\theta = \frac{9}{\sqrt{14}\sqrt{45}}$

Problem 6 (10 marks)

This question will help you to interpret sparsity. Suppose the n-vector x is sparse, i.e., has only a few non-zero entries. Give a short sentence or two explaining what this means in each of the following contexts.

(a) **[2 marks]** x represents the daily cash flow of some business over n days.

(b) **[2 marks]** x represents the annual dollar value purchases by a customer of n products or services.

(c) **[2 marks]** x represents a portfolio, say, the dollar value holdings of n stocks.

(d) **[2 marks]** x represents a bill of materials for a project, i.e., the amounts of n materials needed.

(e) **[2 marks]** x represents a monochrome image, i.e., the brightness values of n pixels.

Solution.

(a) On most days the business neither receives nor makes cash payments.

(b) The customer has purchased only a few of the n products.

(c) The portfolio invests (long or short) in only a small number of stocks.

(d) The project requires only a few types of materials.

(e) Most pixels have a brightness value of zero, so the image consists of mostly black space. This is the case, for example, in astronomy, where many imaging methods leverage the assumption that there are a few light sources against a dark sky.

Problem 7 (10 marks)

In a study to determine the most prevalent symptoms of Covid-19 researchers combined data from various medical institutes to form a 10-vector s. This vector records whether each of 10 different symptoms is present in a Covid patient, with $s_i = 1$ meaning the patient has the symptom and $s_i = 0$ meaning he/she does not. Express the following using vector notation.

(a) [**4 marks**] The total number of symptoms the patient has

(b) [**6 marks**] The patient exhibits two out of the first five symptoms

Solution.

(a) The total number of symptoms the patient has is $\mathbf{1}^T s$

(b) The patient exhibits two out of the first five symptoms can be expressed as $a^T s = 2$, where the 10-vector a is given by $a = (\mathbf{1}_5, \mathbf{0}_5)$. (The subscripts give the dimensions.)

Problem 8 (10 marks)

Suppose the n-vector w is the word count vector associated with a document and a dictionary of n words. For simplicity we will assume that all words in the document appear in the dictionary.

(a) [**3 marks**] What is $\mathbf{1}^T w$?

(b) [**3 marks**] What does $w_{282} = 0$ mean?

(c) [**4 marks**] Let h be the n-vector that gives the histogram of the word counts, i.e., h_i is the fraction of the words in the document that are word i. Use vector notation to express h in terms of w . (You can assume that the document contains at least one word.)

Solution.

(a) $\mathbf{1}^T w$ is total number of words in the document.

(b) $w_{282} = 0$ means that word 282 does not appear in the document.

(c) $h = w/(\mathbf{1}^T w)$. We simply divide the word count vector by the total number of words in the document.

Problem 9 (5 marks)

Consider the 5 Wikipedia pages in table 1 given below.

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	0.095	0.130	0.153	0.170
Memorial Day	0.095	0	0.122	0.147	0.164
Academy A.	0.130	0.122	0	0.108	0.164
Golden Globe A.	0.153	0.147	0.108	0	0.181
Super Bowl	0.170	0.164	0.164	0.181	0

Table 1: Pairwise word count histogram distances between five Wikipedia articles

What is the nearest neighbor of (the word count histogram vector of) 'Veterans Day' among the others? Does the answer make sense?

Solution.

Its nearest neighbor is the histogram vector associated with 'Memorial Day'. (We can see this by scanning the first row or column of table 1.) Yes, it makes sense — these two pages describe holidays, whereas the others do not.

Problem 10 (15 marks)

Let a and b be different n-vectors. The line passing through a and b is given by the set of vectors of the form $(1-\theta)a + \theta b$, where θ is a scalar that determines the particular point on the line. Let x be any n-vector. Find a formula for the point p on the line that is closest to x. The point p is called the projection of x onto the line. Show that $(p - x) \perp (a - b)$, and draw a simple picture illustrating this in 2-D.

Hint: Work with the square of the distance between a point on the line and x, i.e., $||(1-\theta)a + \theta b - x||^2$. Expand this, and minimize over θ.

Solution.

The squared distance of p to x is

$$\begin{aligned} ||(1-\theta)a + \theta b - x||^2 &= ||a - x + \theta(b-a)||^2 \\ &= ||a-x||^2 + 2\theta(a-x)^T(b-a) + \theta^2||b-a||^2 \end{aligned}$$

This is a quadratic function of θ. We can minimize it using ordinary calculus by setting the derivative to zero:

$$2(a-x)^T(b-a) + 2\theta||b-a||^2 = 0$$

which yield $\theta = (x-a)^T(b-a)/||b-a||^2$. (Recall that $b \neq a$, so the denominator is nonzero). This yields the closest point on the line,

$$p = a + \tfrac{(x-a)^T(b-a)}{||b-a||^2}(b-a)$$

Let us show that $(p-x) \perp (a-b)$:

$$\begin{aligned} (p-x)^T(a-b) &= (a + \frac{(x-a)^T(b-a)}{||b-a||^2}(b-a))^T(a-b) \\ &= (a-x)^T(a-b) + \frac{(x-a)^T(b-a)}{||b-a||^2}(b-a)^T(a-b) \\ &= (a-x)^T(a-b) - \frac{(x-a)^T(b-a)}{||b-a||^2}||b-a||^2 \\ &= (a-x)^T(a-b) - (x-a)^T(b-a) \\ &= 0 \end{aligned}$$

On the third line we used $(b-a)^T(a-b) = -||b-a||^2$.

Problem 11 (15 marks)

Suppose x is an n-vector and α and β are scalars.

(a) **[7 marks]** Show that $\mathbf{avg}(\alpha x + \beta \mathbf{1}) = \alpha\ \mathbf{avg}(x) + \beta$

(b) **[8 marks]** Show that $\mathbf{std}(\alpha x + \beta \mathbf{1}) = |\alpha|\ \mathbf{std}(x)$

Solution.
Suppose x is an n-vector and α, β are scalars.

(a) First we find the average of $\alpha x + \beta \mathbf{1}$:

$$\mathbf{avg}(\alpha x + \beta \mathbf{1}) = \mathbf{1}^T(\alpha x + \beta \mathbf{1})/\text{n} = (\alpha \mathbf{1}^T x + \beta \mathbf{1}^T \mathbf{1})/\text{n} = \alpha\ \mathbf{avg}(x) + \beta,$$

Where we use $\mathbf{1}^T\mathbf{1} = n$

(b) Using the definition of $\mathbf{std}(x)$ and part (a),

$$\begin{aligned}\mathbf{std}(\alpha x + \beta \mathbf{1}) &= \mathbf{rms}(\alpha x + \beta \mathbf{1} - (\alpha \mathbf{avg}(x) + \beta)\mathbf{1}) \\ &= \mathbf{rms}(\alpha x - \alpha \mathbf{avg}(x)\mathbf{1}) \\ &= |\alpha| \mathbf{rms}(x - \mathbf{avg}(x)\mathbf{1}) \\ &= |\alpha| \mathbf{std}(x)\end{aligned}$$

Problem 12 (5 marks)

An intern at a quantitative hedge fund examines the daily returns of around 400 stocks over one year (which has 250 trading days). She tells her supervisor that she has discovered that the returns of one of the stocks, Packages Ltd., can be expressed as a linear combination of the others, which include many stocks that are unrelated to Packages Ltd. (say, in a different type of business or sector).

Her supervisor then says: "It is overwhelmingly unlikely that a linear combination of the returns of unrelated companies can reproduce the daily return of Packages Ltd. So you've made a mistake in your calculations."

Is the supervisor right? Did the intern make a mistake? Give a very brief explanation.

Solution.

The supervisor is wrong, and the intern is most likely correct. She is examining 400 250-vectors, each representing the daily returns of a particular stock over one year. By the independence-dimension inequality, any set of $n + 1$ or more n vectors is linearly dependent, so we know that this set of vectors is linearly dependent. That is, there exist some vectors that can be expressed as a linear combination of the others. It is quite likely that the returns of any particular stock, such as Packages, can be expressed as a linear combination of the returns other stocks.

Even if Packages Ltd's return can be expressed as a linear combination of the others, by the independence-dimension inequality, this fact is not as useful as it might seems. For example, Packages Ltd's future returns would not be given by the same linear combination of other asset returns. So although the intern is right, and the supervisor is wrong, the observation cannot be monetized.

— End of Assignment —