![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**] $-2u + 4v - 3w$

Problem 2 (5 marks)
Find x, y, z such that $(x - y, x + y, z - 1) = (4, 2, 3)$.

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**] $v^T w$

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

Problem 5 (5 marks)

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

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.

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

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.)

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?

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.

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 $\mathbf{std}(\alpha x + \beta\mathbf{1}) = |\alpha|\mathbf{std}(x)$

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. So you've made a mistake in your calculations."

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

— End of Assignment —