



College of Engineering & Applied Sciences

CSPB 2820

Linear Algebra With Computer Science Applications

Study Guide 3 - Norm And Distance

UNIVERSITY OF COLORADO

2024

Linear Algebra With Computer Science Applications - Study Guide 3 - Norm And Distance

1 Study Guide 3	3
Norm And Distance.....	3
Problem 1.....	4
Problem 2.....	7
Problem 3.....	10
Problem 4.....	12
Problem 5.....	15
Problem 6.....	18
Problem 7.....	20
Problem 8.....	22
Problem 9.....	25



Study Guide 3

Norm And Distance

Study Guide Instructions

- Submit your work in Gradescope as a PDF - you will identify where your “questions are.”
- Identify the question number as you submit. Since we grade "blind" if the questions are NOT identified, the work WILL NOT BE GRADED and a 0 will be recorded. Always leave enough time to identify the questions when submitting.
- One section per page (if a page or less) - We prefer to grade the main solution in a single page, extra work can be included on the following page.
- Long instructions may be removed to fit on a single page.
- **Do not start a new question in the middle of a page.**
- Solutions to book questions are provided for reference.
- You may NOT submit given solutions - this includes minor modifications - as your own.
- Solutions that do not show individual engagement with the solutions will be marked as no credit and can be considered a violation of honor code.
- If you use the given solutions you must reference or explain how you used them, in particular...

Method Selection

For full credit, EACH book exercise in the Study Guides must use one or more of the following methods and FOR EACH QUESTION. Identify the number the method by number to ensure full credit.

- **Method 1** - Provide original examples which demonstrate the ideas of the exercise in addition to your solution.
- **Method 2** - Include and discuss the specific topics needed from the chapter and how they relate to the question.
- **Method 3** - Include original Python code, of reasonable length (as screenshot or text) to show how the topic or concept was explored.
- **Method 4** - Expand the given solution in a significant way, with additional steps and comments. All steps are justified. This is a good method for a proof for which you are only given a basic outline.
- **Method 5** - Attempt the exercise without looking at the solution and then the solution is used to check work. Words are used to describe the results.
- **Method 6** - Provide an analysis of the strategies used to understand the exercise, describing in detail what was challenging, who helped you or what resources were used. The process of understanding is described.

Problem 1

Problem Statement

Select one page or section of Chapter Three of VMLS to annotate. Include a screenshot of your annotation here.

For Chapter Three of VMLS, I am annotating **page 57** of VMLS. The annotations for this problem can be seen on the following page.



3.4 Angle

57

which looks more intimidating. This inequality is attributed to the mathematician Augustin-Louis Cauchy; Hermann Schwarz gave the derivation given below.

The Cauchy-Schwarz inequality can be shown as follows. The inequality clearly holds if $a = 0$ or $b = 0$ (in this case, both sides of the inequality are zero). So we suppose now that $a \neq 0$, $b \neq 0$, and define $\alpha = \|a\|$, $\beta = \|b\|$. We observe that

$$\begin{aligned} 0 &= 2\|a\|^2\|b\|^2 - 2\|a\|\|b\|\alpha^T b \\ 2\|a\|^2\|b\|^2 &= 2\|a\|\|b\|\alpha^T b \\ \|a\|\|b\| &\geq \alpha^T b, \alpha^T b \leq \|a\|\|b\| \end{aligned} \quad \begin{aligned} 0 &\leq \|\beta a - \alpha b\|^2 \\ &= \|\beta a\|^2 - 2(\beta a)^T(\alpha b) + \|\alpha b\|^2 \\ &= \beta^2\|a\|^2 - 2\beta\alpha(a^T b) + \alpha^2\|b\|^2 \\ &= \|b\|^2\|a\|^2 - 2\|b\|\|a\|(\alpha^T b) + \|a\|^2\|b\|^2 \\ &= 2\|a\|^2\|b\|^2 - 2\|a\|\|b\|(\alpha^T b). \end{aligned}$$

Dividing by $2\|a\|\|b\|$ yields $\alpha^T b \leq \|a\|\|b\|$. Applying this inequality to $-a$ and b we obtain $-\alpha^T b \leq \|a\|\|b\|$. Putting these two inequalities together we get the Cauchy-Schwarz inequality, $|\alpha^T b| \leq \|a\|\|b\|$.

This argument also reveals the conditions on a and b under which they satisfy the Cauchy-Schwarz inequality with equality. This occurs only if $\|\beta a - \alpha b\| = 0$, i.e., $\beta a = \alpha b$. This means that each vector is a scalar multiple of the other (in the case when they are nonzero). This statement remains true when either a or b is zero. So the Cauchy-Schwarz inequality holds with equality when one of the vectors is a multiple of the other; in all other cases, it holds with strict inequality.

Verification of triangle inequality. We can use the Cauchy-Schwarz inequality to verify the triangle inequality. Let a and b be any vectors. Then

$$\begin{aligned} \|a+b\|^2 &= \|a\|^2 + 2a^T b + \|b\|^2 \\ a^T b &= \|a\|\|b\| \longrightarrow \leq \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 \\ &= (\|a\| + \|b\|)^2 \end{aligned} \quad \begin{aligned} &\text{When } \theta = \pi/2 \rightarrow \text{Pythag. Thm.} \\ &x^2 + 2xy + y^2 = (x+y)^2 \end{aligned}$$

where we used the Cauchy-Schwarz inequality in the second line. Taking the squareroot we get the triangle inequality, $\|a+b\| \leq \|a\| + \|b\|$.

Angle between vectors. The angle between two nonzero vectors a , b is defined as

$$\cos^{-1}(\theta) \equiv \arccos(\theta) \rightarrow \theta = \arccos\left(\frac{a^T b}{\|a\|\|b\|}\right)$$

where \arccos denotes the inverse cosine, normalized to lie in the interval $[0, \pi]$. In other words, we define θ as the unique number between 0 and π that satisfies

$$a^T b = \|a\|\|b\|\cos\theta.$$

The angle between a and b is written as $\angle(a, b)$, and is sometimes expressed in degrees. (The default angle unit is radians; 360° is 2π radians.) For example, $\angle(a, b) = 60^\circ$ means $\angle(a, b) = \pi/3$, i.e., $a^T b = (1/2)\|a\|\|b\|$.

The angle coincides with the usual notion of angle between vectors, when they have dimension two or three, and they are thought of as displacements from a

$0 < \theta < \pi/2$
↳ Acute
 $\pi/2 < \theta < \pi$
↳ obtuse

$$\begin{aligned} a^T b &= \|a\|\|b\|\cos(\theta) \rightarrow \frac{a^T b}{\|a\|\|b\|} = \cos(\theta) \rightarrow \theta = \cos^{-1}\left(\frac{a^T b}{\|a\|\|b\|}\right) \\ \theta &= \frac{n\pi}{2} \text{ w/ } n = 2k+1 \Rightarrow \perp \text{ or perpendicular, orthogonal} \\ \theta &= n\pi \text{ w/ } n = 2k \Rightarrow \parallel \text{ or parallel, aligned} \end{aligned}$$

Problem 1 Summary

Procedure

- Annotate a page of the textbook and highlight the important aspects from the page

Key Concepts

- This page encapsulates important aspects of the Triangle Inequality
- This page also encapsulates how to calculate the angle between two vectors

Variations

- We could be asked to annotate a different section or page of the textbook



Problem 2

Problem Statement

Solve the Chapter 3 Random exercise from the video and Piazza in your own words here.

Original Question:

Norm of sum. Use formulas (3.1) and (3.6) to show the following:

- (a) $a \perp b$ if and only if $\|a + b\| = \sqrt{\|a\|^2 + \|b\|^2}$.
- (b) Nonzero vectors a and b make an acute angle if and only if $\|a + b\| > \sqrt{\|a\|^2 + \|b\|^2}$.
- (c) Nonzero vectors a and b make an obtuse angle if and only if $\|a + b\| < \sqrt{\|a\|^2 + \|b\|^2}$.

Draw a picture illustrating each case in 2-D.

Solution - Part (a)

For this problem I will be using **Method 4**.

Equation (3.1) from VMLS states that for two vectors a and b

$$\|a + b\| = \sqrt{\|a\|^2 + 2a^T b + \|b\|^2}. \quad (1)$$

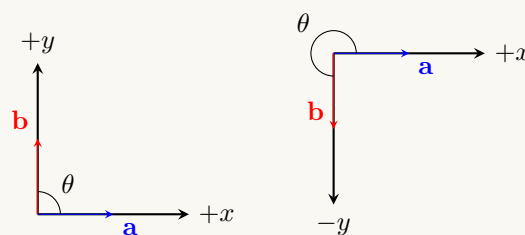
Equation (3.6) from VMLS states that for two vectors a and b

$$\|a + b\|^2 = \|a\|^2 + 2\|a\|\|b\|\cos(\theta) + \|b\|^2. \quad (2)$$

Taking the square root on each side of equation (2) we have

$$\|a + b\| = \sqrt{\|a\|^2 + 2\|a\|\|b\|\cos(\theta) + \|b\|^2}. \quad (3)$$

Two vectors are said to be orthogonal if the angle (θ) between them is $n(90)^\circ$ or $\frac{n\pi}{2}$ radians with $n = 2k + 1$. Graphically this can be seen below



In the above image, we see that the angle between these two vectors is $\frac{n\pi}{2}$ radians with $n = 2k + 1$. If we insert this angle into equation (3) we then see that

$$\|a + b\| = \sqrt{\|a\|^2 + 2\|a\|\|b\|\cos(n\pi/2) + \|b\|^2} \quad (\text{Equation (3) with } \theta = n\pi/2) \quad (4)$$

$$= \sqrt{\|a\|^2 + 2\|a\|\|b\|(0) + \|b\|^2} \quad (\cos(n\pi/2) = 0) \quad (5)$$

$$= \sqrt{\|a\|^2 + \|b\|^2} \quad (\text{Simplification by multiplication}) \quad (6)$$

From equation (6), we can see that this is only possible if $\theta = \frac{n\pi}{2}$ with $n = 2k + 1$ and k being a nonnegative integer. Therefore we can say $\|a + b\| = \sqrt{\|a\|^2 + \|b\|^2}$ if and only if $a \perp b$.

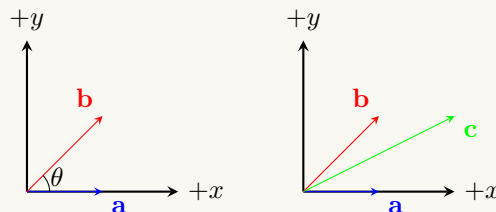
Solution - Part (b)

For this problem I will be using **Method 4**.

Re-using the results from part (a) of this problem, we can re-write equation (3.1) from VMLS as

$$\|a + b\| = \sqrt{\|a\|^2 + 2\|a\|\|b\|\cos(\theta) + \|b\|^2}. \quad (7)$$

An angle, in this case we will represent it as θ , is said to be acute if that angle $0 < \theta < \pi/2$ radians. In the image below, we see an example of an acute angle as well as the superposition of two vectors a and b .



For angles $0 < \theta < \pi/2$ radians we can say $0 < \cos(\theta) < 1$. Letting $\alpha = 2\|a\|\|b\|\cos(\theta)$ this means we can say $\approx 2\|a\|\|b\| < \alpha < 0$. Regardless of the angle for θ we can say $\alpha > 0$. Intuitively, we can see that the length of the vector c (defined as $a + b$) is greater than that of the sum of the two vectors a and b 's lengths. Mathematically this can be seen as

$$\|a + b\| > \sqrt{\|a\|^2 + \|b\|^2}. \quad (8)$$

This of course only is possible if the angle between the vectors a and b is acute ($0 < \theta < \pi/2$).

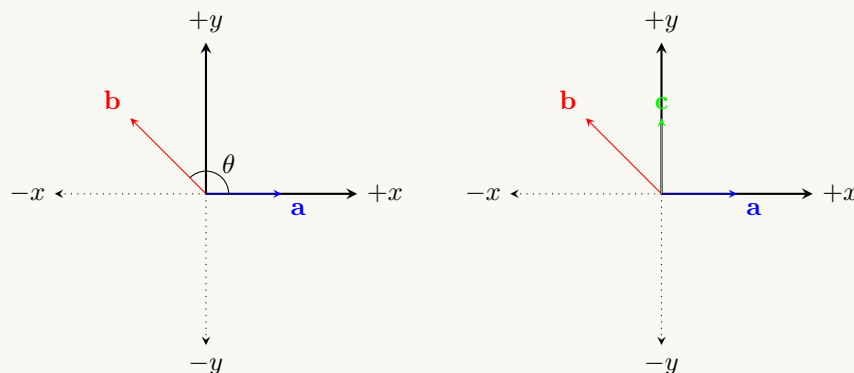
Solution - Part (c)

For this problem I will be using **Method 4**.

We once again use equation (3.1) from VMLS in a new form

$$\|a + b\| = \sqrt{\|a\|^2 + 2\|a\|\|b\|\cos(\theta) + \|b\|^2}. \quad (9)$$

An angle θ is said to be obtuse if $\pi/2 < \theta < \pi$ radians. The image below shows an example of an obtuse angle and the superposition of the two vectors a and b .



For angles $\pi/2 < \theta < \pi$ radians we can say $-1 < \cos(\theta) < 0$. Letting $\alpha = 2\|a\|\|b\|\cos(\theta)$ this means we can say $\approx -2\|a\|\|b\| < \alpha < 0$. Regardless of the angle for θ we can say $\alpha < 0$. Intuitively, we can see that the length of the vector c (defined as $a + b$) is less than that of the sum of the two vectors a and b 's lengths. Mathematically this can be seen as

$$\|a + b\| < \sqrt{\|a\|^2 + \|b\|^2}. \quad (10)$$

This of course only is possible if the angle between the vectors a and b is obtuse ($\pi/2 < \theta < \pi$).

Problem 2 Summary

Procedure

(a) Part (a)

- Use the formula for the sum of norms and show that the middle term in the formula is zero when the two vectors are perpendicular ($\theta = \frac{n\pi}{2}$)

(b) Part (b)

- Use the formula for the sum of norms and show that the middle term in the formula will evaluate to less than $2\|a\|\|b\|$ when the angle between the vectors is acute ($\theta < \frac{\pi}{2}$)
- Show that $\|a + b\|$ is then greater than the right hand side of the expression

(c) Part (c)

- Use the formula for the sum of norms and show that the middle term in the formula will evaluate to greater than $2\|a\|\|b\|$ when the angle between the vectors is obtuse ($\theta > \frac{\pi}{2}$)
- Show that $\|a + b\|$ is then less than the right hand side of the expression

Key Concepts

- The formula for the norm of one vector is

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

and the formula for finding the norm of two vectors is

$$\|x + y\| = \sqrt{\|x\|^2 + 2x^T y + \|y\|^2}.$$

- When two vectors are perpendicular to one another, then we can say that the relationship between the sum of norms is then

$$\|a + b\| = \sqrt{\|a\|^2 + \|b\|^2}.$$

- When the angle between two vectors is acute, we can then say that the relationship between the sum of norms is then

$$\|a + b\| > \sqrt{\|a\|^2 + \|b\|^2}.$$

- When the angle between two vectors is obtuse, we can then say that the relationship between the sum of norms is then

$$\|a + b\| < \sqrt{\|a\|^2 + \|b\|^2}.$$

Variations

- The only way that this problem could be changed is if the expression that we are trying to prove is different

- We would then have to use the same procedure in evaluating different values of θ and whatnot for the expression

Problem 3

Problem Statement

Explain the solution to 3.1 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Original Question

Distance between Boolean vectors. Suppose that x and y are Boolean n -vectors, which means that each of their entries is either 0 or 1. What is their distance $\|x - y\|$?

Solution

For this problem I will be using **Method 4**.

VMLS Solution:

For each i , $x_i - y_i$ is either $-1, 0$, or 1 ; $(x_i - y_i)^2$ is zero (if $x_i = y_i$) or 1 (if $x_i \neq y_i$). Summing over i , we find that $\|x - y\|^2$ is the total number of entries in which x and y differ. The distance is then the square root of the number of entries in which x and y differ.

Explanation:

Given two vectors x and y we know that there are a limited number of possibilities for the vectors in this scenario. Because of this we know the values of $x_i - y_i$ can only be

$$x_i - y_i = -1 \quad (\text{Lowest value. Occurs when } x_i = 0 \text{ and } y_i = 1) \quad (1)$$

$$= 0 \quad (\text{Median value. Occurs when } x_i = y_i) \quad (2)$$

$$= 1 \quad (\text{Highest value. Occurs when } x_i = 1 \text{ and } y_i = 0). \quad (3)$$

This means for each element the only possible values for $(x_i - y_i)^2$ are

$$(x_i - y_i)^2 = 0 \quad (\text{Lowest value. Occurs when } x_i = y_i) \quad (4)$$

$$= 1 \quad (\text{Highest value. Occurs when } x_i \neq y_i). \quad (5)$$

Regardless of the value of $x_i - y_i$ in equations (1)-(3), the distance between these boolean vectors is then

$$\|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (6)$$

Where i is the index of the given element of a vector and n is the length of the vectors.

Problem 3 Summary

Procedure

- Determine what the possible values of the entries can be between the two entries of the vectors
- Determine what the distance between the entries can be for the given values

Key Concepts

- Boolean vectors are vectors where the entries are either a 0 or 1
- The distance between two vectors is calculated with

$$\mathbf{dist}(a, b) = \|a - b\|$$

- The distance between entries of boolean vectors can either be 0 or 1

Variations

- We could be given a different vector to determine what the distance between these entries would be
 - In this case we would have to determine the possible values of each entry and then what the possible distance between those entries would be



Problem 4

Problem Statement

Explain the solution to 3.2 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Original Question

RMS value and average of block vectors. Let x be a block vector with two vector elements, $x = (a, b)$, where a and b are vectors of size n and m , respectively.

- (a) Express $\mathbf{rms}(x)$ in terms of $\mathbf{rms}(a)$, $\mathbf{rms}(b)$, m , and n .
- (b) Express \mathbf{avg} in terms of $\mathbf{avg}(a)$, $\mathbf{avg}(b)$, m , and n .

Solution - Part (a)

For this problem I will be using **Method 4**.

VMLS Solution:

- (a) The RMS value of the vector $x = (a_1, \dots, a_n, b_1, \dots, b_m)$ is

$$\begin{aligned}\mathbf{rms}(x) &= \left(\frac{a_1^2 + \dots + a_n^2 + b_1^2 + \dots + b_m^2}{n + m} \right)^{1/2} \\ &= \left(\frac{n(a_1^2 + \dots + a_n^2)/n + m(b_1^2 + \dots + b_m^2)/m}{n + m} \right)^{1/2} \\ &= \left(\frac{n\mathbf{rms}(a)^2 + m\mathbf{rms}(b)^2}{n + m} \right)^{1/2}.\end{aligned}$$

Explanation:

In its simplest terms, the root mean square of a vector μ can be defined as

$$\mathbf{rms}(\mu) = \sqrt{\frac{\mu_1^2 + \dots + \mu_\alpha^2}{\alpha}} = \frac{\|\mu\|}{\sqrt{\alpha}}. \quad (1)$$

The vector x that was given to us in the problem statement is a concatenated vector of two vectors a and b where the length of x is defined as α . Concisely, we can rewrite α as

$$\alpha = n + m \quad (2)$$

where n is the vector length of a and m is the vector length of b . This means, if we re-write equation (1) for that of vector x we will have

$$\mathbf{rms}(x) = \sqrt{\frac{x_1^2 + \dots + x_\alpha^2}{\alpha}} = \left(\frac{a_1^2 + \dots + a_n^2 + b_1^2 + \dots + b_m^2}{n + m} \right)^{1/2} \quad (3)$$

where we have used the definition of an inner product for a concatenated vector in the numerator of equation (3). To get equation (3) written in terms of the \mathbf{rms} of both sub vectors a and b we first need to *sneakily* multiply by 1 for both the inner product of a and b in the numerator of equation (3). Applying a completeness relation to the inner products of both vectors a and b in equation (3) we now have

$$\mathbf{rms}(x) = \left(\frac{n(a_1^2 + \dots + a_n^2)/n + m(b_1^2 + \dots + b_m^2)/m}{n + m} \right)^{1/2} \quad (4)$$

where all we have done to equation (3) is multiplying by a constant value of 1 to both the inner products of a and b with the lengths of their vectors respectively. Observing equation (4) more closely we can see that we can right the square of the \mathbf{rms} of a vector generically as

$$\mathbf{rms}^2(\mu) = \frac{\mu_1^2 + \cdots + \mu_\alpha^2}{\alpha}. \quad (5)$$

Applying equation (5) to both vectors a and b in equation (3) we have the $\mathbf{rms}(x)$ in terms of the $\mathbf{rms}(a)$ and $\mathbf{rms}(b)$ to be

$$\mathbf{rms}(x) = \left(\frac{n\mathbf{rms}(a)^2 + m\mathbf{rms}(b)^2}{n+m} \right)^{1/2}. \quad (6)$$

Equation (6) represents the $\mathbf{rms}(x)$ in terms of the \mathbf{rms} of both sub vectors a and b .

Solution - Part (b)

For this problem I will be using **Method 4**.

VMLS Solution:

(b) The average of the vector $x = (a_1, \dots, a_n, b_1, \dots, b_m)$ is

$$\mathbf{avg}(x) = \frac{a_1 + \cdots + a_n + b_1 + \cdots + b_m}{n+m} = \frac{n\mathbf{avg}(a) + m\mathbf{avg}(b)}{n+m}.$$

Explanation:

We define \mathbf{avg} of some arbitrary vector μ with length α to be

$$\mathbf{avg}(\mu) = \frac{\mu_1 + \cdots + \mu_\alpha}{\alpha}. \quad (7)$$

In the context of our concatenated vector x that was given to us in the problem statement, the lengths of sub vectors a and b are n and m respectively. This means the value of α is the same as that of equation (2) from part (a). With this in mind, equation (7) now becomes

$$\mathbf{avg}(x) = \frac{a_1 + \cdots + a_n + b_1 + \cdots + b_m}{n+m} \quad (8)$$

for vector x with sub vectors a and b . Similar to part (a), we now apply a completeness relation to both vectors a and b found in equation (8) with the lengths of each respective vector to get

$$\mathbf{avg}(x) = \frac{n(a_1 + \cdots + a_n)/n + m(b_1 + \cdots + b_m)/m}{n+m}. \quad (9)$$

Utilizing the generic definition for the average of a vector in equation (7) we can re-write equation (9) to now be in terms of the averages of both a and b . Namely,

$$\mathbf{avg}(x) = \frac{n\mathbf{avg}(a) + m\mathbf{avg}(b)}{n+m} \quad (10)$$

where equation (10) is the desired result that we have been working towards.

Problem 4 Summary

Procedure

(a) Part (a)

- Use the general form of **rms** for one vector and apply it to both vectors
- Put this form into the expression
- Manipulate this form so that it can match the form of that is requested of us by multiplying by 1
- Simplify this form so that it is written in terms of **rms**

(b) Part (b)

- Use the same process as in part (a) with **avg** instead of **rms**

Key Concepts

- We can write the **rms** and **avg** for concatenated vectors in the form of the respective values for each independent vector
- **rms** of a vector is calculated with

$$\mathbf{rms}(x) = \frac{\|x\|}{\sqrt{n}}$$

- **avg** of a vector is calculated with

$$\mathbf{avg}(x) = \frac{x_1 + \cdots + x_n}{n}$$

Variations

- We could be asked to write a different expression for a concatenated vector in the form of values for independent vectors
 - In this case we would have to use the definition of whatever we are trying to write and then apply it to each individual vector

Problem 5

Problem Statement

Explain the solution to 3.4 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Original Question

Norm identities. Verify that the following identities hold for any two vectors a and b of the same size.

(a) $(a + b)^T(a - b) = \|a\|^2 - \|b\|^2.$

(b) $\|a + b\|^2 + \|a - b\|^2 = 2(\|a\|^2 + \|b\|^2).$ This is called the *parallelogram law*.

Solution - Part (a)

For this problem I will be using **Method 4**.

VMLS Solution:

(a)

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

Explanation:

To verify this identity, we have to take advantage of some properties pertaining to vectors as well as inner products of vectors. We first take advantage of the *distributive* property of vector addition. Namely,

$$(a + b)^T c = a^T c + b^T c. \quad (1)$$

We also have to use the *commutative* property of inner products, namely

$$a^T b = b^T a. \quad (2)$$

Using these properties we can go on to show

$$(a + b)^T(a - b) = a^T(a - b) + b^T(a - b) \quad (\text{Distributive property of inner products}) \quad (3)$$

$$= a^T a - a^T b + b^T a - b^T b \quad (\text{Distributive property of inner products}) \quad (4)$$

$$= a^T a - a^T b + a^T b - b^T b \quad (\text{Commutative property of inner products}) \quad (5)$$

$$= a^T a - b^T b \quad (\text{Simplification by addition}) \quad (6)$$

$$= \|a\|^2 - \|b\|^2 \quad (\text{Definition of inner product}). \quad (7)$$

We have shown what was asked of us to show.

Solution - Part (b)

For this problem I will be using **Method 4**.

VMLS Solution:

(b)

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

Explanation:

To verify this identity, we need to use the definition of norm. The norm of a vector is defined as

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2} \quad (8)$$

for a vector x with length n . Extrapolating from equation (8) we can then say

$$\begin{aligned}
\|a+b\|^2 + \|a-b\|^2 &= (a+b)^T(a+b) + (a-b)^T(a-b) && \text{(Distributivity of norms)} \\
&= a^T a + b^T a + a^T b + b^T b + a^T a - b^T a - a^T b + b^T b && \text{(Distributivity of norms)} \\
&= a^T a + b^T b + a^T a + b^T b && \text{(Simplification by subtraction)} \\
&= 2a^T a + 2b^T b && \text{(Simplification by addition)} \\
&= 2\|a\|^2 + 2\|b\|^2 && \text{(Definition of inner products).}
\end{aligned}$$

(9)
(10)
(11)
(12)
(13)

We have shown what was asked of us to show.



Problem 5 Summary

Procedure

(a) Part (a)

- Apply the distributive property for transposes to the term $(a + b)^T(a - b)$
- Apply the commutative property for inner products
- Simplify the terms and use the definition norm to reach the conclusion

(b) Part (b)

- Apply the definition of norms to the term $\|a + b\|^2 + \|a - b\|^2$
- Use the distributive property of transposes and simplify terms in the following expressions
- Use the definition of inner products to simplify the final expression

Key Concepts

- Distributivity applies to transposes
- Commutativity applies to inner products
- We use the definition of norm to simplify terms in the expression
- We can say the following expressions are true

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

Variations

- We could be given different expressions to prove
 - In this case we would use the same properties and work through the algebra to prove the expression

Problem 6

Problem Statement

Explain the solution to 3.10 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Original Question

Nearest neighbor document. Consider the 5 Wikipedia pages in table 3.1 on page 51. What is the nearest neighbor of (the word count histogram vector of) ‘Veterans Day’ among the others? Does the answer make sense?

Solution

For this problem I will be using **Method 4**.

VMLS 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 3.1.) Yes, it makes sense — these two pages describe holidays, whereas the others do not.

Explanation:

We define the nearest neighbor for a vector x with a total of m neighbors to be z_j . In mathematical form this is

$$\|x - z_j\| \leq \|x - z_i\|, \quad i = 1, \dots, m. \quad (1)$$

Quantitatively, the distance between x (Veteran’s day) and it’s neighbors (z_i) is a scalar value that represents the distance between the two respective vectors. For a neighbor to be considered vector x ’s closest neighbor (z_j), the distance between the two vectors must be less than or equal to that of all the other individual neighbors. Upon observing table 3.1 on page 51, it is painstakingly obvious that the distance between Veteran’s day and the other neighbors that is the smallest is **Memorial Day**. Mathematically this is

$$|x - z_j| = 0.095 \quad (2)$$

where z_j is indeed Memorial Day. Verifying this mathematically we see

$$|x - z_{MD}| = 0.095 \quad (\text{Memorial Day}) \quad (3)$$

$$|x - z_{AA}| = 0.130 \quad (\text{Academy Awards}) \quad (4)$$

$$|x - z_{GGA}| = 0.153 \quad (\text{Golden Globe Awards}) \quad (5)$$

$$|x - z_{SB}| = 0.170 \quad (\text{Super Bowl}) \quad (6)$$

where when this is compared to the others we see

$$|x - z_{MD}| = 0.095 \leq |x - z_{AA}| = 0.130 \quad (7)$$

$$|x - z_{MD}| = 0.095 \leq |x - z_{GGA}| = 0.153 \quad (8)$$

$$|x - z_{MD}| = 0.095 \leq |x - z_{SB}| = 0.170. \quad (9)$$

and hence, we have shown that Veteran’s Day closest neighbor is Memorial Day.

Problem 6 Summary

Procedure

- Use the definition of nearest neighbor to determine what the nearest neighbor of a specific vector is

Key Concepts

- Nearest neighbor is defined to be the vector that is the closest vector to another given vector
- The nearest neighbor z_j to a vector x with a total of n vectors is

$$\|x - z_j\| \leq \|x - z_i\|, \quad i = 1, \dots, n$$

Variations

- We could be given a different data set to examine and determine what its nearest neighbor is
 - In this case we would use the definition of nearest neighbor to determine what the nearest neighbor of a vector is



Problem 7

Problem Statement

Explain the solution to 3.20 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Original Question

Regression model sensitivity. Consider the regression model $\hat{y} = x^T \beta + v$, where \hat{y} is the prediction, x is a feature vector, β is a coefficient vector, and v is the offset term. If x and \tilde{x} are feature vectors with corresponding predictions \hat{y} and \tilde{y} , show that $|\hat{y} - \tilde{y}| \leq \|\beta\| \|x - \tilde{x}\|$. This means that when $\|\beta\|$ is small, the prediction is not very sensitive to a change in the feature vector.

Solution

For this problem I will be using **Method 4**.

VMLS Solution:

Solution. The inequality follows the Cauchy-Schwarz inequality:

$$|\hat{y} - \tilde{y}| = |\beta^T x + v - (\beta^T \tilde{x} + v)| = |\beta^T (x - \tilde{x})| \leq \|\beta\| \|x - \tilde{x}\|.$$

Explanation:

From the problem statement, we can say in general for this regression model

$$\alpha = \gamma^T \delta + v \tag{1}$$

where α is the prediction, γ is the coefficient vector, δ is the feature vector, and v is the offset term. Using equation (1) we can extrapolate to say for this specific problem that

$$\hat{y} = \beta^T x + v \tag{2}$$

$$\tilde{y} = \beta^T \tilde{x} + v. \tag{3}$$

Taking the results from equations (2) and (3) we can then say

$$|\hat{y} - \tilde{y}| = |\beta^T x + v - (\beta^T \tilde{x} + v)| \tag{4} \quad \text{(Substitution)}$$

$$= |\beta^T x + v - \beta^T \tilde{x} - v| \tag{5} \quad \text{(Distributivity)}$$

$$= |\beta^T x - \beta^T \tilde{x} + v - v| \tag{6} \quad \text{(Commutativity)}$$

$$= |\beta^T x - \beta^T \tilde{x}| \tag{7} \quad \text{(Simplification by subtraction)}$$

$$= |\beta^T (x - \tilde{x})| \tag{8} \quad \text{(Factoring)}$$

$$\leq \|\beta\| \|x - \tilde{x}\| \tag{9} \quad \text{(Commutativity of norms).}$$

Where equation (9) represents the desired inequality that we have been asked to show.

Problem 7 Summary

Procedure

- Use the definition of the Cauchy Schwarz inequality and apply it to the two regression models

Key Concepts

- A regression model is of the form

$$\alpha = \gamma^T \delta + v$$

- The Cauchy Schwarz inequality is

$$\|a^T b\| \leq \|a\| \|b\|$$

- It can be shown that for two regression models the Cauchy Schwarz inequality is

$$|\hat{y} - \tilde{y}| \leq \|\beta\| \|x - \tilde{x}\|$$

Variations

- We could be asked to prove a different relationship
 - We would then use the same properties and definitions just with a different premise



Problem 8

Problem Statement

Explain the solution to 3.23 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Original Question

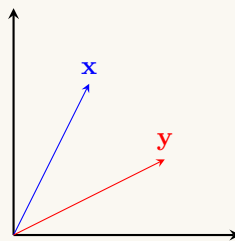
Angle between two nonnegative vectors. Let x and y be two nonzero n -vectors with nonnegative entries, i.e., each $x_i \geq 0$ and each $y_i \geq 0$. Show that the angle between x and y lies between 0 and 90° . Draw a picture for the case when $n = 2$, and give a short geometric explanation. When are x and y orthogonal?

Solution

For this problem I will be using **Method 4**.

VMLS Solution:

The angle lies between 0 and 90° because the inner product of two elementwise nonnegative vectors, which is a sum of products of nonnegative numbers, is nonnegative. When $n = 2$, the two vectors lie in the positive orthant, also known as quadrant I. The maximum angle between any such vectors is clearly 90° . This is illustrated below.



Two nonnegative vectors are orthogonal if and only if they have a complementary zero-nonzero pattern, i.e., $x_k = 0$ whenever $y_k > 0$ and $y_k = 0$ whenever $x_k > 0$.

Explanation:

Mathematically, the inner product between two vectors x and y can be defined as

$$x^T y = \|x\| \|y\| \cos(\theta) \quad (1)$$

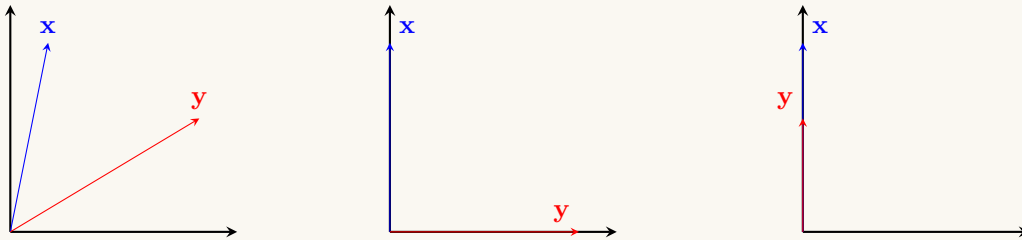
where θ is defined as the angle between two vectors. Solving for the angle θ in equation (1) yields

$$\theta = \arccos\left(\frac{x^T y}{\|x\| \|y\|}\right). \quad (2)$$

In equation (2), it is a condition of trigonometric functions that the numerator found in arccos must be less than that of the denominator. Because the quantities $x^T y$ and $\|x\|$ as well as $\|y\|$ are positive values for vectors with nonnegative entries, this constitutes that the value that is computed for arccos must be

$$0 \leq \arccos\left(\frac{x^T y}{\|x\| \|y\|}\right) \leq \pi/2 \quad , \quad 0 \leq \theta \leq \pi/2. \quad (3)$$

The vectors will be orthogonal when the inner product between the two vectors is 0, or in terms of angles, when the angle between the vectors is $\pi/2$. Because of the minimum and maximum angle θ is either 0 or $\pi/2$, we can confidently say that these vectors will lie in the **first quadrant**. These vectors could resemble something like what is illustrated below.



The image that is furthest to the left is an example of two vectors where the angle between them is greater than 0 and less than $\pi/2$. The middle image is an example of two vectors where two vectors are orthogonal (perpendicular) where the angle between them is $\pi/2$. The image that is furthest to the right is an example where the vectors are parallel to one another and the angle between them is 0.



Problem 8 Summary

Procedure

- Use the definition of inner product and solve it for the angle between the vectors
- Show what the vectors will look like for given values of θ

Key Concepts

- When vectors are orthogonal, the inner product of the two vectors is zero and the angle between them is $\pi/2$
- When vectors are parallel, the inner product is non zero and the angle between them is less than $\pi/2$
- When vectors are not orthogonal and parallel, then the angle between them is not 0 or $\pi/2$

Variations

- We could be given a different set of orientations for vectors
 - We would have to figure out the angle between the vectors for this new orientation



Problem 9

Problem Statement

Explain the solution to 3.24 here in your own words. (Since you are given a solution, you will be graded on your ability to explain). Create your own example for part (a).

Original Question

Distance versus angle nearest neighbor. Suppose z_1, \dots, z_m is a collection of n -vectors, and x is another n -vector. The vector z_j is the (distance) nearest neighbor of x (among the given vectors) if

$$\|x - z_j\| \leq \|x - z_i\|, \quad i = 1, \dots, m,$$

i.e., x has smallest distance to z_j . We say that z_j is the *angle nearest neighbor* of x if

$$\angle(x, z_j) \leq \angle(x, z_i), \quad i = 1, \dots, m,$$

i.e., x has smallest angle to z_j .

- Give a simple specific numerical example where the (distance) nearest neighbor is not the same as the angle nearest neighbor.
- Now suppose that the vectors z_1, \dots, z_m are normalized, which means that $\|z_i\| = 1$, $i = 1, \dots, m$. Show that in this case the distance nearest neighbor and the angle nearest neighbor are always the same. *Hint.* You can use the fact that arccos is a decreasing function, i.e., for any u and v with $-1 \leq u < v \leq 1$, we have $\arccos(u) > \arccos(v)$.

Solution - Part (a)

For this problem I will be using **Method 4**.

For this example, we will choose $n = 1$. We then choose $z_1 = 2$, $z_2 = -5$, and $x = 4$. We then have the results

$$\|x - z_1\| = 2, \quad \|x - z_2\| = 9, \quad \angle(x, z_1) = 0, \quad \angle(x, z_2) = \pi. \quad (1)$$

The nearest neighbor to x is then z_1 and the angle nearest neighbor is z_1 .

Solution - Part (b)

For this problem I will be using **Method 4**.

To show what is asked of us, we do the following

$$\|x - z_j\| \leq \|x - z_i\| \Leftrightarrow \|x - z_j\|^2 \leq \|x - z_i\|^2 \quad (\text{Premise}) \quad (2)$$

$$\Leftrightarrow \|x\|^2 - 2x^T z_j + \|z_j\|^2 \leq \|x\|^2 - 2x^T z_i + \|z_i\|^2 \quad (\text{Substitution}) \quad (3)$$

$$\Leftrightarrow -2x^T z_j \leq -2x^T z_i \quad (\text{Simplification by subtraction}) \quad (4)$$

$$\Leftrightarrow x^T z_j \geq x^T z_i \quad (\text{Simplification by division}) \quad (5)$$

$$\Leftrightarrow \frac{x^T z_j}{\|x\|\|z_j\|} \geq \frac{x^T z_i}{\|x\|\|z_i\|} \quad (\cos(\theta) \text{ of inner products}) \quad (6)$$

$$\Leftrightarrow \arccos\left(\frac{x^T z_j}{\|x\|\|z_j\|}\right) \leq \arccos\left(\frac{x^T z_i}{\|x\|\|z_i\|}\right) \quad (\theta \text{ of vectors}) \quad (7)$$

$$\Leftrightarrow \langle x, z_j \rangle \leq \langle x, z_i \rangle \quad (\text{Definition of inner product}). \quad (8)$$

Hence, the nearest neighbor and nearest neighbor angle will always be the same.

Problem 9 Summary

Procedure

(a) Part (a)

- Use the definition of nearest neighbor and evaluate what the nearest neighbor is for each entry

(b) Part (b)

- Use definition of norm with the nearest neighbor definition and the appropriate properties to reach the conclusion

Key Concepts

- We can find relationships between inner products and nearest neighbor

Variations

- We could be given a different set of properties to prove
 - We would then use the same properties and procedures to reach the conclusion

