

Homework 3: Vectors and the Dot Product Solutions

EECS 245, Spring 2026 at the University of Michigan

due Sunday, May 17th, 2026 at 11:59PM Ann Arbor Time

Write your solutions to the following problems either by writing them on a piece of paper or on a tablet and scanning your answers as a PDF. Note that you are not allowed to use LaTeX, Google Docs, or any other digital document creation software to type your answers. Homeworks are due to Gradescope by 11:59PM on the due date. See the [syllabus](#) for details on the slip day policy.

Homework will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should always explain and justify your conclusions, using sound reasoning. Your goal should be to convince the reader of your assertions. If a question does not require explanation, it will be explicitly stated.

Before proceeding, make sure you're familiar with the [collaboration policy](#).

Total Points: $10 + 14 + 9 + 7 + 15 + 12 + 10 + 6 = 83$

Problem 1: Homework 2 Solutions Review (10 pts)

Review the solutions to Homework 2. Pick **two problem parts** (for example, Problem 2a and Problem 4b) from Homework 2 in which your solutions have the most room for improvement, i.e., where they have unsound reasoning, could be significantly more efficient or clearer, etc. **Include a screenshot of your solution to each problem part**, and in a few sentences, explain what was deficient and how it could be fixed.

Alternatively, if you think one of your solutions is significantly better than the posted one, copy it here and explain why you think it is better. If you didn't do Homework 2, choose two problem parts from it that look challenging to you, and in a few sentences, explain the key ideas behind their solutions in your own words.

Solution:

Problem 2: Parallelogram Law (14 pts)

a) (4 pts) Let $\vec{u} = \begin{bmatrix} 3 \\ -6 \\ 0 \\ 2 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 2 \\ 1 \\ 4 \\ -2 \end{bmatrix}$. Compute the following quantities:

(i) $\|\vec{u}\|$

(ii) $\|\vec{v}\|$

(iii) $\|\vec{u} + \vec{v}\|$

(iv) $\|\vec{u} - \vec{v}\|$

Solution:

(i) $\|\vec{u}\| = \sqrt{3^2 + (-6)^2 + 0^2 + 2^2} = \sqrt{9 + 36 + 0 + 4} = \sqrt{49} = 7$

(ii) $\|\vec{v}\| = \sqrt{2^2 + 1^2 + 4^2 + (-2)^2} = \sqrt{4 + 1 + 16 + 4} = \sqrt{25} = 5$

(iii)

$$\begin{aligned}\|\vec{u} + \vec{v}\| &= \sqrt{(3+2)^2 + (-6+1)^2 + (0+4)^2 + (2-2)^2} \\ &= \sqrt{25 + 25 + 16 + 0} \\ &= \sqrt{66}\end{aligned}$$

(iv)

$$\begin{aligned}\|\vec{u} - \vec{v}\| &= \sqrt{(3-2)^2 + (-6-1)^2 + (0-4)^2 + (2-(-2))^2} \\ &= \sqrt{1 + 49 + 16 + 16} \\ &= \sqrt{82}\end{aligned}$$

b) (3 pts) Using the same vectors as in part a), compute the angle between \vec{u} and \vec{v} . Leave your answer in terms of \cos^{-1} .

Solution:

Since $\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos \theta$, we have:

$$\vec{u} \cdot \vec{v} = 3 \cdot 2 + (-6) \cdot 1 + 0 \cdot 4 + 2 \cdot (-2) = 6 - 6 + 0 - 4 = -4,$$

$$\|\vec{u}\| = \sqrt{3^2 + (-6)^2 + 0^2 + 2^2} = \sqrt{49} = 7,$$

$$\|\vec{v}\| = \sqrt{2^2 + 1^2 + 4^2 + (-2)^2} = \sqrt{25} = 5,$$

$$\cos \theta = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|} = \frac{-4}{7 \cdot 5} = -\frac{4}{35}.$$

$$\theta = \cos^{-1}\left(-\frac{4}{35}\right)$$

c) (4 pts) Now, suppose $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ are any two vectors in \mathbb{R}^n . Prove that:

$$\|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2 = 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2$$

The statement above is called the **parallelogram law** of vectors.

*Hint: The point of part a) was to give you a feel for which quantities are involved in this statement. Your proof should not use these values in particular. Instead, **start with the left-hand side** of the equation and use the properties of the dot product introduced in [Chapter 3.3](#).*

Solution:

Since $\|\vec{u}\|^2 = \vec{u} \cdot \vec{u}$, we can expand both $\|\vec{u} + \vec{v}\|^2$ and $\|\vec{u} - \vec{v}\|^2$ using the distributive property of the dot product, as mentioned in the hint. Below, we color $\vec{u} \cdot \vec{u}$ in orange, $\vec{v} \cdot \vec{v}$ in blue, and $\vec{u} \cdot \vec{v}$ in magenta.

$$\begin{aligned} \|\vec{u} + \vec{v}\|^2 &= (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) \\ &= \vec{u} \cdot \vec{u} + \vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{u} + \vec{v} \cdot \vec{v} \\ &= \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} + 2\vec{u} \cdot \vec{v}. \end{aligned}$$

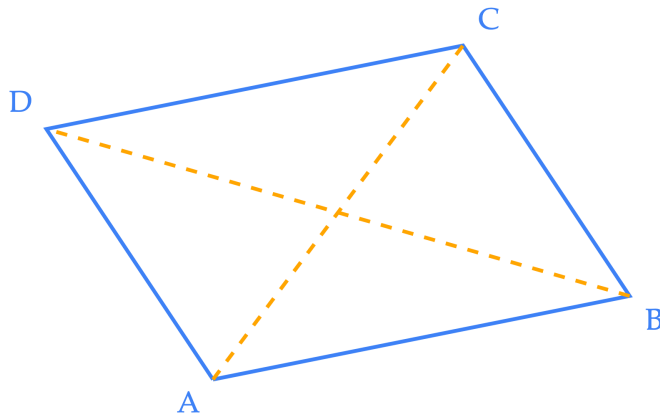
$$\begin{aligned} \|\vec{u} - \vec{v}\|^2 &= (\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) \\ &= \vec{u} \cdot \vec{u} - \vec{u} \cdot \vec{v} - \vec{v} \cdot \vec{u} + \vec{v} \cdot \vec{v} \\ &= \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\vec{u} \cdot \vec{v}. \end{aligned}$$

Note that the $+2\vec{u} \cdot \vec{v}$ from the first expansion and the $-2\vec{u} \cdot \vec{v}$ from the second cancel when we add them. Thus,

$$\begin{aligned} \|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2 &= (\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} + 2\vec{u} \cdot \vec{v}) \\ &\quad + (\vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\vec{u} \cdot \vec{v}) \\ &= 2\vec{u} \cdot \vec{u} + 2\vec{v} \cdot \vec{v} \\ &= 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2. \end{aligned}$$

d) (3 pts) Why is the equality from part c) called the parallelogram law? Let's explore.

Suppose points $A, B, C,$ and D in \mathbb{R}^n form a parallelogram: a polygon with four sides where opposite sides are parallel and equal in length.



Using the results of the previous part of this problem, prove that the sum of the squares of the **side lengths of the parallelogram** is equal to the sum of the squares of the **diagonals**. In other words, prove that:

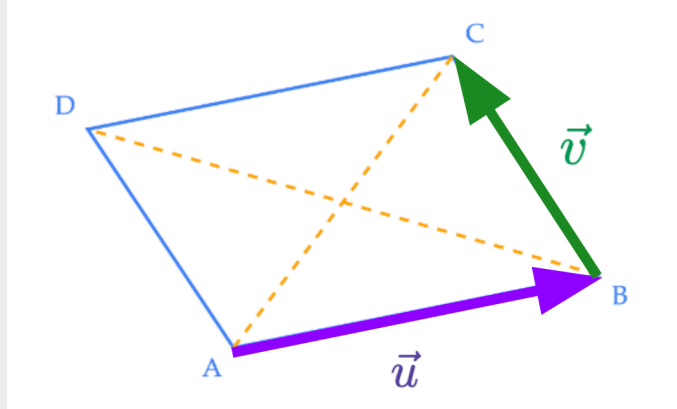
$$(AB)^2 + (BC)^2 + (CD)^2 + (DA)^2 = (AC)^2 + (BD)^2$$

where AB represents the length of the segment from point A to point B , etc.

Hint: Define two vectors, \vec{u} and \vec{v} , and explain why the result from the previous part of this problem implies the desired equality. This is mostly an English problem.

Solution:

Let \vec{u} be the vector along \overrightarrow{AB} (purple arrow in the figure) and \vec{v} be the vector along \overrightarrow{BC} (green arrow).



Opposite sides of a parallelogram are equal and parallel, so the remaining sides point as

$$\overrightarrow{CD} = -\vec{u}, \quad \overrightarrow{DA} = -\vec{v}.$$

Next, read each diagonal as a head-to-tail sum of side vectors.

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC} = \vec{u} + \vec{v}, \quad \overrightarrow{BD} = \overrightarrow{BC} + \overrightarrow{CD} = \vec{v} - \vec{u}.$$

Now translate side lengths into norms of the corresponding vectors.

$$(AB) = \|\vec{u}\|, \quad (BC) = \|\vec{v}\|, \quad (CD) = \|-\vec{u}\| = \|\vec{u}\|, \quad (DA) = \|-\vec{v}\| = \|\vec{v}\|.$$

Squaring and summing the four sides gives

$$\begin{aligned} (AB)^2 + (BC)^2 + (CD)^2 + (DA)^2 &= \|\vec{u}\|^2 + \|\vec{v}\|^2 + \|\vec{u}\|^2 + \|\vec{v}\|^2 \\ &= 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2. \end{aligned}$$

Do the same for the diagonals, using the vector expressions above.

$$(AC) = \|\vec{u} + \vec{v}\|, \quad (BD) = \|\vec{v} - \vec{u}\|.$$

Squaring and adding the diagonals yields

$$(AC)^2 + (BD)^2 = \|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2.$$

By the parallelogram law proven in part (c),

$$\|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2 = 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2.$$

Combining with the side-length sum computed above,

$$\boxed{(AB)^2 + (BC)^2 + (CD)^2 + (DA)^2 = (AC)^2 + (BD)^2}$$

Problem 3: Linear Combinations (9 pts)

As we saw in [Chapter 3.1](#), a **linear combination** of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_d \in \mathbb{R}^n$ is a vector of the form

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \dots + a_d\vec{v}_d$$

where a_1, a_2, \dots, a_d are scalars.

Much of our study of linear algebra involves understanding **the set of possible linear combinations** of a given set of vectors. As the notes detail, our multiple linear regression problem boils down to finding the best possible linear combination of the features, so it's important that we understand how linear combinations work.

$$\text{Let } \vec{v}_1 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}, \text{ and } \vec{x} = \begin{bmatrix} -6 \\ 1 \\ 4 \end{bmatrix}.$$

You can find an interactive, three-dimensional visualization of these four vectors at this link:

<https://eecs245.org/resources/homeworks/hw03/hw03-problem-2.html>

We recommend you have this visual open while you work through this problem.

a) (4 pts) Find constants a , b , and c such that

$$a\vec{v}_1 + b\vec{v}_2 + c\vec{v}_3 = \vec{x}$$

In other words, write \vec{x} as a linear combination of \vec{v}_1 , \vec{v}_2 , and \vec{v}_3 .

Hint: Start by writing out the equation as a system of equations. Then, use your favorite method for solving systems of equations to find a , b , and c . We reviewed how to solve systems of equations in Lab 3.

Solution: $a = -3, b = 0, c = 2$

We can express the equation above as a system of three equations in the three unknowns a, b, c .

$$a\vec{v}_1 + b\vec{v}_2 + c\vec{v}_3 = \vec{x}$$
$$a \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} + b \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} + c \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} -6 \\ 1 \\ 4 \end{bmatrix}$$

Equating components gives the system

$$2a - b = -6 \quad (1)$$

$$3a + 2b + 5c = 1 \quad (2)$$

$$-b + 2c = 4 \quad (3)$$

In (1), isolate b :

$$b = 6 + 2a \quad (4)$$

Substitute this expression for b into (3):

$$-(6 + 2a) + 2c = 4$$

$$2c = 10 + 2a$$

$$c = 5 + a \quad (5)$$

Now substitute (4) and (5) into (2):

$$3a + 2(6 + 2a) + 5(5 + a) = 1$$

$$3a + 12 + 4a + 25 + 5a = 1$$

$$12a + 37 = 1$$

$$12a = -36$$

$$a = -3 \quad (6)$$

Back-substitute to get b and c :

$$b = 6 + 2(-3) = 0$$

$$c = 5 + (-3) = 2$$

So: to verify that these values are correct, compute the linear combination:

$$a\vec{v}_1 + b\vec{v}_2 + c\vec{v}_3 = (-3) \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} -6 \\ -9 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 10 \\ 4 \end{bmatrix}$$

$$= \begin{bmatrix} -6 \\ 1 \\ 4 \end{bmatrix} = \vec{x}$$

$$\boxed{a = -3, b = 0, c = 2}$$

b) (2 pts) Try and find constants d and e such that

$$d\vec{v}_1 + e\vec{v}_3 = \vec{x}$$

If you are able to find constants d and e , **explain why**, even though there are two unknowns but three equations for them. If you are unable to find constants d and e , **explain why** no solution exists.

Solution: $d = -3, e = 2$

This is a system of three equations and two unknowns, since there are only two vectors in the linear combination. We'll start by solving it the same way we did in part (a).

$$d\vec{v}_1 + e\vec{v}_3 = \vec{x}$$
$$d \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} + e \begin{bmatrix} 0 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} -6 \\ 1 \\ 4 \end{bmatrix}$$

Equating components gives

$$2d = -6 \tag{1}$$

$$3d + 5e = 1 \tag{2}$$

$$2e = 4 \tag{3}$$

We can solve directly from (1) and (3):

$$d = -3 \tag{4}$$

$$e = 2 \tag{5}$$

Now check equation (2) with these values:

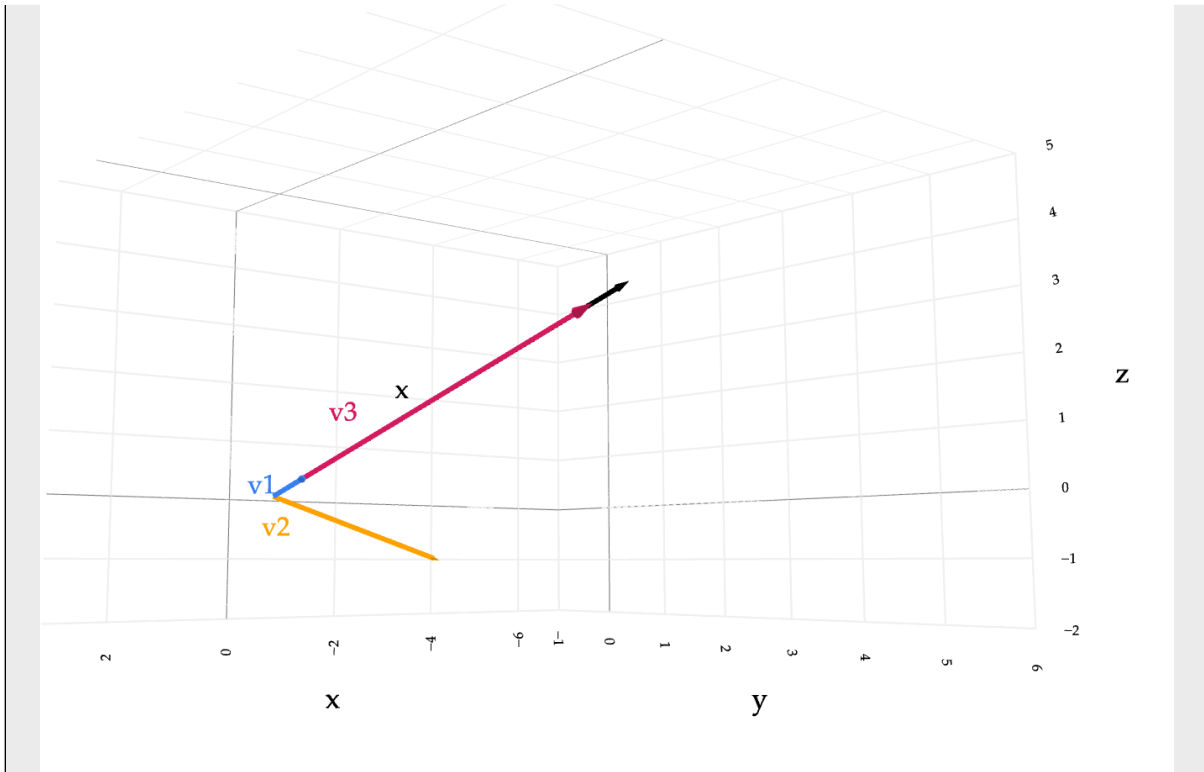
$$3(-3) + 5(2) = -9 + 10 = 1$$

so (2) holds as well. Notice we did not need (2) to *find* d and e ; it served as a consistency check, which it passes, so a solution exists.

You could also have seen this without re-solving: in part (a), the coefficient on \vec{v}_2 was $b = 0$. Since \vec{v}_2 does not appear in the combination here, the same coefficients on \vec{v}_1 and \vec{v}_3 work, namely $d = a = -3$ and $e = c = 2$.

$$\boxed{d = -3, e = 2}$$

Geometrically, \vec{x} lies in the plane spanned by \vec{v}_1 and \vec{v}_3 (the set $\text{span}\{\vec{v}_1, \vec{v}_3\}$), so we did not need \vec{v}_2 to reach \vec{x} . (See next page)



Visual note: In the view above, the arrows for \vec{v}_1 , \vec{v}_3 , and \vec{x} look almost colinear only because of the camera angle. What matters is that they are *coplanar*: $\vec{x} \in \text{span}\{\vec{v}_1, \vec{v}_3\}$. The arrow for \vec{v}_2 is not in that plane, which is why we did not need it in the combination. To see this clearly, open the interactive viewer and rotate the scene:

<https://eecs245.org/resources/homeworks/hw03/hw03-problem-2.html>

c) (3 pts) Try and find constants p and q such that

$$p\vec{v}_1 + q\vec{v}_2 = \vec{x}$$

If you are able to find constants p and q , **explain why**, even though there are two unknowns but three equations for them. If you are unable to find constants p and q , **explain why** no solution exists.

Solution: No solution.

This is a system of three equations in two unknowns, because the linear combination uses only two vectors. We'll attempt to solve it the same way as in part (a).

$$p\vec{v}_1 + q\vec{v}_2 = \vec{x}$$
$$p \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} + q \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -6 \\ 1 \\ 4 \end{bmatrix}$$

Equating components gives the system

$$2p - q = -6 \tag{1}$$

$$3p + 2q = 1 \tag{2}$$

$$-q = 4 \tag{3}$$

From (3) we get $q = -4$. Use (1) with $q = -4$:

$$2p - (-4) = -6$$

$$2p + 4 = -6$$

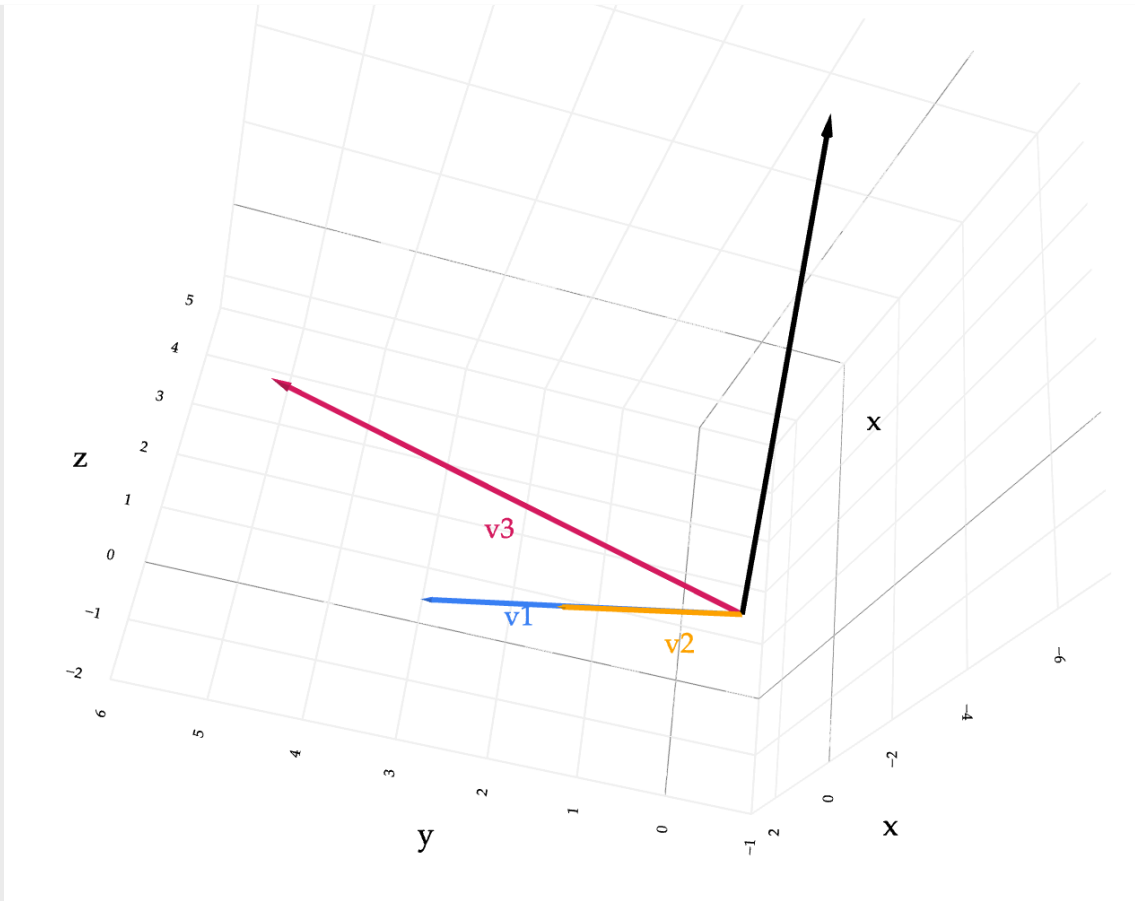
$$2p = -10$$

$$p = -5$$

Now check (2) with $p = -5$, $q = -4$:

$$3(-5) + 2(-4) = -15 - 8 = -23 \neq 1$$

Since (2) is violated, the system is inconsistent and there are no constants p, q satisfying all three equations.



Why this also makes sense visually. From the 3D visualization, \vec{x} lies in the plane generated by linear combinations of \vec{v}_1 and \vec{v}_3 ($\text{span}\{\vec{v}_1, \vec{v}_3\}$). Trying to use only \vec{v}_1 and \vec{v}_2 keeps you in the different plane $\text{span}\{\vec{v}_1, \vec{v}_2\}$, which does not contain \vec{x} . Hence no solution with $p\vec{v}_1 + q\vec{v}_2 = \vec{x}$.

No solution

Problem 4: Correlation (7 pts)

In [Chapter 2.4](#), you were told that the correlation coefficient, r , ranges between -1 and 1 , where -1 implies a perfect negative linear association and 1 implies a perfect positive linear association. However, you were never given a proof of the fact that $-1 \leq r \leq 1$.

Here, you will prove this fact, given your newfound understanding of vectors, the dot product, and angles.

a) (2 pts) Let \vec{x} and \vec{y} be two vectors in \mathbb{R}^n . We define the "mean-centered" version of \vec{x} to be:

$$\vec{x}_c = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is the mean of the components of \vec{x} . The mean-centered version of \vec{y} , named \vec{y}_c , is defined similarly.

Express $\vec{x}_c \cdot \vec{y}_c$ using summation notation.

Solution:

$$\begin{aligned} \vec{x}_c \cdot \vec{y}_c &= \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix} \cdot \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} \\ &= (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y}) \\ &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \end{aligned}$$

b) (3 pts) Prove that:

$$r = \frac{\vec{x}_c \cdot \vec{y}_c}{\|\vec{x}_c\| \|\vec{y}_c\|}$$

Do so by starting with the right-hand side of the equation, expanding it, and simplifying it until you reach the definition of r .

Solution:

From part (a), the dot product of the mean-centered vectors is

$$\vec{x}_c \cdot \vec{y}_c = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

and the norms are

$$\|\vec{x}_c\| = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \|\vec{y}_c\| = \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Start from the right-hand side and substitute these definitions:

$$\frac{\vec{x}_c \cdot \vec{y}_c}{\|\vec{x}_c\| \|\vec{y}_c\|} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Now rewrite each square root as \sqrt{n} times the square root of the average (this is just $\sqrt{\sum} = \sqrt{n}\sqrt{\text{average}}$):

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(\sqrt{n} \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}) (\sqrt{n} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2})}$$

Let

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

so each denominator factor is $\sqrt{n} \sigma_x$ and $\sqrt{n} \sigma_y$. This gives

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$$

Finally, pull out the $1/n$ and group terms to match the correlation definition (average of standardized products):

$$\frac{\vec{x}_c \cdot \vec{y}_c}{\|\vec{x}_c\| \|\vec{y}_c\|} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) = r$$

- c) (2 pts) In 1-2 English sentences, explain why the result from part b) implies that $-1 \leq r \leq 1$.

Solution:

The result from part b) implies that $-1 \leq r \leq 1$ because the right-hand side of the equation is the cosine of the angle between the mean-centered vectors, and the cosine of an angle is bounded between -1 and 1 .

Problem 5: Projections (15 pts)

In [Chapter 3.4](#), we study the concept of **projecting** one vector onto one or more other vectors. In this problem, you'll see how this concept can be thought of in terms of our friend from the first two weeks of the course: calculus.

Let \vec{x} and \vec{y} be two vectors in \mathbb{R}^n . Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$, defined as:

$$f(k) = \|\vec{y} - k\vec{x}\|^2$$

By $\mathbb{R} \rightarrow \mathbb{R}$, we mean that f takes in a single real number (i.e. a scalar, **not** a vector) and outputs a single real number. This means that we can find $\frac{df}{dk}$, the derivative of f with respect to k .

Note that $k\vec{x}$ is a vector that points in the same direction (or the opposite direction) as \vec{x} .

a) (4 pts) Rewrite $f(k)$ using the properties of the dot product from [Chapter 3.3](#). Then, show that:

$$\frac{df}{dk} = -2\vec{x} \cdot \vec{y} + 2k\vec{x} \cdot \vec{x}$$

Solution:

Start by rewriting the squared norm as a dot product and expanding. We color the dot-product pieces to track them: $\vec{y} \cdot \vec{y}$ (orange), $\vec{x} \cdot \vec{x}$ (blue), and $\vec{x} \cdot \vec{y}$ (magenta).

$$\begin{aligned} f(k) &= \|\vec{y} - k\vec{x}\|^2 \\ &= (\vec{y} - k\vec{x}) \cdot (\vec{y} - k\vec{x}) \\ &= \vec{y} \cdot \vec{y} - k(\vec{y} \cdot \vec{x}) - k(\vec{x} \cdot \vec{y}) + k^2(\vec{x} \cdot \vec{x}) \\ &= \|\vec{y}\|^2 - 2k(\vec{x} \cdot \vec{y}) + k^2(\vec{x} \cdot \vec{x}) \quad (\text{since } \vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}). \end{aligned}$$

Now differentiate term-by-term with respect to k (the dot products are constants with respect to k):

$$\frac{df}{dk} = 0 - 2(\vec{x} \cdot \vec{y}) + 2k(\vec{x} \cdot \vec{x}).$$

$$\boxed{\frac{df}{dk} = -2\vec{x} \cdot \vec{y} + 2k(\vec{x} \cdot \vec{x})}$$

b) (2 pts) Find k^* , the value of k that minimizes $f(k)$. A second derivative test is not necessary.

Solution:

From part **a**), we know that $\frac{df}{dk} = -2 \vec{x} \cdot \vec{y} + 2k (\vec{x} \cdot \vec{x})$. Let's set this derivative to 0 and solve for k .

$$\begin{aligned} -2 \vec{x} \cdot \vec{y} + 2k (\vec{x} \cdot \vec{x}) &= 0 \\ 2k (\vec{x} \cdot \vec{x}) &= 2 \vec{x} \cdot \vec{y} \\ k^* &= \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} \end{aligned}$$

This gives the minimizer when $\vec{x} \cdot \vec{x} > 0$. If $\vec{x} \cdot \vec{x} = 0$, the only possibility is $\vec{x} = \vec{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

In that case,

$$f(k) = \|\vec{y} - k\vec{0}\|^2 = \|\vec{y}\|^2$$

which does not depend on k . Thus every value of k yields the same objective value, and any k minimizes f .

$$k^* = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} \text{ when } \vec{x} \neq \vec{0}$$

c) (3 pts) Show that the vectors $k^* \vec{x}$ and $\vec{y} - k^* \vec{x}$ are orthogonal.

Solution:

From part **b**), $k^* = \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}}$ when $\vec{x} \neq \vec{0}$. To check whether two vectors are orthogonal, we compute their dot product; they are orthogonal if and only if the dot product equals 0.

$$\begin{aligned} (k^* \vec{x}) \cdot (\vec{y} - k^* \vec{x}) &= (k^* \vec{x}) \cdot \vec{y} - (k^* \vec{x}) \cdot (k^* \vec{x}) \\ &= k^* (\vec{x} \cdot \vec{y}) - (k^*)^2 (\vec{x} \cdot \vec{x}) \\ &= \frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} (\vec{x} \cdot \vec{y}) - \left(\frac{\vec{x} \cdot \vec{y}}{\vec{x} \cdot \vec{x}} \right)^2 (\vec{x} \cdot \vec{x}) \\ &= \frac{(\vec{x} \cdot \vec{y})^2}{\vec{x} \cdot \vec{x}} - \frac{(\vec{x} \cdot \vec{y})^2}{\vec{x} \cdot \vec{x}} \\ &= 0 \end{aligned}$$

Since the dot product is 0, the vectors $k^* \vec{x}$ and $\vec{y} - k^* \vec{x}$ are orthogonal.

d) (4 pts) Now, let's study a seemingly unrelated problem. Suppose we're given a dataset $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ and we'd like to find the optimal model parameter, w , for a simple

linear model **with no intercept term**,

$$h(x_i) = wx_i$$

Find the value of w that minimizes the average loss (i.e. empirical risk) when using squared loss. A second derivative test is not necessary. (To be clear, the solution to this problem does not involve linear algebra.)

Solution: First, let's start by writing down an expression for mean squared error, $R_{\text{sq}}(w)$.

$$R_{\text{sq}}(w) = \frac{1}{n} \sum_{i=1}^n (y_i - wx_i)^2$$

Now, let's find $\frac{dR_{\text{sq}}}{dw}$. Differentiate term by term; for each i ,

$$\frac{d}{dw} (y_i - wx_i)^2 = 2(y_i - wx_i)(-x_i)$$

Then,

$$\begin{aligned} \frac{dR}{dw} &= \frac{1}{n} \sum_{i=1}^n 2(y_i - wx_i)(-x_i) \\ &= -\frac{2}{n} \sum_{i=1}^n x_i y_i + \frac{2}{n} w \sum_{i=1}^n x_i^2 \end{aligned}$$

Next, we'll set $\frac{dR_{\text{sq}}}{dw} = 0$ and solve for the corresponding value of w , called w^* .

$$\begin{aligned} -\sum_{i=1}^n x_i y_i + w^* \sum_{i=1}^n x_i^2 &= 0 \\ w^* \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \\ w^* &= \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \end{aligned}$$

$$w^* = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

e) (2 pts) Suppose $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is a vector of all the x_i values in the dataset and $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ is a vector

of all the y_i values in the dataset.

Then, notice that w^* (the optimal slope) from part **d**) is the same formula as k^* (the optimal stretching factor) from part **b**)! This is not a coincidence: the problem in part **d**) is equivalent to the problem stated at the start of this question, it's just stated differently! This may be a bit confusing, since:

- The problem at the start involves two vectors, \vec{x} and \vec{y} , which live in \mathbb{R}^n , and n may be very large (could be 100-dimensional space!).
- The problem in part **d**) involves a dataset of n points, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, but the points themselves along with the line $h(x_i) = wx_i$ are drawn in \mathbb{R}^2 .

In the vector view, we're finding the best scalar multiple of a vector in \mathbb{R}^n to make it as close as possible to another vector in \mathbb{R}^n . In the regression view, we're fitting a line (through the origin) to points in \mathbb{R}^2 .

Let's make the connection between the two viewpoints more explicit. In part **d**), once we pick a value of w , the predictions for each x_i are of the form wx_i . A vector of predictions, \vec{p} , might look like:

$$\vec{p} = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_n) \end{bmatrix} = \begin{bmatrix} wx_1 \\ wx_2 \\ \vdots \\ wx_n \end{bmatrix} = w \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = w\vec{x}$$

Given this, in 1-2 English sentences, explain why finding the w that minimizes

$$R_{\text{sq}}(w) = \frac{1}{n} \sum_{i=1}^n (y_i - wx_i)^2$$

is equivalent to finding the scalar k that minimizes

$$f(k) = \|\vec{y} - k\vec{x}\|^2.$$

Solution: Minimizing

$$R_{\text{sq}}(w) = \frac{1}{n} \sum_{i=1}^n (y_i - wx_i)^2$$

is equivalent to minimizing

$$\|\vec{y} - w\vec{x}\|^2$$

because the vector of residuals has components $y_i - wx_i$, and squaring and summing those components is exactly the squared Euclidean norm of $\vec{y} - w\vec{x}$. Thus, choosing the best slope w in the regression problem is the same as choosing the best scalar k that makes $k\vec{x}$ as close as possible to \vec{y} .

Problem 6: Norms (12 pts)

In [the last section of Chapter 3.2](#), we introduced the concept of vector norms *other than* the “default” Euclidean norm. Each of those norms describes a different way of measuring the length of a vector — just like how different loss functions described different ways of measuring the error of a prediction.

- a) (2 pts) In this part only, let $\vec{v} = \begin{bmatrix} 3 \\ -6 \\ 0 \\ 2 \end{bmatrix}$. Compute $\|\vec{v}\|_2$, $\|\vec{v}\|_1$, and $\|\vec{v}\|_\infty$.

Solution:

$$\begin{aligned}\|\vec{v}\|_2 &= \sqrt{3^2 + (-6)^2 + 0^2 + 2^2} = \sqrt{9 + 36 + 0 + 4} = \sqrt{49} = 7, \\ \|\vec{v}\|_1 &= |3| + |-6| + |0| + |2| = 3 + 6 + 0 + 2 = 11, \\ \|\vec{v}\|_\infty &= \max\{|3|, |-6|, |0|, |2|\} = \max\{3, 6, 0, 2\} = 6.\end{aligned}$$

- b) (3 pts) In Problem 1, we introduced the parallelogram law, which states that

$$\|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2 = 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2$$

In general, the parallelogram law only holds for the L_2 norm, not necessarily other norms.

Find a counterexample involving two vectors \vec{u} and \vec{v} such that the parallelogram law **does not hold** for the L_1 norm.

Solution:

We want to check whether the parallelogram law holds for the L_1 norm

$$\|\vec{u} + \vec{v}\|_1^2 + \|\vec{u} - \vec{v}\|_1^2 \stackrel{?}{=} 2\|\vec{u}\|_1^2 + 2\|\vec{v}\|_1^2$$

Choose

$$\vec{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Left-hand side

$$\vec{u} + \vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \|\vec{u} + \vec{v}\|_1 = |1| + |1| = 2, \quad \|\vec{u} + \vec{v}\|_1^2 = 2^2 = 4$$

$$\vec{u} - \vec{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \|\vec{u} - \vec{v}\|_1 = |1| + |-1| = 2, \quad \|\vec{u} - \vec{v}\|_1^2 = 2^2 = 4$$

$$\|\vec{u} + \vec{v}\|_1^2 + \|\vec{u} - \vec{v}\|_1^2 = 4 + 4 = 8$$

Right-hand side

$$\|\vec{u}\|_1 = |1| + |0| = 1, \quad \|\vec{v}\|_1 = |0| + |1| = 1$$

$$2\|\vec{u}\|_1^2 + 2\|\vec{v}\|_1^2 = 2 \cdot 1^2 + 2 \cdot 1^2 = 2 + 2 = 4$$

Since $8 \neq 4$, the parallelogram law does not hold for the L_1 norm with this choice of \vec{u}, \vec{v}

c) (3 pts) Prove that

$$\|\vec{v}\|_2 \leq \sqrt{n} \|\vec{v}\|_\infty$$

Hint: Start by writing out the definition of the L_2 norm, and then square it to remove the square root. You will have a sum of n terms. Explain why each of those n terms is less than or equal to $\|\vec{v}\|_\infty^2$. This is most of the way to the proof, but there's still some work you'll need to do after you get to that point.

Solution: Start from the definition of the L_2 norm

$$\|\vec{v}\|_2^2 = \sum_{i=1}^n v_i^2$$

For each index i , by the definition of the L_∞ norm we have $|v_i| \leq \|\vec{v}\|_\infty$

$$v_i^2 \leq \|\vec{v}\|_\infty^2 \quad \text{for each } i$$

Sum these inequalities over $i = 1, \dots, n$

$$\sum_{i=1}^n v_i^2 \leq \sum_{i=1}^n \|\vec{v}\|_\infty^2 = n \|\vec{v}\|_\infty^2$$

Thus

$$\|\vec{v}\|_2^2 \leq n \|\vec{v}\|_\infty^2$$

Both sides are nonnegative, so taking square roots preserves the inequality

$$\|\vec{v}\|_2 \leq \sqrt{n} \|\vec{v}\|_\infty$$

d) (4 pts) Prove that

$$\|\vec{v}\|_2 \leq \|\vec{v}\|_1$$

Hint: Start with the fact that

$$\|\vec{v}\|_1^2 = (|v_1| + |v_2| + \dots + |v_n|)^2$$

Solution:

As the hint implies, let's start by expanding the definition of $\|\vec{v}\|_1^2$. We write the L_1 norm as one square to avoid carrying a square root

$$\|\vec{v}\|_1^2 = (|v_1| + |v_2| + \cdots + |v_n|)^2$$

For intuition, recall the perspective from part (c): by definition, the L_∞ norm is the largest absolute value among the components, so $\|\vec{v}\|_\infty \geq |v_i|$ for every i , and this remains true after squaring because both sides are nonnegative. Here, we will instead use the hint and expand the L_1 square into "square" terms and "cross" terms

$$\|\vec{v}\|_1^2 = \underbrace{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}_{\text{squares}} + \underbrace{2(|v_1||v_2| + |v_1||v_3| + \cdots + |v_{n-1}||v_n|)}_{\text{cross terms}}$$

Each cross term $|v_i||v_j|$ is nonnegative, so the cross-term sum shown above is ≥ 0

$$\|\vec{v}\|_1^2 \geq |v_1|^2 + |v_2|^2 + \cdots + |v_n|^2$$

The right-hand side is exactly the square of the L_2 norm

$$|v_1|^2 + \cdots + |v_n|^2 = \|\vec{v}\|_2^2$$

Therefore

$$\|\vec{v}\|_1^2 \geq \|\vec{v}\|_2^2$$

Both sides are nonnegative, so taking square roots preserves the inequality

$$\|\vec{v}\|_2 \leq \|\vec{v}\|_1$$

Problem 7: Neighbors (10 pts)

This problem involves writing code and submitting it to the Gradescope autograder.

There are two ways to access the supplemental Jupyter Notebook:

- **Option 1 (preferred):** Set up a Jupyter Notebook environment locally, use `git` to clone our course repository, and open `homeworks/hw03/hw03.ipynb`. For instructions on how to do this, see the [Environment Setup](#) page of the course website.
- **Option 2:** Click [here](#) to open `hw03.ipynb` on DataHub. Before doing so, read the instructions on the [Environment Setup](#) page on how to use the DataHub.

To receive credit for the programming portion of the homework, you'll need to submit your completed notebook to the autograder on Gradescope. Your submission time for Homework 3 is the **latter** of your PDF and code submission times.

Problem 8: Feedback (6 pts)

We'd like to get your feedback on how the course has been going so far, now that we're a few weeks in.

You can find the survey [at this link](#). It is **not anonymous**, but it links to an anonymous feedback form if you'd like to provide some feedback anonymously.

Thank you for your feedback — it's helping shape our brand-new course.