

# Midterm 1 **Solutions**

EECS 245, Winter 2026 at the University of Michigan

**Name:** \_\_\_\_\_

**uniqname:** \_\_\_\_\_

**UMID:** \_\_\_\_\_

**Room:**  1670 BBB (big room)  1690 BBB  Other

## Instructions

- This exam consists of 7 problems, worth a total of 100 points, spread across 12 pages (6 sheets of paper).
- You have 120 minutes to complete this exam, unless you have extended-time accommodations through SSD.
- Write your uniqname in the top right corner of each page.
- For free response problems, **you must show all of your work**, and **circle** your final answer. We will not grade work that appears elsewhere, and you may lose points if your work is not shown.
- For multiple choice problems, completely fill in bubbles and square boxes; if we cannot tell which option(s) you selected, you may lose points.
  - A bubble means that you should only select one choice.
  - A square box means you should select all that apply.
- You may refer to **one double-sided 8.5x11" handwritten notes sheet**. Other than that, you may not refer to any other resources or technology during the exam (no phones, watches, or calculators).

You are to abide by the University of Michigan/Engineering Honor Code. To receive a grade, please sign below to signify that you have kept the Honor Code pledge.

*I have neither given nor received aid on this exam, nor have I concealed any violations of the Honor Code.*

This page has been intentionally left blank.

**Tip:** Skim through the entire exam before starting to work on it.

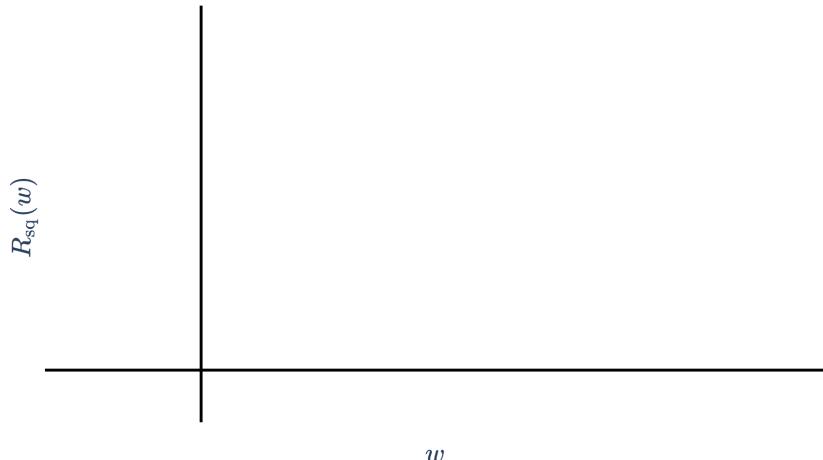
**Problem 1 (16 pts)**

Consider a dataset of  $n$  values,  $y_1, y_2, \dots, y_n$ , with:

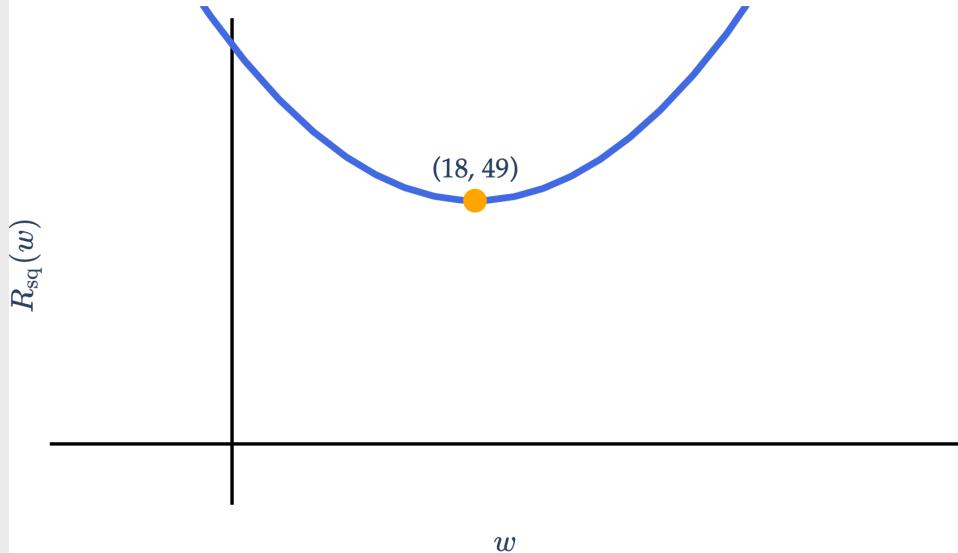
- a mean of  $\bar{y} = 18$
- a median of 15
- a standard deviation of  $\sigma_y = 7$

a) (4 pts) In the space provided, sketch the graph of  $R_{\text{sq}}(w)$ , the mean squared error of a constant prediction  $w$  on the dataset. For full credit:

- The shape of the graph must be correct.
- You must clearly label the coordinates of the **minimum point** on the graph.



**Solution:**



Recall that

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

is a parabola, minimized at  $w = \bar{y}$ . When  $w = \bar{y}$ ,

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

is the variance of the dataset. Here, the mean is 18 and the variance is 49, so the minimum point is at (18, 49).

**b) (6 pts)** Which of the following quantities is **guaranteed** to be equal to 0? Select all that apply.

- $\frac{1}{n} \sum_{i=1}^n (y_i - 15)$
- $\frac{1}{n} \sum_{i=1}^n (y_i - 18)$
- $\frac{1}{n} \sum_{i=1}^n (y_i - 15)^2$
- $\frac{1}{n} \sum_{i=1}^n (y_i - 18)^2$
- $\frac{1}{n} \sum_{i=1}^n (y_i - 15)^2 - 7^2$
- $\frac{1}{n} \sum_{i=1}^n (y_i - 18)^2 - 7^2$

**Solution:**

There are two key ideas at play here:

- The mean is the unique point in the dataset such that the sum of deviations from the mean is 0. In other words,

$$\sum_{i=1}^n (y_i - \bar{y}) = \sum_{i=1}^n y_i - n\bar{y} = n\bar{y} - n\bar{y} = 0$$

- The variance of a dataset is the average of the squared deviations from the mean. In other words,

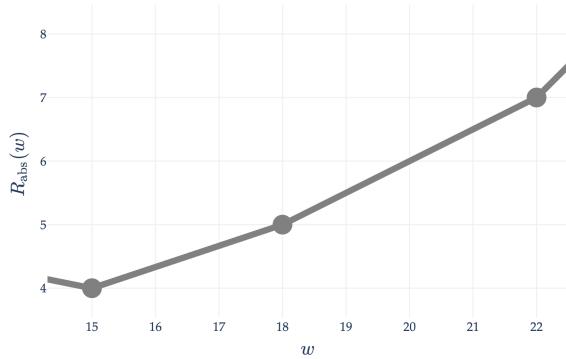
$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Equivalently, this is the value of  $R_{\text{sq}}(w)$  when  $w = \bar{y}$ .

With this in mind, let's look at the options:

1. **(False)**  $\frac{1}{n} \sum_{i=1}^n (y_i - 15)$ : This is the average of the deviations from the median, which is not 0. This is only true for the mean.
2. **(True)**  $\frac{1}{n} \sum_{i=1}^n (y_i - 18)$ : This is the average of the deviations from the mean, which is 0. This is only true for the mean.
3. **(False)**  $\frac{1}{n} \sum_{i=1}^n (y_i - 15)^2$ : This is the function  $R_{\text{sq}}(w)$  when  $w = 15$ . As we see in the solution to part a), this is not 0.
4. **(False)**  $\frac{1}{n} \sum_{i=1}^n (y_i - 18)^2$ : This is the function  $R_{\text{sq}}(w)$  when  $w = 18$ , i.e. it is the variance of the dataset. As we see in the solution to part a), this is also not zero — here, it is  $\sigma_y^2 = 7^2 = 49$ . One point of confusion may be that  $w = \bar{y}$  is the point at which  $R_{\text{sq}}(w)$  is minimized and  $R_{\text{sq}}(w)$  has a **derivative** of 0, but  $R_{\text{sq}}(\bar{y}) \neq 0$  in general.
5. **(False)**  $\frac{1}{n} \sum_{i=1}^n (y_i - 15)^2 - 7^2$ : This would be true if the 15 were replaced with the mean, 18, but it is not.
6. **(True)**  $\frac{1}{n} \sum_{i=1}^n (y_i - 18)^2 - 7^2$ : This is the variance of the dataset minus the variance of the dataset, which indeed is 0.

c) (6 pts) Recall that  $R_{\text{abs}}(w)$  is the mean absolute error of a constant prediction  $w$  on the dataset. A snippet of the graph of  $R_{\text{abs}}(w)$  is shown below.



For clarity, the circles at  $(15, 4)$ ,  $(18, 5)$ , and  $(22, 7)$  indicate the points at which the slope of  $R_{\text{abs}}(w)$  changes.

Given that there are  $n = 72$  values in the dataset, how many values in the dataset are equal to **18**? Show your work and  circle your final answer, which should be an integer with no variables.

**Solution:** The number of values in the dataset that are equal to 18 is 6. Recall, the slope of  $R_{\text{abs}}(w)$  at any  $w$  that is not a data point is:

$$\frac{d}{dw} R_{\text{abs}}(w) = \frac{\# \text{ left of } w - \# \text{ right of } w}{n}$$

There are two line segments of interest here: the one between  $w = 15$  and  $w = 18$ , and the one between  $w = 18$  and  $w = 22$ . We have two ways of computing the slope of each one: by using slope  $= \frac{\text{change in } y}{\text{change in } x}$  and by using the formula above. We'll use both formulas on each line segment.

- **Between  $w = 15$  and  $w = 18$ :**

- Method 1: Using slope  $= \frac{\text{change in } y}{\text{change in } x}$ , the graph rises from  $(15, 4)$  to  $(18, 5)$ , which gives a slope of

$$s_1 = \frac{5 - 4}{18 - 15} = \frac{1}{3}$$

- Method 2: Using the formula for the slope of  $R_{\text{abs}}(w)$ , let  $l$  be the number of values in the dataset less than or equal to 15. Then, the slope in this interval is

$$s_1 = \frac{l - (72 - l)}{72} = \frac{2l - 72}{72}$$

At this point, we have enough information to solve for  $l$ :

$$\frac{2l - 72}{72} = \frac{1}{3} \implies l = 48$$

- **Between  $w = 18$  and  $w = 22$ :**

- Method 1:

$$s_2 = \frac{7 - 5}{22 - 18} = \frac{2}{4} = \frac{1}{2}$$

- Method 2: Let  $k$  be the number of values in the dataset **equal to** 18. Ultimately, this is what we're trying to find. Then, the number of values in the dataset less than or equal to 18 is  $l + k$ . In this interval, the slope is

$$s_2 = \frac{(l + k) - (72 - (l + k))}{72} = \frac{2(l + k) - 72}{72}$$

So, we need to solve for  $k$  in

$$\frac{2(l + k) - 72}{72}$$

But, we know that  $l = 48$ , so

$$\frac{2(48 + k) - 72}{72} = \frac{1}{2} \implies 96 + 2k - 72 = 36 \implies 2k = 12 \implies k = 6$$

Therefore, there are 6 values in the dataset that are equal to 18.

## Problem 2 (14 pts)

Suppose we'd like to fit a simple linear regression model to a dataset of  $n$  points,  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , by minimizing mean squared error.

Suppose  $w_0^*$  and  $w_1^*$  are the optimal intercept and slope parameters, respectively, and let

$$M = \frac{1}{n} \sum_{i=1}^n (y_i - (w_0^* + w_1^* x_i))^2$$

Finally, let  $\sigma_x$  and  $\sigma_y$  be the standard deviations of the  $x$ -values and  $y$ -values in the dataset, respectively. Assume that  $\sigma_x > 0$  and  $\sigma_y > 0$ .

a) (5 pts) Which of the following is the relationship between  $M$  and  $\sigma_y^2$ ? Select an answer and provide a brief explanation in the box provided.

$M \leq \sigma_y^2$      $M = \sigma_y^2$      $M \geq \sigma_y^2$     Impossible to tell

**Solution:**  $M$  is the mean squared error of the best simple linear regression model for the dataset; it minimizes the mean squared error among all models of the form

$$h(x_i) = w_0 + w_1 x_i$$

The constant model,  $h(x_i) = w$ , can be thought of as a more restrictive version of the simple linear regression model, in that it has an intercept  $w$  and slope of 0. So, the best simple linear regression model is at least as good as the best constant model, when both are measured by mean squared error. If the  $x$  and  $y$  values in the dataset have no linear association, meaning the correlation coefficient  $r$  is 0, then the best simple linear regression model is the same as the best constant model; otherwise, the best simple linear regression model is better, since it has all of the flexibility of the constant model, and more. The first section of [Chapter 2.5](#) discusses this idea further.

b) (5 pts) Suppose that  $M = 0$ . What is the value of  $r$ , the correlation coefficient between the  $x$ -values and  $y$ -values in the dataset?  your final answer and provide a brief explanation. If there are multiple possible values, state them all.

**Solution:**  $r = 1$  or  $r = -1$ .

The only case in which  $M = 0$  is when the best simple linear regression model makes 0 errors, i.e. it passes through every point in the dataset. This happens when the  $x$  and  $y$  values in the dataset have a perfect linear association, meaning  $r = 1$  (positive linear association) or  $r = -1$  (negative linear association).

c) (2 pts) True or False: It is possible for there to be multiple pairs of (intercept, slope) with a mean squared error of  $M$ .

True    False

**Solution:** The values of  $w_0^*$  and  $w_1^*$  are unique. We've seen several formulas for them in the notes; they are the unique minimizers of

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n} \sum_{i=1}^n (y_i - (w_0 + w_1 x_i))^2$$

d) (2 pts) True or False: It is possible for there to be multiple pairs of (intercept, slope) with a mean squared error of  $M + 1$ .

True  False

**Solution:** The values of  $w_0$  and  $w_1$  that minimize  $R_{\text{sq}}(w_0, w_1)$  are unique, but we're not discussing the minimizers here, so that fact is irrelevant.

Instead, it's asking whether it's possible for there to be multiple pairs of  $(w_0, w_1)$  with a mean squared error of something bigger than  $M$ . The  $+1$  is not important; we could have stated  $+17$  or  $+3\pi^2$  and the question would be the same.

Recall from [Chapter 2.3](#) that the graph of  $R_{\text{sq}}(w_0, w_1)$  looks like a bowl in  $\mathbb{R}^3$ . While there's only one point at which the bowl is minimized, for any height ( $z$ -value) greater than  $M$ , there are infinitely many pairs of  $(w_0, w_1)$  that give that height. To see this, imagine slicing the bowl with the plane  $z = M + 1$ . This slice is an ellipse (stretched circle), upon which infinitely many combinations of  $(w_0, w_1)$  lie.

So, yes, it is possible for there to be multiple pairs of  $(w_0, w_1)$  with a mean squared error of  $M + 1$  — in fact, that's guaranteed.

### Problem 3 (12 pts)

Consider the following two planes,  $P_1$  and  $P_2$ , in  $\mathbb{R}^3$ .

- $P_1$  is the plane spanned by the vectors  $\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$  and  $\begin{bmatrix} 6 \\ -4 \\ -3 \end{bmatrix}$ .
- $P_2$  is the plane defined by the equation  $5x + 3y - z = 0$ .

a) (6 pts) Find the equation of  $P_1$  in standard form, i.e.  $ax + by + cz + d = 0$ . Show your work and circle your final answer.

**Solution:**  $2x - 3y + 8z = 0$ .

As discussed in [Chapter 4.4](#), the solution is to take the cross product of the two vectors used to span the plane; this will give us a vector  $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$  that is orthogonal to both vectors, and therefore both will satisfy  $ax + by + cz + d = 0$ . We know  $d = 0$  since the span of a set of vectors must contain the origin.

$$\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \times \begin{bmatrix} 6 \\ -4 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \cdot (-3) - 0 \cdot (-4) \\ 0 \cdot 6 - 3 \cdot (-3) \\ 3 \cdot (-4) - 2 \cdot 6 \end{bmatrix} = \begin{bmatrix} -6 \\ 9 \\ -24 \end{bmatrix}$$

So, the equation of  $P_1$  is  $-6x + 9y - 24z = 0$ , or simplified,  $2x - 3y + 8z = 0$ . To verify, we should plug in both vectors to make sure they satisfy the equation:

$$2(3) - 3(2) + 8(0) = 6 - 6 + 0 = 0, \quad 2(6) - 3(-4) + 8(-3) = 12 + 12 - 24 = 0$$

b) (6 pts) Planes  $P_1$  and  $P_2$  intersect at a line. Find the equation of this line in parametric form. Show your work and circle your final answer. *Hint: This can be done without knowing the answer to the previous part.*

**Solution:**

$$L = t \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \quad t \in \mathbb{R}$$

(where the direction vector could be scaled by any non-zero scalar)

There are a few possible techniques here.

1. We can find the intersection of the two planes by solving the system of equations:

$$\begin{aligned} 5x + 3y - z &= 0 \\ 2x - 3y + 8z &= 0 \end{aligned}$$

Adding both equations gives

$$7x + 7z = 0 \implies z = -x$$

We know that the system will have infinitely many solutions, so we can let our “parameter” be  $x$ . So far, we know two of the three components of the line:  $x$  is the free variable, and  $z = -x$ . Finally, let’s solve for  $y$  in terms of  $x$ .

$$5x + 3y + x = 0 \implies 6x + 3y = 0 \implies y = -2x$$

Therefore, the parametric equation of the line is

$$L = \begin{bmatrix} x \\ -2x \\ -x \end{bmatrix} = x \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \quad x \in \mathbb{R}$$

2. Another solution is to recognize that any point on the first plane can be written as a linear combination of the two vectors that span the plane, i.e.

$$s \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 6 \\ -4 \\ -3 \end{bmatrix} = \begin{bmatrix} 3s + 6t \\ 2s - 4t \\ -3t \end{bmatrix}$$

Any vector on the first plane can be written in the form above. For a vector to be in both planes (i.e. in the intersection), it must be able to be written in the form above **and** satisfy the equation of the second plane,  $5x + 3y - z = 0$ .

$$5(3s + 6t) + 3(2s - 4t) - (-3t) = 0$$

$$15s + 30t + 6s - 12t + 3t = 0$$

$$21s + 21t = 0$$

$$t = -s$$

So, as long as we pick  $s$  and  $t$  such that  $t = -s$ , the resulting vector,  $\begin{bmatrix} 3s + 6t \\ 2s - 4t \\ -3t \end{bmatrix}$ , will be in both planes. There are infinitely many pairs of such  $s$  and  $t$  – 1 and  $-1$ , 2 and  $-2$ , etc. – and these fill out the line of intersection. To find one of them, let  $s = 1$  and  $t = -1$ :

$$\begin{bmatrix} 3(1) + 6(-1) \\ 2(1) - 4(-1) \\ -3(-1) \end{bmatrix} = \begin{bmatrix} 3 - 6 \\ 2 + 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 6 \\ 3 \end{bmatrix}$$

Therefore, the parametric equation of the line is

$$L = t \begin{bmatrix} -3 \\ 6 \\ 3 \end{bmatrix}, \quad t \in \mathbb{R}$$

which is equivalent to

$$L = t \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \quad t \in \mathbb{R}$$

This is the same line we found earlier, just with a scaled direction vector, which doesn't change the line.

3. A final solution is to (1) find a vector that is perpendicular to each plane (i.e. a normal vector), and (2) take the cross product of those two vectors. This will give us a vector that is in both planes, and therefore spans the intersecting line, which we know must also pass through the origin.

$$\begin{bmatrix} 5 \\ 3 \\ -1 \end{bmatrix} \times \begin{bmatrix} 2 \\ -3 \\ 8 \end{bmatrix} = \begin{bmatrix} 3 \cdot 8 - (-1) \cdot (-3) \\ (-1) \cdot 2 - 5 \cdot 8 \\ 5 \cdot (-3) - 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} 21 \\ -42 \\ -21 \end{bmatrix} = 21 \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$$

So, once again, we find that  $\begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$  is a direction vector for the line of intersection.

**Problem 4 (12 pts)**

Suppose  $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$  and  $\vec{u} + \vec{v} + \vec{w} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$ . Assume that none of  $\vec{u}$ ,  $\vec{v}$ , or  $\vec{w}$  are the zero vector,  $\vec{0}$ .

For each statement below, identify whether it is **impossible**, **possible**, or **guaranteed**, and provide a brief explanation in the box provided.

a) (4 pts)  $\vec{u}$  and  $\vec{v}$  are orthogonal.

Impossible  Possible  Guaranteed

**Solution:** This is possible.

There is nothing stopping  $\vec{u}$  and  $\vec{v}$  from being orthogonal. For example, let  $\vec{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ .

Then,  $\vec{u} \cdot \vec{v} = 0 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0$ , so  $\vec{u}$  and  $\vec{v}$  are orthogonal, and we can still find a  $\vec{w}$

such that  $\vec{u} + \vec{v} + \vec{w} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$ . For example, let  $\vec{w} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} - \vec{u} - \vec{v} = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$ .

However, it's not guaranteed:  $\vec{v} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$  and  $\vec{w} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$  satisfy  $\vec{u} + \vec{v} + \vec{w} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$ , but  $\vec{u}$  and

$\vec{v}$  are not orthogonal.

So, it is possible for  $\vec{u}$  and  $\vec{v}$  to be orthogonal.

b) (4 pts) The set  $\{\vec{u}, \vec{v}, \vec{w}\}$  is linearly dependent.

Impossible  Possible  Guaranteed

**Solution:** This is guaranteed.

$$\vec{u} + \vec{v} + \vec{w} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}$$

Since  $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ , we can subtract  $4\vec{u}$  from both sides to get

$$\vec{u} + \vec{v} + \vec{w} - 4\vec{u} = \vec{w} - 3\vec{u} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} - 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Or, equivalently,

$$-3\vec{u} + \vec{v} + \vec{w} = \vec{0}$$

This is a non-trivial linear combination of  $\vec{u}$ ,  $\vec{v}$ , and  $\vec{w}$  that equals the zero vector, so the set  $\{\vec{u}, \vec{v}, \vec{w}\}$  is linearly dependent. Equivalently, we could say  $\vec{w} = 3\vec{u} - \vec{v}$ , which means  $\vec{w}$  is a linear combination of  $\vec{u}$  and  $\vec{v}$ , which also means the set is linearly dependent.

c) (4 pts)  $\vec{u}$ ,  $\vec{v}$ , and  $\vec{w}$  all have the same norm (length).

Impossible  Possible  Guaranteed

**Solution:** This is impossible.

Recall that the triangle inequality states that for any two vectors  $\vec{v}$  and  $\vec{w}$ ,

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$$

We know that  $\|\vec{u}\| = 1$ , so in order for the statement to be possible, we'd need both  $\|\vec{v}\| = 1$  and  $\|\vec{w}\| = 1$ . But,  $\vec{v} + \vec{w} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$ , so  $\|\vec{v} + \vec{w}\| = \sqrt{3^2 + 0^2 + 0^2} = \sqrt{9} = 3$ . In the triangle inequality, this would mean

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\| \implies 3 \leq 2$$

This is a contradiction, so it is impossible for both  $\vec{v}$  and  $\vec{w}$  to have a norm of 1, and therefore impossible for all three vectors to have the same norm.

**Problem 5 (12 pts)**

Suppose  $\vec{u}, \vec{v} \in \mathbb{R}^n$ . Let  $\vec{p}$  be the projection of  $\vec{u}$  onto  $\vec{v}$ . Furthermore, we know that:

$$\underbrace{\|\vec{v}\| = 2}_{\text{length of } \vec{v} \text{ (not } \vec{u})} \quad \|\vec{p}\| = 3$$

a) (6 pts) Find  $|\vec{u} \cdot \vec{v}|$ . Show your work and **circle** your final answer, which should be a number with no variables.

**Solution:**  $|\vec{u} \cdot \vec{v}| = 6$ .

Let's start with the formula for  $\vec{p}$ .

$$\vec{p} = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v} = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v}$$

We know that  $\|\vec{p}\| = 3$ , so let's try and find the magnitude of  $\vec{p}$  in the formula above, which will allow us to learn more about  $\vec{u} \cdot \vec{v}$ .

The key to remember that  $\|kx\| = |k|\|x\|$  for any scalar  $k$  and vector  $x$ . The absolute value is necessary because the scalar  $k$  could be negative, but the length of a vector is always non-negative.

$$\|\vec{p}\| = \left| \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \right| \|\vec{v}\| = \left| \frac{\vec{u} \cdot \vec{v}}{2^2} \right| 2 = \left| \frac{\vec{u} \cdot \vec{v}}{4} \right| 2 = \frac{|\vec{u} \cdot \vec{v}|}{2}$$

So, we know that  $\frac{|\vec{u} \cdot \vec{v}|}{2} = 3$ , which means that  $|\vec{u} \cdot \vec{v}| = 6$ .

b) (6 pts) For each pair of vectors, determine whether they are orthogonal, linearly dependent, or neither. Make sure to select **one bubble per row**.

	pair of vectors	orthogonal	linearly dependent	neither
(i)	$\vec{u}$ and $\vec{u} - \vec{p}$	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
(ii)	$\vec{u}$ and $\vec{v} - \vec{p}$	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
(iii)	$\vec{v}$ and $\vec{u} - \vec{p}$	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
(iv)	$\vec{v}$ and $\vec{v} - \vec{p}$	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
(v)	$\vec{p}$ and $\vec{u} - \vec{p}$	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
(vi)	$\vec{p}$ and $\vec{v} - \vec{p}$	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

**Solution:** The key fact about orthogonality when it comes to projections is that the error vector — here,  $\vec{e} = \vec{u} - \vec{p}$  — is orthogonal to the vector we're projecting onto,  $\vec{v}$ .

This means that  $\vec{v}$  and  $\vec{u} - \vec{p}$  are orthogonal (iii). But,  $\vec{p}$  is a scalar multiple of  $\vec{v}$ , so  $\vec{p}$  and  $\vec{u} - \vec{p}$  are also orthogonal (v).

Remember that  $\vec{p}$  is a scalar multiple of  $\vec{v}$ , so  $\vec{v} - \vec{p}$  is a scalar multiple of  $\vec{v}$  too. So,  $\vec{v}$  and  $\vec{v} - \vec{p}$  are linearly dependent (iv), as are  $\vec{p}$  and  $\vec{v} - \vec{p}$  (vi).

Now, we need to address (i) and (ii), which ask about  $\vec{u}$ 's relation to  $\vec{u} - \vec{p}$  and  $\vec{v} - \vec{p}$ , respectively.  $\vec{u} - \vec{p}$  is the error vector of the projection, which in general is orthogonal to  $\vec{v}$  and neither orthogonal nor linearly dependent with  $\vec{u}$ .

The only possible “edge case” here is when  $\vec{u}$  and  $\vec{v}$  are orthogonal, in which case  $\vec{p} = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v} = \frac{0}{\vec{v} \cdot \vec{v}} \vec{v} = \vec{0}$ , which would mean that  $\vec{u}$  and  $\vec{v} - \vec{p}$  are orthogonal and  $\vec{u}$  and  $\vec{u} - \vec{p}$  are the same vector and thus linearly dependent. However, we know that  $\vec{p} \neq \vec{0}$  since  $\|\vec{p}\| = 3 > 0$ . So, this edge case doesn't apply to this problem, and therefore  $\vec{u}$  and  $\vec{u} - \vec{p}$  are neither orthogonal nor linearly dependent (i), and same with  $\vec{u}$  and  $\vec{v} - \vec{p}$  (ii).

**Problem 6 (14 pts)**

Suppose  $\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{x}_5, \vec{x}_6$  are 6 vectors in  $\mathbb{R}^9$  such that

$$S = \text{span}(\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{x}_5, \vec{x}_6\})$$

is a **4-dimensional** subspace of  $\mathbb{R}^9$ .

a) (2 pts) True or False: The set  $\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{x}_5, \vec{x}_6\}$  is linearly independent.  
 True    False

**Solution:** This is false.

If these vectors were linearly independent, they would span a 6-dimensional subspace of  $\mathbb{R}^9$ ; since they only span a 4-dimensional subspace, they must be linearly dependent, and two of them are “redundant”.

b) (4 pts) Consider the statement:

“There exists a vector  $\vec{b} \in \mathbb{R}^9$  such that the number of ways to write  $\vec{b}$  as a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$  is \_\_\_\_.”

In each part below, a possible way to fill in the blank is given. Determine whether the statement that results from filling in the blank is **True** or **False**.

- (i) zero  
 True    False
- (ii) exactly one  
 True    False
- (iii) exactly two  
 True    False
- (iv) infinite  
 True    False

**Solution:**

- **(True) zero ways:**  $S$ , the set of all linear combinations of  $\vec{x}_1, \dots, \vec{x}_6$ , is a 4-dimensional subspace of  $\mathbb{R}^9$ . Since  $S$  isn't all of  $\mathbb{R}^9$ , there are plenty of vectors  $\vec{b} \in \mathbb{R}^9$  that are not in  $S$ , and therefore can't be written as a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$ . So, it's true that there are some  $\vec{b}$ 's such that there are zero ways to write  $\vec{b}$  as a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$ .
- **(False) exactly one way:** Linear combinations are only unique if the spanning vectors are linearly independent. Since  $\vec{x}_1, \dots, \vec{x}_6$  are linearly dependent, there is a non-trivial linear combination of them that equals the zero vector, which could be added to any other existing linear combination that sums to  $\vec{b}$  to "invent" a new, different-looking linear combination that sums to  $\vec{b}$ . So, it's false that there is exactly one way to write  $\vec{b}$  as a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$  — if there is one way, there are infinitely many.
- **(False) exactly two ways:** Same logic as above. If this thinking is a bit confusing, see the solution to part c).
- **(True) infinite ways:** For any vector  $\vec{b} \in S$ , there are infinitely many ways to write  $\vec{b}$  as a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$ .

c) (4 pts) Suppose  $\vec{b}$  is some vector in  $S$  such that both of the following equations are true:

$$\begin{aligned}\vec{b} &= 4\vec{x}_1 - 2\vec{x}_2 + 6\vec{x}_3 \\ \vec{b} &= 3\vec{x}_1 + 3\vec{x}_3 - \vec{x}_5\end{aligned}$$

State **one** other linear combination of  $\vec{x}_1, \dots, \vec{x}_6$  that is equal to  $\vec{b}$ . Fill in each box with a number with no variables.

$$\vec{b} = \boxed{\mathbf{7/2}} \vec{x}_1 + \boxed{\mathbf{-1}} \vec{x}_2 + \boxed{\mathbf{9/2}} \vec{x}_3 + \boxed{\mathbf{0}} \vec{x}_4 + \boxed{\mathbf{-1/2}} \vec{x}_5 + \boxed{\mathbf{0}} \vec{x}_6$$

**Solution:**

Arguably, answering part **c)** may have helped clarify the answer to part **b)**. Let's try adding the two representation of  $\vec{b}$  together.

$$\begin{aligned}\vec{b} &= 4\vec{x}_1 - 2\vec{x}_2 + 6\vec{x}_3 \\ \vec{b} &= 3\vec{x}_1 + 3\vec{x}_3 - \vec{x}_5 \\ \implies 2\vec{b} &= 7\vec{x}_1 - 2\vec{x}_2 + 9\vec{x}_3 - \vec{x}_5\end{aligned}$$

Dividing both sides by 2 gives us

$$\boxed{\vec{b} = \frac{7}{2}\vec{x}_1 - \vec{x}_2 + \frac{9}{2}\vec{x}_3 - \frac{1}{2}\vec{x}_5}$$

This is not the only possible answer, but it's probably the easiest one. For example, you could repeat this process with one of the original two  $\vec{b}$ 's along with the new representation of  $\vec{b}$  to get another valid representation of  $\vec{b}$ .

You also could have subtracted the two representations of  $\vec{b}$  to get a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$  that equals the zero vector, which could be added to any other existing linear combination that sums to  $\vec{b}$  to "invent" a new, different-looking linear combination that sums to  $\vec{b}$  (as we said in the solution to part **b)**). If you did this, you'd find that

$$\vec{0} = \vec{x}_1 - 2\vec{x}_2 + 3\vec{x}_3 + \vec{x}_5$$

This must mean that

$$\vec{b} + \vec{0} = (4\vec{x}_1 - 2\vec{x}_2 + 6\vec{x}_3) + (\vec{x}_1 - 2\vec{x}_2 + 3\vec{x}_3 + \vec{x}_5) = 5\vec{x}_1 - 4\vec{x}_2 + 9\vec{x}_3 + \vec{x}_5$$

is another way to represent  $\vec{b}$  as a linear combination of  $\vec{x}_1, \dots, \vec{x}_6$ , and so is

$$\vec{b} + 245(\vec{0}) = (4\vec{x}_1 - 2\vec{x}_2 + 6\vec{x}_3) + 245(\vec{x}_1 - 2\vec{x}_2 + 3\vec{x}_3 + \vec{x}_5)$$

(for instance).

**d)** (4 pts) Let  $T = \text{span}(\{\vec{x}_1, \vec{x}_2, \vec{x}_3\})$  and  $U = \text{span}(\{\vec{x}_4, \vec{x}_5, \vec{x}_6\})$ . Suppose  $W$  is the **intersection** of  $T$  and  $U$ , i.e.  $W = T \cap U$ .  $W$  is also a subspace of  $\mathbb{R}^9$ .

What are the smallest and largest possible values of  $\dim(W)$ , the dimension of  $W$ ? Give your answers as integers.

smallest possible value =

largest possible value =

**Solution:**  $T$  and  $U$  are each individually at most 3-dimensional, since they are each spanned by 3 vectors. So, the intersection  $W$  must be at most 3-dimensional. This means the possible dimensions to consider are 3, 2, 1, or 0. Let's reason about them, starting with 3.

To give examples, we'll use the standard basis vectors  $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_9$  of  $\mathbb{R}^9$ . In  $\mathbb{R}^2$ ,  $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  and  $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ , so (for instance) in  $\mathbb{R}^9$ ,

$$\vec{e}_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

- Could  $\dim(W) = 3$ ? **No.** If  $\dim(W) = 3$ , it would mean that  $T$  and  $U$  are both **the same** 3-dimensional subspace of  $\mathbb{R}^9$ , and intersect everywhere. But if that were the case, then  $S = \text{span}(\{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \vec{x}_5, \vec{x}_6\})$  would be a 3-dimensional subspace of  $\mathbb{R}^9$ , which contradicts the problem statement that  $S$  is a 4-dimensional subspace of  $\mathbb{R}^9$ . So,  $\dim(W) < 3$ , and the maximum possible value is something less than 3.
- Could  $\dim(W) = 2$ ? **Yes**, and all smaller values are also possible.  $T$  and  $U$  could overlap in a 2-dimensional subspace of  $\mathbb{R}^9$ , while each containing a direction that the other doesn't.

For example,  $T$  could be the span of  $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$  and  $U$  could be the span of  $\{\vec{e}_2, \vec{e}_3, \vec{e}_4\}$ . In this case,  $W$  would be the span of  $\{\vec{e}_2, \vec{e}_3\}$ , which is a 2-dimensional subspace of  $\mathbb{R}^9$ .

- Could  $\dim(W) = 1$ ? **Yes.** For example,  $T$  could be the span of  $\{\vec{e}_1, \vec{e}_2\}$  and  $U$  could be the span of  $\{\vec{e}_2, \vec{e}_3, \vec{e}_4\}$ . In this case,  $W$  would be the span of  $\{\vec{e}_2\}$ , which is 1-dimensional, while  $S$  would still be 4-dimensional. (In this example, we said that  $T$  is the span of  $\vec{e}_1$  and  $\vec{e}_2$  though we defined it in the problem statement to be the span of three vectors. No problem — just pick the third vector to be a linear combination of  $\vec{e}_1$  and  $\vec{e}_2$ . That is,  $\vec{x}_1 = \vec{e}_1$ ,  $\vec{x}_2 = \vec{e}_2$ , and  $\vec{x}_3 = \vec{e}_1 + \vec{e}_2$  would work as an example.)
- Could  $\dim(W) = 0$ ? **Yes.**  $T$  and  $U$  could be two completely disjoint subspaces, except for  $\vec{0}$ , which is in every subspace.

For example, let  $\vec{x}_1 = \vec{e}_1$ ,  $\vec{x}_2 = \vec{e}_2$ ,  $\vec{x}_3 = \vec{e}_1 + \vec{e}_2$ , which makes  $T$  a 2-dimensional subspace of  $\mathbb{R}^9$ , and  $\vec{x}_4 = \vec{e}_3$ ,  $\vec{x}_5 = \vec{e}_4$ ,  $\vec{x}_6 = \vec{e}_3 + \vec{e}_4$ , which makes  $U$  a 2-dimensional subspace of  $\mathbb{R}^9$ . In this case,  $W$  would be the set  $\{\vec{0}\}$ , while  $S$  would still be 4-dimensional.

So, the smallest possible value of  $\dim(W)$  is  $\boxed{0}$ , and the largest possible value is  $\boxed{2}$ .

**Problem 7 (20 pts)**

Suppose we'd like to find the optimal constant parameter,  $w^*$ , for the constant model  $h(x_i) = w$ , given a dataset of  $n$  points  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . To do so, we use the **sub-squared** loss function,  $L_{ss}$ , defined below.

$$L_{ss}(y_i, w) = (\sqrt{y_i} - \sqrt{w})^2$$

This requires us to assume that all  $y_i \geq 0$ , as are all possible values of  $w$ .

a) (6 pts) Find  $\frac{d}{dw} R_{ss}(w)$ , the derivative of **average** sub-squared loss (i.e. the empirical risk) with respect to  $w$ . Show your work and **circle** your final answer, which should be an expression in terms of the  $y_i$ 's,  $n$ , and/or any constants. *Hint: The derivative of  $f(x) = \sqrt{x}$  is  $\frac{d}{dx} \sqrt{x} = \frac{1}{2\sqrt{x}}$ .*

**Solution:**

The definition of average sub-squared loss is

$$R_{ss}(w) = \frac{1}{n} \sum_{i=1}^n L_{ss}(y_i, w) = \frac{1}{n} \sum_{i=1}^n (\sqrt{y_i} - \sqrt{w})^2$$

Then,

$$\begin{aligned} \frac{d}{dw} R_{ss}(w) &= \frac{d}{dw} \left( \frac{1}{n} \sum_{i=1}^n (\sqrt{y_i} - \sqrt{w})^2 \right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{d}{dw} [(\sqrt{y_i} - \sqrt{w})^2] \\ &= \frac{1}{n} \sum_{i=1}^n 2(\sqrt{y_i} - \sqrt{w}) \cdot \frac{d}{dw} (\sqrt{y_i} - \sqrt{w}) \\ &= \frac{1}{n} \sum_{i=1}^n 2(\sqrt{y_i} - \sqrt{w}) \cdot \left( 0 - \frac{1}{2\sqrt{w}} \right) \\ &= \boxed{-\frac{1}{n} \sum_{i=1}^n \frac{\sqrt{y_i} - \sqrt{w}}{\sqrt{w}}} \end{aligned}$$

b) (6 pts) Show that the value of  $w^*$  that minimizes average sub-squared loss is

$$w^* = \left( \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \right)^2$$

**Solution:** We've found that

$$\frac{d}{dw} R_{ss}(w) = -\frac{1}{n} \sum_{i=1}^n \frac{\sqrt{y_i} - \sqrt{w}}{\sqrt{w}}$$

To find  $w^*$ , we need to set this expression equal to 0 and solve for  $w$ .

$$\begin{aligned} -\frac{1}{n} \sum_{i=1}^n \frac{\sqrt{y_i} - \sqrt{w}}{\sqrt{w}} &= 0 \\ \sum_{i=1}^n \frac{\sqrt{y_i} - \sqrt{w}}{\sqrt{w}} &= 0 \\ \sum_{i=1}^n (\sqrt{y_i} - \sqrt{w}) &= 0 \\ \sum_{i=1}^n \sqrt{y_i} - n\sqrt{w} &= 0 \\ \sqrt{w} &= \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \\ w &= \left( \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \right)^2 \end{aligned}$$

So, the value of  $w^*$  that minimizes average sub-squared loss is

$$w^* = \left( \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \right)^2$$

c) (6 pts) Using the Cauchy-Schwarz inequality, prove that

$$\left( \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \right)^2 \leq \bar{y}$$

where  $\bar{y}$  is the mean of the  $y_i$ 's.

*Solutions that do not use the Cauchy-Schwarz inequality will not receive credit.*

**Solution:** The Cauchy-Schwarz inequality states that

$$|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\| \|\vec{v}\|$$

for any two vectors  $\vec{u}, \vec{v} \in \mathbb{R}^n$ . The problem boils down to constructing  $\vec{u}$  and  $\vec{v}$  such that the Cauchy-Schwarz inequality, for that  $\vec{u}$  and  $\vec{v}$ , implies the inequality we're trying to prove.

For hints on how to proceed, let's expand the definition of  $\bar{y}$  in the inequality we're trying to prove.

$$\left( \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \right)^2 \leq \frac{1}{n} \sum_{i=1}^n y_i$$

On the left, we have a sum of  $\sqrt{y_i}$ 's, and on the right, we have a sum of  $y_i$ 's. We know that in the norm of a vector, the individual components are squared, which would allow us to turn  $\sqrt{y_i}$  into  $y_i$ . So, one possible path forward is

$$\vec{u} = \begin{bmatrix} \sqrt{y_1} \\ \sqrt{y_2} \\ \vdots \\ \sqrt{y_n} \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The dot product of  $\vec{u}$  and  $\vec{v}$  is  $\sum_{i=1}^n \sqrt{y_i}$ , which seems promising. Let's plug  $\vec{u}$  and  $\vec{v}$  into the Cauchy-Schwarz inequality and see what we get.

$$\begin{aligned} |\vec{u} \cdot \vec{v}| &\leq \|\vec{u}\| \|\vec{v}\| \\ \left| \sum_{i=1}^n \sqrt{y_i} \right| &\leq \left\| \begin{bmatrix} \sqrt{y_1} \\ \sqrt{y_2} \\ \vdots \\ \sqrt{y_n} \end{bmatrix} \right\| \left\| \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\| \\ \sum_{i=1}^n \sqrt{y_i} &\leq \sqrt{\sum_{i=1}^n y_i} \sqrt{n} \end{aligned}$$

Seems like we're getting somewhere. Let's square both sides.

$$\left( \sum_{i=1}^n \sqrt{y_i} \right)^2 \leq \left( \sqrt{\sum_{i=1}^n y_i} \sqrt{n} \right)^2$$

$$\left( \sum_{i=1}^n \sqrt{y_i} \right)^2 \leq n \sum_{i=1}^n y_i$$

Now, all that's left is to divide both sides by  $n^2$ .

$$\left( \sum_{i=1}^n \sqrt{y_i} \right)^2 \leq n \sum_{i=1}^n y_i$$

$$\frac{\left( \sum_{i=1}^n \sqrt{y_i} \right)^2}{n^2} \leq \frac{n \sum_{i=1}^n y_i}{n^2}$$

$$\left( \frac{1}{n} \sum_{i=1}^n \sqrt{y_i} \right)^2 \leq \frac{1}{n} \sum_{i=1}^n y_i$$

This is exactly the inequality we're trying to prove, so we're done!

d) (2 pts) What is the value of  $w$  that minimizes the following function:

$$R(w) = \frac{1}{n} \sum_{i=1}^n (y_i^4 - w^4)^2$$

*Hint: This can be done without using any calculus — don't try and take the derivative.*

○  $\left( \frac{1}{n} \sum_{i=1}^n y_i \right)^4$    ○  $\left( \frac{1}{n} \sum_{i=1}^n y_i^{1/4} \right)^4$    ●  $\left( \frac{1}{n} \sum_{i=1}^n y_i^4 \right)^{1/4}$    ○  $\left( \frac{1}{n} \sum_{i=1}^n y_i^{1/2} \right)^{1/4}$    ○  $\left( \frac{1}{n} \sum_{i=1}^n y_i^4 \right)^{1/2}$

**Solution:**

The idea here is to make a substitution that reduces the problem to one we've already seen — the problem of minimizing mean squared error for the constant model.

Let  $z_i = y_i^4$ , and let  $t = w^4$ . Then,

$$\frac{1}{n} \sum_{i=1}^n (z_i - t)^2 = \frac{1}{n} \sum_{i=1}^n (y_i^4 - w^4)^2$$

What is  $t^*$ , the minimizer of  $\frac{1}{n} \sum_{i=1}^n (z_i - t)^2$ ? That's  $\bar{z}$ , which is

$$t^* = \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i = \frac{1}{n} \sum_{i=1}^n y_i^4$$

But,  $t = w^4$ , so  $w = t^{1/4}$ , meaning

$$w^* = \left[ \left( \frac{1}{n} \sum_{i=1}^n y_i^4 \right) \right]^{1/4}$$

Notice how this relates to parts **a**) and **b**) — those could have been solved the same way, if you wrote  $\sqrt{x}$  as  $x^{1/2}$ .

Congrats on finishing Midterm 1!

Feel free to draw us a picture about EECS 245 in the box below (or use it for scratch work).