Mock Midterm 1 Solutions

EECS 245

Name: ______________________________

uniqname: ______________________________

UMID: ______________________________

Room: ◯ 1014 DOW ◯ At Home

Instructions

- This exam consists of 7 questions. **On the real midterm, we will also state the total number of points for each question. We make no guarantees on the number of questions, points, or specific questions on the real midterm.**
- 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 in the space provided.
- For free response problems, you must show all of your work (unless otherwise specified), 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 a single two-sided 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.

Problem 1: Doubling Down

Suppose we'd like to find the optimal 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 **doubly squared** loss function, L_{ds}, defined below.

$$L_{\text{ds}}(y_i, w) = (y_i^2 - w^2)^2$$

a) Find $\frac{\mathrm{d}R_{\text{ds}}}{\mathrm{d}w}$, the derivative of average doubly squared loss (i.e. the empirical risk) with respect to w.

Solution: Recall that

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

Differentiate term-by-term:

$$\begin{aligned}\frac{\mathrm{d}R_{\text{ds}}}{\mathrm{d}w} &= \frac{1}{n}\sum_{i=1}^{n} 2(y_i^2 - w^2)(-2w) \\ &= -\frac{4w}{n}\sum_{i=1}^{n}(y_i^2 - w^2)\end{aligned}$$

b) Show that the value of w that minimizes average doubly squared loss is

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

Solution: Using the derivative from part **a)**,

$$\frac{\mathrm{d}R_{\text{ds}}}{\mathrm{d}w} = -\frac{4w}{n}\sum_{i=1}^{n}(y_i^2 - w^2)$$

Critical points satisfy

$$w = 0$$

or

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

Since the loss depends on w^2, the minimum occurs when w^2 equals this quantity. One minimizing value is

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

Problem 2: Absolutely...

Consider a dataset of 3 values, $y_1 < y_2 < y_3$, with a mean of 2. Let

$$Y_{\text{abs}}(w) = \frac{1}{3}\sum_{i=1}^{3} |y_i - w|$$

represent the mean absolute error of a constant prediction w on this dataset of 3 values.

Similarly, consider another dataset of 5 values, $z_1 < z_2 < z_3 < z_4 < z_5$, with a mean of 12. Let

$$Z_{\text{abs}}(w) = \frac{1}{5}\sum_{i=1}^{5} |z_i - w|$$

represent the mean absolute error of a constant prediction w on this dataset of 5 values.

Suppose that $y_3 < z_1$, and that $T_{\text{abs}}(w)$ represents the mean absolute error of a constant prediction w on the combined dataset of 8 values, $y_1, y_2, y_3, z_1, z_2, z_3, z_4, z_5$.

a) Fill in the blanks to complete the sentence:

___(i)___ minimizes $Y_{\text{abs}}(w)$, ___(ii)___ minimizes $Z_{\text{abs}}(w)$, and ___(iii)___ minimizes $T_{\text{abs}}(w)$.

Note that in the options below, $[a, b]$ represents the range of values between a and b, including both a and b.

(i) ○ y_1 ○ any value in $[y_1, y_2]$ ● y_2 ○ y_3 ○ z_1

(ii) ○ z_1 ○ z_2 ○ any value in $[z_2, z_3]$ ○ any value in $[z_3, z_4]$ ● z_3

(iii) ○ y_2 ○ y_3 ○ any value in $[y_3, z_1]$ ● any value in $[z_1, z_2]$
○ any value in $[z_2, z_3]$

Solution: Mean absolute error is minimized by a median.

- The median of $y_1 < y_2 < y_3$ is y_2
- The median of $z_1 < z_2 < z_3 < z_4 < z_5$ is z_3
- In the combined sorted list, the 4th and 5th values are z_1 and z_2, so any value in $[z_1, z_2]$ minimizes $T_{\text{abs}}(w)$

b) For any value w, it's true that

$$T_{\text{abs}}(w) = \alpha Y_{\text{abs}}(w) + \beta Z_{\text{abs}}(w)$$

for some constants α and β. Determine the values of α and β. Both answers should be integers or simplified fractions with no variables.

$\alpha =$ $\frac{3}{8}$ $\beta =$ $\frac{5}{8}$

Solution: The combined dataset has 8 values, so

$$\begin{aligned} T_{\text{abs}}(w) &= \frac{1}{8}\left(\sum_{i=1}^{3}|y_i - w| + \sum_{i=1}^{5}|z_i - w|\right) \\ &= \frac{3}{8}\left(\frac{1}{3}\sum_{i=1}^{3}|y_i - w|\right) + \frac{5}{8}\left(\frac{1}{5}\sum_{i=1}^{5}|z_i - w|\right) \\ &= \frac{3}{8}Y_{\text{abs}}(w) + \frac{5}{8}Z_{\text{abs}}(w) \end{aligned}$$

So

$$\alpha = \frac{3}{8} \qquad \beta = \frac{5}{8}$$

c) Show that $Y_{\text{abs}}(z_1) = z_1 - 2$.
Hint: Use the fact that you know the mean of y_1, y_2, y_3.

Solution: Since $y_3 < z_1$, all three y-values are to the left of z_1. So

$$\begin{aligned} Y_{\text{abs}}(z_1) &= \frac{1}{3}\left((z_1 - y_1) + (z_1 - y_2) + (z_1 - y_3)\right) \\ &= \frac{1}{3}\left(3z_1 - (y_1 + y_2 + y_3)\right) \end{aligned}$$

The mean of y_1, y_2, y_3 is 2, so

$$\frac{y_1 + y_2 + y_3}{3} = 2$$

Substitute this in:

$$Y_{\text{abs}}(z_1) = z_1 - 2$$

d) Suppose the minimum possible **output** of $T_{\text{abs}}(w)$ in the full dataset of 8 values is 6. What is the value of z_1?
Hint: You'll need to use the answers to the previous parts.

○ -2 ○ 0 ○ 2 ● 3 ○ 5 ○ 6 ○ 7 ○ 9

Solution: From part **a)**, any value in $[z_1, z_2]$ minimizes $T_{\text{abs}}(w)$, so in particular

$$T_{\text{abs}}(z_1) = 6$$

Using part **b)**,

$$6 = \frac{3}{8}Y_{\text{abs}}(z_1) + \frac{5}{8}Z_{\text{abs}}(z_1)$$

From part **c)**,

$$Y_{\text{abs}}(z_1) = z_1 - 2$$

Also,

$$\begin{aligned} Z_{\text{abs}}(z_1) &= \frac{1}{5}\left((z_2 - z_1) + (z_3 - z_1) + (z_4 - z_1) + (z_5 - z_1)\right) \\ &= \frac{1}{5}\left((z_1 + z_2 + z_3 + z_4 + z_5) - 5z_1\right) \\ &= 12 - z_1 \end{aligned}$$

So

$$\begin{aligned} 6 &= \frac{3}{8}(z_1 - 2) + \frac{5}{8}(12 - z_1) \\ 48 &= 3z_1 - 6 + 60 - 5z_1 \\ 48 &= 54 - 2z_1 \\ z_1 &= 3 \end{aligned}$$

Problem 3: Switcheroo

Consider the following datasets, both consisting of $n = 8$ points.

- "Old" dataset: $(\mathbf{3}, \mathbf{8}), (\mathbf{7}, \mathbf{2}), (x_3, y_3), \dots, (x_8, y_8)$
- "New" dataset: $(\mathbf{3}, \mathbf{2}), (\mathbf{7}, \mathbf{8}), (x_3, y_3), \dots, (x_8, y_8)$

Note that the only difference between the datasets is that the first two y-values have been swapped.

a) Which of the following quantities **are guaranteed to be different** between the old and new datasets? Select all that apply.

- [ ] The mean of the x-values, $\bar{x}$
- [ ] The mean of the y-values, $\bar{y}$
- [ ] The variance of the x-values, σ_x^2
- [ ] The variance of the y-values, σ_y^2
- [x] The correlation coefficient between the x-values and the y-values, r

Solution: Swapping two y-values does not change the list of x-values, so $\bar{x}$ and σ_x^2 stay the same. It also does not change the multiset of y-values, so $\bar{y}$ and σ_y^2 stay the same. What can change is the pairing between the x-values and the y-values. That changes the covariance term, so the correlation coefficient r is the only quantity here that is guaranteed to be different.

b) Let m_{old} and m_{new} be the slopes of the regression lines fit to the old and new datasets, respectively. Given that $\sigma_x^2 = 50$, find the value of $\boxed{m_{\text{new}} - m_{\text{old}}}$. Your final answer should be a number with no variables.

Solution: Recall the slope formula

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

The denominator is the same for both datasets, so

$$m_{\text{new}} - m_{\text{old}} = \frac{\Delta}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where Δ is the change in the numerator.
Only the first two points change, so

$$\begin{aligned}\Delta &= (3 - \bar{x})(2) + (7 - \bar{x})(8) - (3 - \bar{x})(8) - (7 - \bar{x})(2) \\ &= 24\end{aligned}$$

Also,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = n\sigma_x^2 = 8 \cdot 50 = 400$$

So

$$m_{\text{new}} - m_{\text{old}} = \frac{24}{400} = \frac{3}{50}$$

Problem 4: Variability

Let $\vec{x} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ and $y = \begin{bmatrix} 3 \\ c \end{bmatrix}$, where $c \in \mathbb{R}$ is a constant.

a) Fill in the blanks to complete the sentence:

The value of c that makes $|\vec{x} \cdot \vec{y}|$ as small as possible is ___(i)___; when using that value of c, $\vec{x}$ and $\vec{y}$ are ___(ii)___.

(i) **4**

(ii) **orthogonal** (1-3 words)

Solution: We have

$$\vec{x} \cdot \vec{y} = (-4)(3) + 3c = -12 + 3c$$

The smallest possible value of $|\vec{x} \cdot \vec{y}|$ is 0, so we set

$$-12 + 3c = 0$$

This gives

$$c = 4$$

When the dot product is 0, the vectors are orthogonal.

b) Suppose the projection of $\vec{y}$ onto $\vec{x}$ is $\begin{bmatrix} -12/5 \\ 9/5 \end{bmatrix}$, for some value of c. What is the value of c? Show your work, and circle your final answer, which should be a number with no variables.

Solution: Use the projection formula:

$$\text{proj}_{\vec{x}}\vec{y} = \frac{\vec{y}\cdot\vec{x}}{\vec{x}\cdot\vec{x}}\vec{x}$$

Since

$$\vec{x}\cdot\vec{x} = (-4)^2 + 3^2 = 25$$

and

$$\vec{y}\cdot\vec{x} = -12 + 3c$$

we get

$$\text{proj}_{\vec{x}}\vec{y} = \frac{-12+3c}{25}\begin{bmatrix}-4\\3\end{bmatrix}$$

We are told this equals

$$\begin{bmatrix}-12/5\\9/5\end{bmatrix} = \frac{3}{5}\begin{bmatrix}-4\\3\end{bmatrix}$$

So

$$\frac{-12+3c}{25} = \frac{3}{5}$$

which gives

$$\begin{aligned}-12+3c &= 15\\ 3c &= 27\\ c &= 9\end{aligned}$$

uniqname: ______________________

As a refresher, $\vec{x} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 3 \\ c \end{bmatrix}$, where $c \in \mathbb{R}$ is a constant.

In the next two parts, suppose θ_c is the angle between $\vec{x}$ and $\vec{y}$. As c gets larger and larger, $\cos\theta_c$ gets closer and closer to L, i.e.

$$\lim_{c\to\infty} \cos\theta_c = L$$

c) What is the value of L?

○ $3/4$ ○ $-3/4$ ● $3/5$ ○ $-3/5$ ○ $4/5$ ○ $-4/5$ ○ None of these

Solution: Using the cosine formula,

$$\cos\theta_c = \frac{\vec{x}\cdot\vec{y}}{||\vec{x}||\,||\vec{y}||} = \frac{-12+3c}{5\sqrt{9+c^2}}$$

As $c \to \infty$, the dominant terms are $3c$ in the numerator and $5c$ in the denominator, so

$$L = \lim_{c\to\infty} \cos\theta_c = \frac{3}{5}$$

d) $\cos^{-1}(L)$ is also equal to the angle between $\vec{x}$ and a particular unit vector, $\vec{u}$. Find $\vec{u}$ and explain your answer.
Hint: This is more of a conceptual question than a computational one.

Solution: As c gets very large, the vector

$$\vec{y} = \begin{bmatrix} 3 \\ c \end{bmatrix}$$

points more and more in the vertical direction. So the angle between $\vec{x}$ and $\vec{y}$ approaches the angle between $\vec{x}$ and

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

This is a unit vector, and

$$\frac{\vec{x}\cdot\vec{u}}{||\vec{x}||\,||\vec{u}||} = \frac{3}{5} = L$$

So $\cos^{-1}(L)$ is the angle between $\vec{x}$ and $\vec{u}$.

Problem 5: Well, It Depends

Let $\vec{u} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 4 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 6 \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} 1 \\ -1 \\ \alpha \\ 2 \end{bmatrix}$, where $\alpha \in \mathbb{R}$ is a constant.

In parts **a)**, **b)**, **c)**, and **d)**, suppose $\boxed{\alpha = 3}$. Fill in the blanks to complete each sentence.

a) The two vectors $\{\vec{w}, \vec{u} - \vec{w}\}$ are ___(i)___, and span a ___(ii)___-dimensional subspace of $\mathbb{R}^4$.

(i) ● linearly independent ○ linearly dependent

(ii) ○ 1 ● 2 ○ 3 ○ 4

Solution: When $\alpha = 3$,

$$\vec{w} = \begin{bmatrix} 1 \\ -1 \\ 3 \\ 2 \end{bmatrix} \qquad \vec{u} - \vec{w} = \begin{bmatrix} 1 \\ 1 \\ -4 \\ 2 \end{bmatrix}$$

These are not scalar multiples of each other, so they are linearly independent. Two linearly independent vectors span a 2-dimensional subspace.

b) The two vectors $\{\vec{v}, 2\vec{v}\}$ are ___(i)___, and span a ___(ii)___-dimensional subspace of $\mathbb{R}^4$.

(i) ○ linearly independent ● linearly dependent

(ii) ● 1 ○ 2 ○ 3 ○ 4

Solution: The vector $2\vec{v}$ is a scalar multiple of $\vec{v}$, so the set is linearly dependent. Its span is just the line through $\vec{v}$, so the dimension is 1.

c) The three vectors $\{\vec{u}, \vec{v}, \vec{w}\}$ are ___(i)___, and span a ___(ii)___-dimensional subspace of $\mathbb{R}^4$.

(i) ● linearly independent ○ linearly dependent

(ii) ○ 1 ○ 2 ● 3 ○ 4

Solution: Still with $\alpha = 3$, suppose

$$a\vec{u} + b\vec{v} = \vec{w}$$

Looking at the second coordinate gives

$$b = -1$$

Then the first coordinate gives

$$2a + 3(-1) = 1$$

so $a = 2$. But then the third coordinate would require

$$-a = 3$$

which is impossible. So $\vec{w}$ is not in span$\{\vec{u}, \vec{v}\}$, and the three vectors are linearly independent.
Three linearly independent vectors span a 3-dimensional subspace.

d) The four vectors $\{\vec{u}, \vec{v}, \vec{w}, \vec{w} - \vec{u}\}$ are ___(i)___, and span a ___(ii)___-dimensional subspace of $\mathbb{R}^4$.

(i) ○ linearly independent ● linearly dependent

(ii) ○ 1 ○ 2 ● 3 ○ 4

Solution: The vector $\vec{w} - \vec{u}$ is already a linear combination of $\vec{w}$ and $\vec{u}$, so adding it does not create anything new. That makes the set linearly dependent.
Its span is the same as the span of $\{\vec{u}, \vec{v}, \vec{w}\}$ from part **c)**, which is 3-dimensional.

Now, suppose $\boxed{\alpha = -2}$. Fill in the blanks to complete each sentence.

e) The three vectors $\{\vec{u}, \vec{v}, \vec{w}\}$ are ___(i)___, and span a ___(ii)___-dimensional subspace of $\mathbb{R}^4$.

(i) ○ linearly independent ● linearly dependent

(ii) ○ 1 ● 2 ○ 3 ○ 4

Solution: Now $\alpha = -2$, so

$$\vec{w} = \begin{bmatrix} 1 \\ -1 \\ -2 \\ 2 \end{bmatrix}$$

Check whether $\vec{w}$ is in span$\{\vec{u}, \vec{v}\}$:

$$2\vec{u} - \vec{v} = 2\begin{bmatrix} 2 \\ 0 \\ -1 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -2 \\ 2 \end{bmatrix} = \vec{w}$$

So the set is linearly dependent. Since $\vec{u}$ and $\vec{v}$ are not multiples of each other, they are linearly independent, so the span has dimension 2.

f) The four vectors $\{\vec{u}, \vec{v}, \vec{w}, \vec{w} - \vec{u}\}$ are ___(i)___, and span a ___(ii)___-dimensional subspace of $\mathbb{R}^4$.

(i) ○ linearly independent ● linearly dependent

(ii) ○ 1 ● 2 ○ 3 ○ 4

Solution: From part **e)**, $\vec{w}$ is already in span$\{\vec{u}, \vec{v}\}$. Then $\vec{w} - \vec{u}$ is also in span$\{\vec{u}, \vec{v}\}$. So all four vectors lie in the same 2-dimensional subspace spanned by $\vec{u}$ and $\vec{v}$.
That means the set is linearly dependent and its span has dimension 2.

Problem 6: Covering All the Bases

a) Find a basis for the following subspace of $\mathbb{R}^3$:

$$S = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \mid 6x - y + z = 0 \right\}$$

Solution: We just need two linearly independent vectors that satisfy

$$6x - y + z = 0$$

Two easy choices are

$$\begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Both are in S, and they are not scalar multiples of each other. So one basis is

$$\left\{ \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

b) The equations $6x - y + z = 0$ and $4x + 4y = 0$ intersect in a line in $\mathbb{R}^3$. Find the equation of this line in **parametric** form.

Solution: From

$$4x + 4y = 0$$

we get

$$y = -x$$

Substitute this into

$$6x - y + z = 0$$

to get

$$\begin{aligned} 6x - (-x) + z &= 0 \\ 7x + z &= 0 \\ z &= -7x \end{aligned}$$

Let $x = t$. Then

$$y = -t \qquad z = -7t$$

So the line is

$$L = \left\{ t \begin{bmatrix} 1 \\ -1 \\ -7 \end{bmatrix} : t \in \mathbb{R} \right\}$$

uniqname: ______________________________

Problem 7: Catchy

Suppose $\vec{x} \in \mathbb{R}^n$. Prove that the L_1 norm of a vector is less than or equal to $\sqrt{n}$ times the L_2 norm of a vector, i.e.

$$\|\vec{x}\|_1 \leq \sqrt{n}\|\vec{x}\|_2$$

Hint: Use the Cauchy-Schwarz inequality, which states that for any two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, $|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\|_2\|\vec{v}\|_2$. Most of your job is to choose the right vectors $\vec{u}$ and $\vec{v}$ to apply the Cauchy-Schwarz inequality to.

Solution: Choose

$$\vec{u} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} |x_1| \\ |x_2| \\ \vdots \\ |x_n| \end{bmatrix}$$

Then

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} |x_i| = \|\vec{x}\|_1$$

Apply Cauchy-Schwarz:

$$\|\vec{x}\|_1 = |\vec{u} \cdot \vec{v}| \leq \|\vec{u}\|\|\vec{v}\|$$

Now

$$\|\vec{u}\| = \sqrt{n} \qquad \|\vec{v}\| = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{\sum_{i=1}^{n} x_i^2} = \|\vec{x}\|$$

So

$$\|\vec{x}\|_1 \leq \sqrt{n}\|\vec{x}\|$$