

Midterm 2 Solutions

EECS 245, Fall 2025 at the University of Michigan

Name: _____

username: _____

UMID: _____

Room: ☐ 1365 LCSIB ☐ 2901 BBB

Instructions

- This exam consists of 7 problems, worth a total of 100 points, spread across 12 pages (6 sheets of paper).
- You have 80 minutes to complete this exam, unless you have extended-time accommodations through SSD.
- Write your username in the top right corner of every 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: Getting Started (12 pts)

- a) (3 pts) Let $A = \begin{bmatrix} 2 & 4 \\ -3 & -7 \end{bmatrix}$. Find $\det(A)$, the determinant of A . Give your answer as an integer.

$$\det(A) = \boxed{-2}$$

Solution: The determinant of a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad - bc$. So,

$$\det(A) = (2)(-7) - (4)(-3) = -14 + 12 = \boxed{-2}$$

- b) (3 pts) Using A from part a), find A^{-1} , the inverse of A . Fully simplify your answer, i.e. don't leave any constants out front.

$$A^{-1} = \boxed{\begin{bmatrix} 7/2 & 2 \\ -3/2 & -1 \end{bmatrix}}$$

Solution: The inverse of a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $\frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. So,

$$A^{-1} = \frac{1}{(-2)} \begin{bmatrix} -7 & -4 \\ 3 & 2 \end{bmatrix} = \boxed{\begin{bmatrix} 7/2 & 2 \\ -3/2 & -1 \end{bmatrix}}$$

- c) (2 pts) Let $B = \begin{bmatrix} -1 & 2 & -1 \\ 3 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. What is the **first column** of B^{-1} , the inverse of B ?

☐ $\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$
☐ $\begin{bmatrix} -1 \\ 1/2 \\ -1 \end{bmatrix}$
☐ $\begin{bmatrix} -1 \\ 1/3 \\ 1 \end{bmatrix}$
☒ $\begin{bmatrix} -1/3 \\ 1/3 \\ 0 \end{bmatrix}$
☐ $\begin{bmatrix} 1/3 \\ -1/3 \\ 0 \end{bmatrix}$
☐ B is not invertible

Solution: $\begin{bmatrix} -1/3 \\ 1/3 \\ 0 \end{bmatrix}$.

Remember, B^{-1} is the matrix that satisfies $BB^{-1} = I$. Inverting B is not necessary: instead, all one needs to look for is the vector \vec{v} such that $B\vec{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, since B multiplied by B^{-1} 's first column should give the first column of I .

And indeed, $B \begin{bmatrix} -1/3 \\ 1/3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. So, you could have solved this just by guessing and checking each of the options.

If we asked this as an open-ended question instead, we'd be searching for the vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ such that

$$\begin{bmatrix} -1 & 2 & -1 \\ 3 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

From here, there are two ways to solve for $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

- You could solve the system of equations directly.
- Or, you could notice that $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ must be orthogonal to both the second row and third row of B , which means that its in the same direction as the cross product of the second and third rows. If you compute the cross product of the last two rows, you get

$$\begin{bmatrix} 3(1) - 2(0) \\ 2(0) - 3(1) \\ 3(0) - 3(0) \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}$$

which means that $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = c \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}$ for some constant c . To find c , solve for the c

such that the dot product of $c \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$ (the first row of B) is 1. This gives

$$c = -1/9, \text{ which means that } \begin{bmatrix} x \\ y \\ z \end{bmatrix} = -1/9 \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 1/3 \\ 0 \end{bmatrix}.$$

username: _____

- d) (4 pts) This part is independent of the previous parts (i.e. don't use the specific A or B from above).

Select all true statements below.

- ☐ If A and B are both matrices such that $AB = I$, then A and B are both invertible.
- ☐ If A and B are both invertible matrices, then $(A^T B)^{-1} = ((B^{-1})^T A^{-1})^T$.
- ☒ If A is an invertible matrix, then $\text{rank}(A) = \text{rank}(A^{-1})$.
- ☐ If A , B , and C are all symmetric matrices, then $AB + C$ is also symmetric.

Solution: Only Option 3 is true. Let's look at each statement one by one.

1. **If A and B are both matrices such that $AB = I$, then A and B are both invertible.**

This is **False**, because it's possible for $AB = I$ to be true for two non-square matrices

A and B , meaning they can't be invertible. For example, suppose $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

and $A = B^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Then,

$$AB = B^T B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

2. **If A and B are both invertible matrices, then $(A^T B)^{-1} = ((B^{-1})^T A^{-1})^T$.** This is **False**:

- If we expand the right-hand side, we get

$$((B^{-1})^T A^{-1})^T = \underbrace{(A^{-1})^T ((B^{-1})^T)^T}_{\text{reverse order of product when taking transpose}} = (A^{-1})^T (B^{-1})$$

- This is not the same as $(A^T B)^{-1}$, which is $(A^T B)^{-1} = B^{-1} (A^T)^{-1}$. Note that $(A^{-1})^T = (A^T)^{-1}$, but the reason these two expressions aren't the same is because order matters for matrix multiplication — it's not commutative.

3. **If A is an invertible matrix, then $\text{rank}(A) = \text{rank}(A^{-1})$.** This is **True**. If A is invertible, then $\text{rank}(A) = n$. Then, A^{-1} is also invertible (its inverse is A), so it must have a rank of n as well.

4. **If A , B , and C are all symmetric matrices, then $AB + C$ is also symmetric.** This is **False**. Recall, what makes a matrix A symmetric is that $A = A^T$. Let's take the transpose of $AB + C$ and see if we end up getting back $AB + C$:

$$(AB + C)^T = (AB)^T + C^T = B^T A^T + C^T = BA + C$$

$AB + C$ is only symmetric if $AB + C = BA + C$, i.e. if $AB = BA$, which is not true in general, even if A and B are both symmetric.

Problem 2: Space Jam (20 pts)

$$\text{Let } X = \begin{bmatrix} 1 & -4 & 2 & 2 & 0 \\ 0 & 0 & -3 & 3 & 0 \\ 1 & -4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

a) (4.5 pts) Determine the values of each of the following. Give your answers as integers.

$\dim(\text{colsp}(X)) =$

3

$\dim(\text{nullsp}(X)) =$

2

$\dim(\text{colsp}(X^T)) =$

3

$\dim(\text{nullsp}(X^T)) =$

1

Solution: Recall, the rank-nullity theorem states that for any matrix X ,

$$\text{rank}(X) + \dim(\text{nullsp}(X)) = \text{number of columns of } X$$

where $\text{rank}(X) = \dim(\text{colsp}(X)) = \dim(\text{colsp}(X^T))$.

X has 3 linearly independent columns: columns 1, 4, and 5. These three columns can be used to create the other two columns:

$$\bullet \text{ Column 2} = \begin{bmatrix} -4 \\ 0 \\ -4 \\ 0 \end{bmatrix} = -4 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = (-4) \cdot \text{column 1}$$

$$\bullet \text{ Column 3} = \begin{bmatrix} 2 \\ -3 \\ 4 \\ 0 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \end{bmatrix} = 4 \cdot \text{column 1} - \text{column 4}$$

So, $\text{rank}(X) = 3$, meaning $\dim(\text{colsp}(X)) = \boxed{3}$ and $\dim(\text{colsp}(X^T)) = \boxed{3}$ also.

Since $\text{rank}(X) + \dim(\text{nullsp}(X)) = \text{number of columns of } X$, we have $\dim(\text{nullsp}(X)) = 5 - 3 = \boxed{2}$.

And finally, since $\text{rank}(X^T) + \dim(\text{nullsp}(X^T)) = \text{number of columns of } X^T$, we have $\dim(\text{nullsp}(X^T)) = 5 - 4 = \boxed{1}$.

b) (3.5 pts) Suppose $\vec{y} \in \mathbb{R}^4$. How many solutions $\vec{v} \in \mathbb{R}^5$ are there to the system of equations $X\vec{v} = \vec{y}$? **Select all** possibilities, since the answer may depend on \vec{y} .

☒ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☒ Infinitely many

Solution: When solving $X\vec{v} = \vec{y}$ for \vec{v} , there are two possible cases.

- $\vec{y} \notin \text{colsp}(X)$: This is possible because $\dim(\text{colsp}(X)) = 3$, so the columns don't span all of \mathbb{R}^4 . In this case, \vec{v} has no solutions.
- $\vec{y} \in \text{colsp}(X)$: The columns of X aren't linearly independent, so there are infinitely many ways to write \vec{y} as a linear combination of the columns of X .

c) (6 pts) For some $\vec{y} \in \mathbb{R}^4$, the vector $\vec{w}' = \begin{bmatrix} 8 \\ 0 \\ 0 \\ 3 \\ 11 \end{bmatrix}$ is such that $X\vec{w}'$ is the vector in $\text{colsp}(X)$ that is

closest to \vec{y} . State **one other** vector $\vec{\beta}$ such that $X\vec{\beta} = X\vec{w}'$. Show your work, and circle your final answer, which should be a vector with five entries and no variables.

Solution: There's two ways to approach this problem. The first is adding a vector in $\text{nullsp}(X)$ to \vec{w}' . Why does this work? Let $\vec{\beta} = \vec{w}' + \vec{n}$, where $X\vec{n} = \vec{0}$:

$$\begin{aligned} X\vec{\beta} &= X(\vec{w}' + \vec{n}) \\ &= X\vec{w}' + X\vec{n} \\ &= X\vec{w}' = \vec{y} \end{aligned}$$

So, all we have to do is find a vector in the null space of

$$X = \begin{bmatrix} 1 & -4 & 2 & 2 & 0 \\ 0 & 0 & -3 & 3 & 0 \\ 1 & -4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

One such vector is $\vec{n} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, since $X\vec{n} = \begin{bmatrix} -4 \\ 0 \\ -4 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \\ 4 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \end{bmatrix} = \vec{0}$

This leaves us with $\vec{\beta} = \vec{w}' + \vec{n} = \begin{bmatrix} 8 \\ 0 \\ 0 \\ 3 \\ 11 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \\ 1 \\ 4 \\ 11 \end{bmatrix}$.

The other way is to "tweak" \vec{w}' using the relationships we know about in the columns of

X . Since column 2 = $-4 \cdot$ column 1, and $\vec{w}' = \begin{bmatrix} 8 \\ 0 \\ 0 \\ 3 \\ 11 \end{bmatrix}$, an easy swap is to change w_0 from

8 to 0 and w_1 from 0 to -2 :

$$8 \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = -2 \cdot \begin{bmatrix} -4 \\ 0 \\ -4 \\ 0 \end{bmatrix}$$

Doing this gives $\vec{\beta} = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 3 \\ 11 \end{bmatrix}$.

Recall, $X = \begin{bmatrix} 1 & -4 & 2 & 2 & 0 \\ 0 & 0 & -3 & 3 & 0 \\ 1 & -4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

- d) (6 pts) Find a basis for $\text{nullsp}(X^T)$ (**not** $\text{nullsp}(X)$). Show your work, and circle your final answer, which should be a list of vectors.

Solution:

$$X^T = \begin{bmatrix} 1 & 0 & 1 & 0 \\ -4 & 0 & -4 & 0 \\ 2 & -3 & 4 & 0 \\ 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

From the rank-nullity theorem, we know that our basis will have exactly one vector, so our goal is to find a non-zero vector where $X^T \vec{n} = \vec{0}$.

Column 3 = Column 1 $- \frac{2}{3} \cdot$ Column 2, so one possible basis is $\left\{ \begin{bmatrix} 1 \\ -\frac{2}{3} \\ -1 \\ 0 \end{bmatrix} \right\}$.

Problem 3: Nilpotence (12 pts)

Suppose A is an $n \times n$ matrix such that $A^2 = 0_{n \times n}$, where $0_{n \times n}$ is an $n \times n$ matrix of all zeros.

- a) (6 pts) Prove that if $\vec{x} \in \text{colsp}(A)$, then $\vec{x} \in \text{nullsp}(A)$.

Solution: If $\vec{x} \in \text{colsp}(A)$, then $\vec{x} = A\vec{v}$ for some $\vec{v} \in \mathbb{R}^n$. Then, multiplying both sides of $\vec{x} = A\vec{v}$ by A on the left gives us:

$$A\vec{x} = A(A\vec{v}) = A^2\vec{v} = 0_{n \times n}\vec{v} = \vec{0}$$

Since $\vec{x} = A\vec{v} \implies A\vec{x} = \vec{0}$, we have $\vec{x} \in \text{nullsp}(A)$.

- b) (6 pts) In part a), you showed that $\text{colsp}(A)$ is a subset of $\text{nullsp}(A)$. Using this fact, find the **maximum** possible value of $\text{rank}(A)$. Show your work and circle your final answer, which should be an expression involving n and/or constants.

Solution: In the previous part, we showed that every element in $\text{colsp}(A)$ is also in $\text{nullsp}(A)$. (The converse is not true.) Intuitively, this means that the column space is a subset of the null space, so it's "smaller" than the null space.

This means that

$$\dim(\text{colsp}(A)) \leq \dim(\text{nullsp}(A))$$

or in other words

$$\text{rank}(A) \leq \dim(\text{nullsp}(A))$$

Let's add $\text{rank}(A)$ to both sides of the inequality; this will make the right-hand side look like something involved in the rank-nullity theorem.

$$\text{rank}(A) + \text{rank}(A) \leq \text{rank}(A) + \dim(\text{nullsp}(A)) = n$$

This tells us that $2\text{rank}(A) \leq n$, so $\text{rank}(A) \leq \frac{n}{2}$ and so $\frac{n}{2}$ is the maximum possible value of $\text{rank}(A)$.

Problem 4: Poly Wants a Cracker (18 pts)

Suppose we'd like to fit the model $h(x_i) = w_0 + w_1 x_i + w_2 x_i^2$ by minimizing mean squared error. We use an observation vector $\vec{y} \in \mathbb{R}^n$, but instead of using the regular design matrix X ,

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \vec{x}^{(0)} & \vec{x}^{(1)} & \vec{x}^{(2)} \\ | & | & | \end{bmatrix}$$

we use the **centered** design matrix Z (where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is the mean of the x 's).

$$Z = \begin{bmatrix} 1 & x_1 - \bar{x} & (x_1 - \bar{x})^2 \\ 1 & x_2 - \bar{x} & (x_2 - \bar{x})^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n - \bar{x} & (x_n - \bar{x})^2 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \vec{z}^{(0)} & \vec{z}^{(1)} & \vec{z}^{(2)} \\ | & | & | \end{bmatrix}$$

- a) (6 pts) It turns out that $\text{colsp}(Z) = \text{colsp}(X)$. To show this, fill in the blanks below to express $\vec{z}^{(2)}$ (the third column of Z) as a linear combination of X 's columns. Each box should be filled with an expression involving \bar{x} , n , and/or constants.

$$\vec{z}^{(2)} = \boxed{\bar{x}^2} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \boxed{(-2\bar{x})} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \boxed{\mathbf{1}} \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix}$$

Solution: $\vec{z}^{(2)}$ is the column made up of terms of the form $(x_i - \bar{x})^2$. Note that

$$(x_i - \bar{x})^2 = x_i^2 - 2\bar{x}x_i + \bar{x}^2 = (\bar{x}^2)(1) + (-2\bar{x})(x_i) + (1)(x_i^2)$$

which tells us that

$$\vec{z}^{(2)} = \bar{x}^2 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} - 2\bar{x} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix}$$

- b) (6 pts) In this part only, assume that the values x_1, x_2, \dots, x_n are each either 1 or 0. For some specific values x_1, x_2, \dots, x_n , the matrix P that projects vectors in \mathbb{R}^n onto $\text{colsp}(Z)$ is given by

$$P = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

username: _____

(i) What is the rank of Z ? Give your answer as an integer. $\text{rank}(Z) =$

2

(ii) Which specific values of x_1, x_2, \dots, x_n result in P being the matrix above? Give your answer as a list of values, in the order x_1 , then x_2 , then x_3 , etc. (If there are multiple possible answers, just give one.)

Solution:

First, $\text{rank}(Z) = 2$. We're told in part **a)** that $\text{colsp}(Z) = \text{colsp}(X)$, so $\text{rank}(Z) = \text{rank}(X)$. I find it easier to think in terms of X since the numbers are more straightforward.

Remember, throughout this part, that each x_i is either 1 or 0! This means that the

column $\vec{x}^{(1)} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is made up of 1's and 0's, and the column $\vec{x}^{(2)} = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix}$ is made up

of 1's and 0's in the same positions, since $1^2 = 1$ and $0^2 = 0$.

So, X only really has two unique columns, and its rank is 2. But since $\text{rank}(Z) = \text{rank}(X)$, we have $\text{rank}(Z) = 2$. Z doesn't have any repeated columns, but as we showed above, it's still the case that one of Z 's columns is a linear combination of the other two.

The only case in which $\text{rank}(Z) = 1$ is if all of the x_i are the same, but the matrix P tells us that that is not the case.

Let's now look at the matrix P . Notice that rows 1, 2, and 4 of P are identical, as are rows 3 and 5. Let's imagine some vector $\vec{y} \in \mathbb{R}^5$. What would multiplying P by \vec{y} give us?

$$P\vec{y} = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_4 \\ \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_4 \\ \frac{1}{2}y_3 + \frac{1}{2}y_5 \\ \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_4 \\ \frac{1}{2}y_3 + \frac{1}{2}y_5 \end{bmatrix} = \begin{bmatrix} \text{mean of } y_1, y_2, y_4 \\ \text{mean of } y_1, y_2, y_4 \\ \text{mean of } y_3, y_5 \\ \text{mean of } y_1, y_2, y_4 \\ \text{mean of } y_3, y_5 \end{bmatrix}$$

We know from Chapter 1 that the mean is the constant that minimizes mean squared error. Here, it appears that the prediction returned in \vec{y} is not always the same, but is one of two possibilities — rows 1, 2, and 4 have the same prediction, and rows 3 and 5 have the same prediction. This hints to us that rows 1, 2, and 4 come from the same x_i value, and rows 3 and 5 come from the same x_i value, and the optimal prediction is some **conditional** mean. This resembles [Lab 9, Activity 2](#), on one hot encoding with beef, chicken, and fish.

The above observation alone is enough information to answer the question. The two possible answers are $\boxed{x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 1, x_5 = 0}$ and

$\boxed{x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0, x_5 = 1}$.

Let's dive deeper into the math to confirm this. Let's start with what X would have had to be. (We can work with X instead of Z since both have the same column spaces, so projecting onto either column space will give us the same result; X is just easier to work with.) And, let's drop $\vec{x}^{(2)}$ from X , since including it will prevent $X^T X$ from being invertible while not changing $\text{colsp}(X)$.

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Note that I arbitrarily picked $x_1 = x_2 = x_4 = 1$ and $x_3 = x_5 = 0$, but we could reverse the 1's and 0's and P would turn out to be the same.

The formula for the projection matrix is $P = X(X^T X)^{-1} X^T$. I won't include all of the algebra here, but if you work out $P = X(X^T X)^{-1} X^T$, you'll find that P is indeed the matrix provided in the problem.

Here's one final interpretation of what's going on. Suppose the optimal parameters for this X and some \vec{y} are $\vec{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \end{bmatrix}$, which would lead to a hypothesis function of

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

This hypothesis function only returns one of two values:

- If $x_i = 1$, then $h(1) = w_0^* + w_1^*$
- If $x_i = 0$, then $h(0) = w_0^*$

So, w_0^* is the mean of the y 's when $x_i = 0$, and $w_0^* + w_1^*$ is the mean of the y 's when $x_i = 1$. This is exactly what we see in the matrix P .

Recall, $Z = \begin{bmatrix} 1 & x_1 - \bar{x} & (x_1 - \bar{x})^2 \\ 1 & x_2 - \bar{x} & (x_2 - \bar{x})^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n - \bar{x} & (x_n - \bar{x})^2 \end{bmatrix} = \begin{bmatrix} | & | & | \\ \vec{z}^{(0)} & \vec{z}^{(1)} & \vec{z}^{(2)} \\ | & | & | \end{bmatrix}.$

c) (6 pts) Let $\vec{\beta}^* = \begin{bmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \end{bmatrix}$ be a solution to the normal equations for Z and \vec{y} . Show that

$$\beta_0^* = \bar{y} - \beta_2^* \sigma_x^2$$

where $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ is the variance of the x 's, and \bar{y} is the mean of the y 's. *Hint: Use the fact that $\sum_{i=1}^n (x_i - \bar{x}) = 0$. What is the error vector? Is it orthogonal to something useful?*

Solution: The error vector is $\vec{e} = \vec{y} - Z\vec{\beta}^*$. As we studied in depth, the error vector is orthogonal to every vector in $\text{colsp}(Z)$, i.e. every linear combination of the columns of Z . Z has a column of all 1's, so the error vector is orthogonal to that, too.

$$(\vec{y} - Z\vec{\beta}^*) \cdot \vec{1} = 0$$

We'll proceed by expanding $Z\vec{\beta}^*$ and then plugging the result into the above. This will allow us to solve for β_0^* .

$$\begin{aligned} Z\vec{\beta}^* &= \beta_0^* \vec{z}^{(0)} + \beta_1^* \vec{z}^{(1)} + \beta_2^* \vec{z}^{(2)} \\ &= \beta_0^* \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1^* \begin{bmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \\ \vdots \\ x_n - \bar{x}_n \end{bmatrix} + \beta_2^* \begin{bmatrix} (x_1 - \bar{x}_1)^2 \\ (x_2 - \bar{x}_2)^2 \\ \vdots \\ (x_n - \bar{x}_n)^2 \end{bmatrix} \\ &= \begin{bmatrix} \beta_0^* + \beta_1^*(x_1 - \bar{x}_1) + \beta_2^*(x_1 - \bar{x}_1)^2 \\ \beta_0^* + \beta_1^*(x_2 - \bar{x}_2) + \beta_2^*(x_2 - \bar{x}_2)^2 \\ \vdots \\ \beta_0^* + \beta_1^*(x_n - \bar{x}_n) + \beta_2^*(x_n - \bar{x}_n)^2 \end{bmatrix} \end{aligned}$$

$$(\vec{y} - Z\vec{\beta}^*) \cdot \vec{1} = 0$$

$$\sum_{i=1}^n [y_i - (Z\vec{\beta}^*)_i] = 0$$

$$\sum_{i=1}^n [y_i - \beta_0^* - \beta_1^*(x_i - \bar{x}_i) - \beta_2^*(x_i - \bar{x}_i)^2] = 0$$

$$\underbrace{\sum_{i=1}^n y_i}_{n\bar{y}} - \underbrace{\sum_{i=1}^n \beta_0^*}_{\text{sum of constant}} - \underbrace{\sum_{i=1}^n \beta_1^*(x_i - \bar{x}_i)}_0 - \sum_{i=1}^n \beta_2^*(x_i - \bar{x}_i)^2 = 0$$

$$n\bar{y} - n\beta_0^* - n\beta_2^*\sigma_x^2 = 0$$

$$n\beta_0^* = n\bar{y} - n\beta_2^*\sigma_x^2$$

$$\beta_0^* = \boxed{\bar{y} - \beta_2^*\sigma_x^2}$$

Problem 5: Ortho...dontist? (12 pts)

Let $A = \begin{bmatrix} 1 & 0 \\ 1 & 4 \\ 1 & 4 \\ 1 & 4 \end{bmatrix}$.

- a) (6 pts) Find a matrix Q such that $\text{colsp}(Q) = \text{colsp}(A)$ and $Q^T Q = I$. Show your work and circle your final answer, which should be a matrix with two columns and no variables. *Hint: One of the columns may involve square roots.*

Solution: Since we want $Q^T Q = I$, we're looking for a matrix Q with two columns that are orthogonal to each other and are both unit vectors.

The "standard" way to answer this part is to use the Gram-Schmidt process, first introduced in Homework 7, Problem 4. But, since A only has two columns, it's okay if you forgot about the specifics, and instead realized the core of Gram-Schmidt, which takes advantage of the fact that **the error when projecting \vec{u} onto \vec{v} is orthogonal to \vec{v} .**

Let $\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Then, what's the error vector from projecting $\vec{u} = \begin{bmatrix} 0 \\ 4 \\ 4 \\ 4 \end{bmatrix}$ onto \vec{v} ?

$$\vec{p} = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v} = \frac{0 \cdot 1 + 4 \cdot 1 + 4 \cdot 1 + 4 \cdot 1}{1^2 + 1^2 + 1^2 + 1^2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{12}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{e} = \vec{u} - \vec{p} = \begin{bmatrix} 0 \\ 4 \\ 4 \\ 4 \end{bmatrix} - 3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

This vector \vec{e} is orthogonal to \vec{v} , and together \vec{v} and \vec{e} have the same span as $\text{colsp}(A)$. So, one solution is to form Q by normalizing \vec{v} and \vec{e} . Doing so gives us

$$Q = \begin{bmatrix} 1/\sqrt{2} & -3/\sqrt{10} \\ 1/\sqrt{2} & 1/\sqrt{10} \\ 1/\sqrt{2} & 1/\sqrt{10} \\ 1/\sqrt{2} & 1/\sqrt{10} \end{bmatrix}$$

Another common solution was to observe that the vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ are orthogonal

to each other and span $\text{colsp}(A)$ (meaning, you can make both of A 's columns as linear combinations of these two vectors). Normalizing these two vectors gives us another Q :

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{bmatrix}$$

b) (2 pts) True or False: The matrix Q you found above is an orthogonal matrix.

☐ True ☒ False

Solution: No matter how you find Q in part **a**), the answer is false, because Q is not a square matrix, so it cannot be orthogonal!

For Q to be orthogonal, we'd need **both** $Q^T Q = I$ **and** $Q Q^T = I$. Since Q is not square, these can't both be true at the same time (the dimensions don't match, since the former would be 2×2 while the latter would be 4×4).

- c) (4 pts) Let $R = \begin{bmatrix} r_1 & \boxed{r_2} \\ \boxed{r_3} & r_4 \end{bmatrix}$ be a 2×2 matrix such that $A = QR$, where Q is the matrix you found above.

Find r_2 and r_3 . Give your answers as scalars without variables.

$$r_2 = \boxed{}, \quad r_3 = \boxed{}$$

Solution: We ended up giving full credit to everyone for this problem, since there's no unique answer, and it's difficult to answer this correctly if you found an invalid Q .

The main idea being assessed here, taken from Homework 7, Problem 4, is that if Q is a matrix such that $\text{colsp}(Q) = \text{colsp}(A)$ and $Q^T Q = I$, then

$$A = QR \implies Q^T A = Q^T QR \implies R = Q^T A$$

As we saw in that homework problem, if you use **Gram-Schmidt** to find Q , R is an **upper triangular** matrix, meaning that $r_3 = 0$. (We won't elaborate on this here: read the solutions to Homework 7, Problem 4.)

For two different Q 's, we'll find the corresponding R 's to give you some sample possible answers.

- For $Q = \begin{bmatrix} 1/2 & -3/\sqrt{10} \\ 1/2 & 1/\sqrt{10} \\ 1/2 & 1/\sqrt{10} \\ 1/2 & 1/\sqrt{10} \end{bmatrix}$, which did result from Gram-Schmidt,

$$R = Q^T A = \begin{bmatrix} 2 & 6 \\ 0 & 12/\sqrt{10} \end{bmatrix}$$

This R is upper triangular.

- For $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{bmatrix}$, which did not result from Gram-Schmidt,

$$R = Q^T A = \begin{bmatrix} 1 & 0 \\ \sqrt{3} & 4\sqrt{3} \end{bmatrix}$$

This R is **not** upper triangular.

Problem 6: Quadratus Formulus (14 pts)

Let $f(\vec{x}) = \frac{1}{2}\vec{x}^T S \vec{x} - \vec{b}^T \vec{x}$, where S is a symmetric $n \times n$ matrix and $\vec{b} \in \mathbb{R}^n$.

- a) (4 pts) Find $\nabla f(\vec{x})$, the gradient of $f(\vec{x})$. Show your work, and circle your final answer, which should be an expression in terms of \vec{x} , S , \vec{b} , and/or constants. *Hint: There's no need to re-prove gradient rules from class.*

Solution:

$$\begin{aligned}\nabla f(\vec{x}) &= \nabla_{\vec{x}} \left(\frac{1}{2} \vec{x}^T S \vec{x} \right) - \nabla_{\vec{x}} \left(\vec{b}^T \vec{x} \right) \\ &= \frac{1}{2} (2S\vec{x}) - \vec{b} \\ &= \boxed{S\vec{x} - \vec{b}}\end{aligned}$$

- b) (2 pts) True or False: As long as S is invertible, if $\nabla f(\vec{a}) = \vec{0}$, then \vec{a} is a global minimum of $f(\vec{x})$.

☐ True ☒ False

Solution: In general, this is **false**. Even if S is invertible, $\nabla f(\vec{a}) = \vec{0}$ could mean that \vec{a} is at a local maxima, local minima, or saddle point.

For example, let $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and $S = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, which is an invertible matrix. Then,

$$f(\vec{x}) = \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = x^2 - y^2$$

but $f(\vec{x}) = x^2 - y^2$ has no global minimum, since you can make $f(\vec{x})$ arbitrarily negative by setting $x = 0$ and $y = -\text{large number}$.

- c) (2 pts) True or False: As long as all of the components of S are positive real numbers, if $\nabla f(\vec{a}) = \vec{0}$, then \vec{a} is a global minimum of $f(\vec{x})$.

☐ True ☒ False

Solution: This is also **false**. Even if all of the components of S are positive real numbers, $f(\vec{x})$ may not have a global minimum. As we saw later in the semester, the convexity of f has to do with whether or not S is **positive semidefinite**. But, this was not a concept we knew about on the midterm, so the problem is answerable without that concept.

Instead, the way to think through this is through counterexamples. For example, let $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $S = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$, which is a symmetric matrix with all positive real components. Then,

$$f(\vec{x}) = \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + 4xy + 4y^2 - x = (x + 2y)^2 - x$$

$f(\vec{x})$ has no global minimum, since you can keep decreasing the output by picking a really large positive value of x and set $y = -\frac{x}{2}$, which makes

$$f(\vec{x}) = (x + 2 \cdot -\frac{x}{2})^2 - x = 0 - x = -x$$

- d) (6 pts) We'd like to use gradient descent to minimize $f(\vec{x})$. Suppose $S = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$, and we use a learning rate of $\alpha = 1$. After one iteration of gradient descent, we have $\vec{x}^{(1)} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$. What was our initial guess, $\vec{x}^{(0)}$? Show your work, and circle your final answer, which should be a vector with two entries and no variables.

Solution: The gradient update rule is $\vec{x}^{(t+1)} = \vec{x}^{(t)} - \alpha \nabla f(\vec{x}^{(t)})$. Plugging in $\alpha = 1$ and $t = 0$ simplifies our problem to

$$\begin{aligned}\vec{x}^{(1)} &= \vec{x}^{(0)} - \alpha \nabla f(\vec{x}^{(0)}) \\ &= \vec{x}^{(0)} - (S\vec{x}^{(0)} - \vec{b}) \\ &= \vec{x}^{(0)} - S\vec{x}^{(0)} + \vec{b}\end{aligned}$$

Now, all we need to do is substitute our known vector $\vec{x}^{(1)} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$ and matrix S into the above equation and solve for $\vec{x}^{(0)}$.

$$\begin{aligned}\begin{bmatrix} -2 \\ -4 \end{bmatrix} &= \vec{x}^{(0)} - \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \vec{x}^{(0)} + \begin{bmatrix} 1 \\ -4 \end{bmatrix} \\ \begin{bmatrix} -3 \\ 0 \end{bmatrix} &= \vec{x}^{(0)} - \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \vec{x}^{(0)} \\ \begin{bmatrix} -3 \\ 0 \end{bmatrix} &= \vec{x}^{(0)} - \begin{bmatrix} 2x_1^{(0)} \\ 6x_2^{(0)} \end{bmatrix} \\ \begin{bmatrix} -3 \\ 0 \end{bmatrix} &= \begin{bmatrix} -x_1^{(0)} \\ -5x_2^{(0)} \end{bmatrix} \\ x_1^{(0)} &= 3, x_2^{(0)} = 0\end{aligned}$$

So, our initial guess was

$$\boxed{\vec{x}^{(0)} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}}$$

Problem 7: Complexity (10 pts)

Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function.

- a) (4 pts) Find scalars a and b such that $f(3) \leq af(2) + bf(6)$. Show your work and circle your final answer, which should be a pair of scalars.

Solution: Recall the definition of convexity (which is relevant, since f is told to us to be convex):

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

Matching the right-side of the inequality above to the right-side of the inequality given, we see that $a = 1 - t$ and $b = t$.

So, our job is to find $1 - t$ and t such that

$$3 = (1-t) \cdot 2 + t \cdot 6$$

i.e. to write 3 as a linear combination of 2 and 6.

$$3 = (1-t) \cdot 2 + t \cdot 6 = 2 - 2t + 6t = 2 + 4t \implies t = \frac{3-2}{4} = \frac{1}{4}$$

So, $a = \frac{3}{4}, b = \frac{1}{4}$.

- b) (6 pts) Using the result from part a), prove that $f(3) + f(5) \leq f(2) + f(6)$.

Solution: In part a), we proved

$$f(3) \leq \frac{3}{4}f(2) + \frac{1}{4}f(6)$$

Since there's an $f(5)$ in the left side of expression we want to prove, we need to find an inequality for $f(5)$ in terms of $f(2)$ and $f(6)$.

Trying to match the pattern, let $t = \frac{3}{4}$, and keep $x = 2$ and $y = 6$. Where did $t = \frac{3}{4}$ come from? You could have found it from solving $(1 - t) \cdot 2 + t \cdot 6 = 5$, or by guessing/observing that no other value of t would eventually allow us to add the two inequalities together to get $f(2) + f(6)$ on the right.

$$\begin{aligned} f((1-t)x + ty) &\leq (1-t)f(x) + tf(y) \\ f\left(\left(1 - \frac{3}{4}\right) \cdot 2 + \frac{3}{4} \cdot 6\right) &\leq \left(1 - \frac{3}{4}\right)f(2) + \frac{3}{4}f(6) \\ f(5) &\leq \frac{1}{4}f(2) + \frac{3}{4}f(6) \end{aligned}$$

Let's add this to our previous inequality.

$$\begin{aligned} f(3) + f(5) &\leq \frac{3}{4}f(2) + \frac{1}{4}f(6) + \frac{1}{4}f(2) + \frac{3}{4}f(6) \\ f(3) + f(5) &\leq f(2) + f(6) \end{aligned}$$

as required!

username: _____

(2 pts) Congrats on finishing Midterm 2! Here are two free points.

Feel free to draw us a picture about EECS 245 in the box below.