Question 1

Given a dataset $y_1 \leq y_2 \leq ... \leq y_n$, define the following functions:

$$R_{\text{abs}}(h) = \frac{1}{n}\sum_{i=1}^{n} |y_i - h| \qquad R_{\text{sq}}(h) = \frac{1}{n}\sum_{i=1}^{n} (y_i - h)^2$$

Parts (a), (b), and (c) below concern $R_{\text{abs}}(h)$, while parts (d) and (e) concern $R_{\text{sq}}(h)$.

In each of the options below, you're presented with two expressions. Determine which expression is larger, and determine by how much it is larger. If it's impossible to tell which of the two expressions is larger, write "impossible" in the provided box.

a) Let c be such that $c < c + 1 < y_1$.

Which is larger? ● $R_{\text{abs}}(c)$ ○ $R_{\text{abs}}(c+1)$

By how much? 1

since c and c+1 is less than all points:

$|y_i - c| = (y_i - c)$ $|y_i - (c+1)| = (y_i - (c+1))$

$$R_{abs}(c) - R_{abs}(c+1) = \frac{1}{n}\sum_{i=1}^{n} |y_i - c| - \frac{1}{n}\sum_{i=1}^{n} |y_i - (c+1)|$$
$$= \frac{1}{n}\sum_{i=1}^{n} (y_i - c) - \frac{1}{n}\sum_{i=1}^{n} (y_i - (c+1))$$
$$= \frac{1}{n}\left(\sum_{i=1}^{n} (y_i - c) - \sum_{i=1}^{n} (y_i - (c+1))\right)$$
$$= \frac{1}{n}\left(\sum_{i=1}^{n} (y_i - c) - y_i + (c+1)\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} 1$$
$$= 1$$

b) Let c be such that $y_n < c < c + 2$.

Which is larger? ○ $R_{\text{abs}}(c)$ ● $R_{\text{abs}}(c+2)$

By how much? 2

Similar work as above, except c and c+2 greater than all points so $|y_i - c| = (c - y_i)$ $|y_i - (c+2)| = (c + 2 - y_i)$

c) Suppose $n = 10$. Let c be such that $y_3 < c < c + 1 < y_4$.

Which is larger? ● $R_{\text{abs}}(c)$ ○ $R_{\text{abs}}(c+1)$

By how much? 0.4

For any point between y_3 and y_4

$$R_{abs}(h) = \frac{1}{n}(\# y_i < h - \# y_i > h)$$
$$= \frac{1}{10}(3 - 7)$$
$$= -0.4$$ slope for values b/w y_3 and y_4

$\therefore R_{abs}(c)$ is 0.4 units greater than $R_{abs}(c+1)$

d) Let c be such that $c < y_1$.

Which is larger? ○ $R_{\text{sq}}(c)$ ● $R_{\text{sq}}(c-1)$

By how much? impossible

c and c-1 both less than y_1, but c-1 is further to the left, and therefore has a greater distance from every y_i.
We cannot determine the difference as that depends on the datapoints and value of c

e) Let c be such that $c < y_n$.

Which is larger? ○ $R_{\text{sq}}(c)$ ● $R_{\text{sq}}(c+1)$

By how much? impossible

Similar work as above, but c and c+1 are both greater than y_n, but c+1 is further to the right and therefore has a greater distance from every y_i

Question 2

Suppose you have a dataset of n points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ where the least squares regression line is $H^*(x) = 3x + 7$. Suppose also that $n = 10$, the average x is $\bar{x} = 2$, and the variance of x is $\sigma_x^2 = 7$.

a) Suppose we move just one point up by 35 units: the point $(3, 2)$ becomes $(3, 37)$. Determine the slope and intercept of the new regression line. Show your work, and put a $\boxed{\text{box}}$ around your final answers.

$w_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$ changing (3,2) to (3,37) does not change denominator

$\sum_{i=1}^{n} (x_i - \bar{x})^2 = n\sigma_x^2 = 10 \times 7 = 70$

Original $H^*(x) = 3x + 7$, so original $w_1^* = 3$

One term changes in numerator $(x_i - \bar{x})y_i$. $(3-2)\times 2 \to (3-2)\times 37$, so numerator increases by 35, so slope increases by $\frac{35}{70} = 0.5$, $\therefore$ new $w_1^* = 0.5$

$w_0^* = \bar{y} - w_1^* \cdot \bar{x}$ old $\bar{y} = w_0^* + w_1^* \bar{x} = 7 + 3 \times 2 = 13$

Moving one point up by 35, $13 \times 10 + 35 = 165$, so new $\bar{y} = 16.5$

new $w_0^* = \bar{y} - w_1^* \cdot \bar{x} = 16.5 - 3.5 \times 2 = 9.5$

so the new $\boxed{H^*(x) = 3.5x + 9.5}$

b) Suppose instead we move all points up by 35 units. Determine the slope and intercept of the new regression line. Show your work, and put a $\boxed{\text{box}}$ around your final answers.

Moving all points up by 35 does not change the slope

old $w_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$ new $w_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i + 35)}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$ $\dfrac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i + \sum_{i=1}^{n} (x_i - \bar{x}) \cdot 35}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

$\sum_{i=1}^{n} (x_i - \bar{x}) \cdot 35 = 35 \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \right) = 35(n\bar{x} - n\bar{x}) = 0$ so no increase

Moving all points up by 35 causes the average y to increase by 35, so the intercept increases by 35.

so the new $\boxed{H^*(x) = 3x + 42}$

Question 3

Suppose $M \in \mathbb{R}^{m \times n}$ is a matrix, $\vec{v} \in \mathbb{R}^n$ is a vector, and $s \in \mathbb{R}$ is a scalar.

Determine whether each of the following quantities is a matrix, vector, scalar, or undefined. If the result is a matrix or vector, state its dimensions in the blank provided.

a) $M\vec{v}$

○ matrix of size _______ ● vector of size m×1
○ scalar ○ undefined

M (m×n) $\vec{v}$ (n×1)
inner match, outer result
so we get (m×1) vector

b) vM

○ matrix of size _______ ○ vector of size _______
○ scalar ● undefined

$\vec{v}$ (n×1) M (m×n)
no inner match
so undefined

c) $\vec{v}^2$

○ matrix of size _______ ○ vector of size _______
○ scalar ● undefined

$\vec{v}^2$
v × v
(n×1) (n×1)
no inner match so undefined

d) $M^T M$

● matrix of size n×n ○ vector of size _______
○ scalar ○ undefined

M^T M
(n×m) (m×n)
inner match so get n×n matrix

e) MM^T

● matrix of size m×m ○ vector of size _______
○ scalar ○ undefined

M M^T
(m×n) (n×m)
inner match so get m×m matrix

f) $\vec{v}^T M\vec{v}$

○ matrix of size _______ ○ vector of size _______
○ scalar ● undefined

v^T M v
(1×n) (m×n) (n×1)
no inner match so undefined

g) $(sM\vec{v}) \cdot (sM\vec{v})$

○ matrix of size _______ ○ vector of size _______
● scalar ○ undefined

s M v
1×1 (m×n) (n×1)
so get an m×1 vector
self dot product valid,
so get a scalar

h) $(s\vec{v}^T M^T)^T$

○ matrix of size _______ ● vector of size m×1
○ scalar ○ undefined

s v^T M^T
1×1 (1×n) (n×m)
$(1 \times m)^T$ → (m×1)

i) $\vec{v}^T M^T M\vec{v}$

○ matrix of size _______ ○ vector of size _______
● scalar ○ undefined

v^T M^T M v
(1×n) (n×m) (m×n) (n×1)
(1×m) (m×1) → (1×1)

j) $\vec{v}\vec{v}^T + M^T M$

● matrix of size n×n ○ vector of size _______
○ scalar ○ undefined

v v^T M^T M
(n×1) (1×n) (n×m) (m×n)
(n×n) + (n×n) → (n×n)

Question 4

Consider the vectors $\vec{u}$ and $\vec{v}$, defined below.

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

a) Determine the angle between $\vec{u}$ and $\vec{v}$. Put a $\boxed{\text{box}}$ around your final answer. Leave your answer in the form $\cos^{-1}(\cdot)$.

Recall $u \cdot v = \|u\| \|v\| \cos\theta$

$u \cdot v = (2,-1,-3)\cdot(1,0,4) = 2\times1 + (-1)\times0 + (-3)\times4 = 2+0-12 = -10$

$\|u\| = \sqrt{(2)^2+(-1)^2+(-3)^2} = \sqrt{4+1+9} = \sqrt{14}$

$\|v\| = \sqrt{(1)^2+(0)^2+(4)^2} = \sqrt{1+0+16} = \sqrt{17}$

$\therefore \quad -10 = \sqrt{14}\cdot\sqrt{17}\cdot\cos(\theta)$

$\therefore \quad \theta = \cos^{-1}\left(\frac{-10}{\sqrt{14}\cdot\sqrt{17}}\right)$

b) Determine the projection of $\vec{u}$ onto span$(\vec{v})$. Give your answer in the form of a **vector**. Put a $\boxed{\text{box}}$ around your final answer.

To project $\vec{u}$ onto span $(\vec{v})$ we use the formula : $\frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}} \cdot \vec{v}$

$\vec{u}\cdot\vec{v} = (2,-1,-3)\cdot(1,0,4) = 2\times1 + (-1)\times0 + (-3)\times4 = 2+0-12 = -10$

$\vec{v}\cdot\vec{v} = (1,0,4)\cdot(1,0,4) = 1^2+0^2+4^2 = 1+0+16 = 17$

$\frac{-10}{17}\cdot\vec{v} = \frac{-10}{17}(1,0,4) = \begin{bmatrix} -\frac{10}{17} \\ 0 \\ -\frac{40}{17} \end{bmatrix}$

Question 5

You have data on three different apartments for rent. Each apartment's data is given as a tuple $(x^{(1)}, x^{(2)}, x^{(3)}, y)$, where:

- $x^{(1)}$ is the number of bedrooms,
- $x^{(2)}$ is the floor number,
- $x^{(3)}$ is the proportion of positive Yelp reviews for the apartment complex, and
- y is the monthly rent.

Your data is:

$$(2, 9, 0.5, 500), (3, 8, 0.7, 700), (4, 10, 0.8, 900)$$

a) You want to predict an apartment's rent using a hypothesis function of the form:

$$H(x^{(1)}, x^{(2)}, x^{(3)}) = w_0 + w_1 \cdot \sqrt{x^{(1)}} + w_2 \cdot \left(x^{(2)}\right)^2 + w_3 \cdot x^{(3)}$$

Write down the design matrix, X, and the observation vector, $\vec{y}$. State how you would use these quantities to find the hypothesis function, but don't actually do any calculations.

$$X = \begin{bmatrix} 1 & \uparrow & \uparrow & \uparrow \\ \vdots & \sqrt{x^{(1)}} & (x^{(2)})^2 & x^{(3)} \\ 1 & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & \sqrt{2} & 81 & 0.5 \\ 1 & \sqrt{3} & 64 & 0.7 \\ 1 & 2 & 100 & 0.8 \end{bmatrix}$$

The features have to be transformed according to the hypothesis function

$$\vec{y} = \begin{bmatrix} \uparrow \\ \text{observations} \\ \downarrow \end{bmatrix} = \begin{bmatrix} 500 \\ 700 \\ 900 \end{bmatrix}$$

To use X and $\vec{y}$ to find the prediction rule, we need to solve the system of normal equations $X^TXw = X^Ty$ for the parameter vector $\vec{w}$, which would give us the constants $\vec{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix}$

b) Suppose you'd like to instead fit the hypothesis function:

$$H(x^{(1)}, x^{(2)}, x^{(3)}) = w_1 \cdot \sqrt{x^{(1)}} + w_2 \cdot \left(x^{(2)}\right)^2 + w_3 \cdot x^{(3)}$$

Will your answer to part (a) change? If so, how?

The answer to part (a) would change, we would remove the column of 1s

$$X = \begin{bmatrix} \sqrt{2} & 81 & 0.5 \\ \sqrt{3} & 64 & 0.7 \\ 2 & 100 & 0.8 \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 500 \\ 700 \\ 900 \end{bmatrix}$$