Machine Learning

Homework 01 Solution

Min-Kuan Chang

minkuanc@nchu.edu.tw

EE, College of EECS

Problem 01

- Suppose X be a zero-mean Gaussian random variable with variance σ_X^2. Let $Y = X^2$
  1. What is the pdf of Y?

     $$\Pr\{Y \le y\} = \Pr\{X^2 \le y\} = \Pr\{-\sqrt{y} \le X \le \sqrt{y}\} = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx$$

     The pdf of Y is

     $$f_Y(y) = \frac{\partial \Pr\{Y \le y\}}{\partial y} = f_X(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) = \frac{1}{\sqrt{2\pi\sigma_X^2 y}}\exp\{-y/\sigma_X^2\}$$

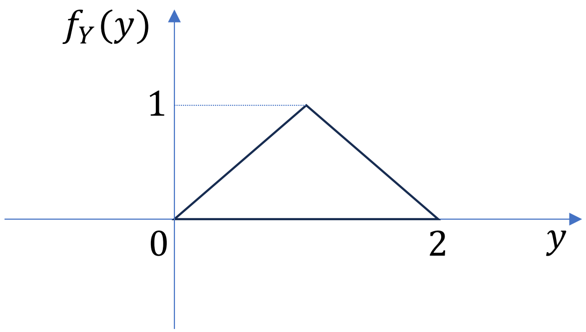
  2. Please calculate $E\{Y\}$ and $E\{Y^2\}$

     $E\{Y\} = \sigma_X^2$

     $E\{Y^2\} = 3\sigma_X^4$

Problem 02

- Suppose X_1 and X_2 be uniform random variables over $(0,1)$ and X_1 and X_2 are independent
- Let $Y = X_1 + X_2$
  1. what is the pdf of Y?
  2. Please calculate $E\{Y\}$ and $E\{Y^2\}$



$E\{Y\} = 1$

$E\{Y^2\} = \frac{1}{3} + 2 * \frac{1}{2} * \frac{1}{2} + \frac{1}{3} = \frac{14}{12} = \frac{7}{6}$

Problem 03

- Suppose $\boldsymbol{y} = [y_1, y_2, y_3]^T$ and $\boldsymbol{x} = [x_1, x_2]^T$ in which $y_1 = x_1^2 - 2x_2^2$, $y_2 = -3x_1^2 + 2x_1x_2 + x_2^2$ and $y_3 = x_1^2 + x_2^2$
  1. Please find $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$

$$\begin{bmatrix} 2x_1 & -6x_1 + 2x_2 & 2x_1 \\ 4x_2 & 2x_1 + 2x_2 & 2x_2 \end{bmatrix}$$

Problem 04

- Let $\mathrm{A} = \begin{bmatrix} -1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 5 \end{bmatrix}$ and $\boldsymbol{x} = [x_1, x_2, x_3]^T$

1. Let $\boldsymbol{y} = \mathrm{A}\boldsymbol{x}$. Please find $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$

$$\mathrm{A} = \begin{bmatrix} -1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 5 \end{bmatrix}$$

2. Let $\boldsymbol{y} = \boldsymbol{x}^T \mathrm{A}\boldsymbol{x}$. Please find $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$

$$2\mathrm{A}\boldsymbol{x}$$

Problem 05

- Suppose we have a dataset

x	(1,2)	(3,2)	(0,3)	(4,1)	(2,0)	(2,3)
label	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2

- Now we have a new data x_{new} =(2,2)
- Use KNN with $k = 3$ to classify this new data
  1. Minkowski distance with $p = 1$
  2. Minkowski distance with $p = 2$
  3. Minkowski distance with $p = \infty$

Problem 05 - Solution

	(1,2)	(3,2)	(0,3)	(4,1)	(2,0)	(2,3)	
x_{new} =(2,2)	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	
$p = 1$	1	1	3	3	2	1	Class 2
$p = 2$	1	1	$\sqrt{5}$	$\sqrt{5}$	2	1	Class 2
$p = \infty$	1	1	2	2	2	1	Class 1

Problem 06

ID	Study Hours/Day	Eat Lunch?	Money on pocket	Age	Have Afternoon Tea?
1	3	No	50	18	No
2	8	Yes	100	19	Yes
3	4	No	130	20	Yes
4	10	No	120	22	No
5	9	Yes	40	18	Yes

- Construct a decision tree with depth 3 to predict whether a person will have afternoon tea
  1. Use information gain
  2. Use gini index
  3. Use information gain ratio

Problem 06 - Solution

- Information gain
  - the first level:
    - study hour/day
      - less than 3

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{1}{5}(0)+\frac{4}{5}\left(-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4}\right)\right)=0.3219$$

      - less than 5

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)\right)=0.02$$

      - less than 9

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)\right)=0.02$$

Problem 06 - Solution

- Information gain
  - the first level:
    - study hour/day
      - less than 10

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{1}{5}(0)+\frac{4}{5}\left(-\frac{2}{4}\log_2\frac{2}{4}-\frac{2}{4}\log_2\frac{2}{4}\right)\right)=0.1710$$

    - eat Lunch?

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}(0)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)\right)=0.42$$

Problem 06 - Solution

- Information gain
  - the first level:
    - money on pocket
      - less than 50?

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{1}{5}\,(0)+\frac{4}{5}\,(1)\right)=0.171$$

      - less than 100?

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}\,(1)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{1}{3}\right)\right)=0.02$$

      - less than 120?

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}\,(1)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{1}{3}\right)\right)=0.02$$

Problem 06 - Solution

- Information gain
  - the first level:
    - money on pocket
      - less than 130?

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{1}{5}(0)+\frac{4}{5}(1)\right)=0.171$$

Problem 06 - Solution

- Information gain
  - the first level:
    - age
      - less than 19

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}(1)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{1}{3}\right)\right)=0.02$$

      - less than 20

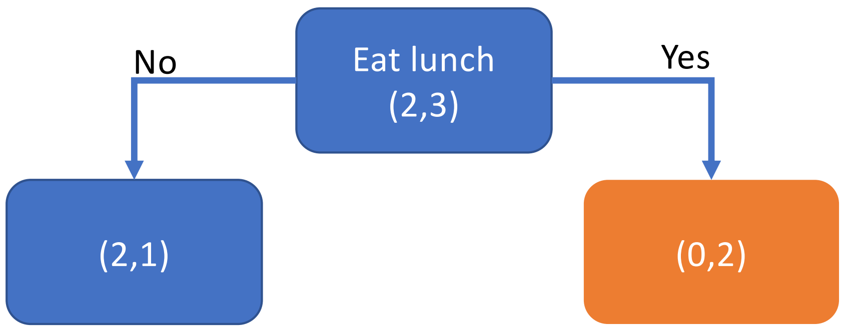
$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{2}{5}(1)+\frac{3}{5}\left(-\frac{2}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{1}{3}\right)\right)=0.02$$

      - less than 22

$$\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)-\left(\frac{1}{5}(0)+\frac{4}{5}\left(-\frac{3}{4}\log_2\frac{1}{4}-\frac{1}{4}\log_2\frac{1}{4}\right)\right)=0.3219$$

Problem 06 - Solution

- Information gain
  - the first level:
    - we can split on eat lunch



Problem 06 - Solution

- Information gain
  - the second level (the left child):

ID	Study Hours/Day	Eat Lunch?	Money on pocket	Age	Have Afternoon Tea?
1	3	No	50	18	No
3	4	No	150	20	Yes
4	10	No	120	22	No

Problem 06 - Solution

- Information gain
  - the second level (the left child):
    - study hour/day
      - less than 4

$$\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)-\left(\frac{1}{3}\,(0)+\frac{2}{3}\,(1)\right)=0.2516$$

      - less than 4

$$\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)-\left(\frac{1}{3}\,(0)+\frac{2}{3}\,(1)\right)=0.2516$$

Problem 06 - Solution

- Information gain
  - the second level (the left child):
    - money on pocket
      - less than 120

$$\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)-\left(\frac{1}{3}(0)+\frac{2}{3}(1)\right)=0.2516$$

      - less than 130

$$\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)-\left(\frac{1}{3}(0)+\frac{2}{3}(0)\right)=0.9183$$

Problem 06 - Solution

- Information gain
  - the second level (the left child):
    - age
      - less than 20

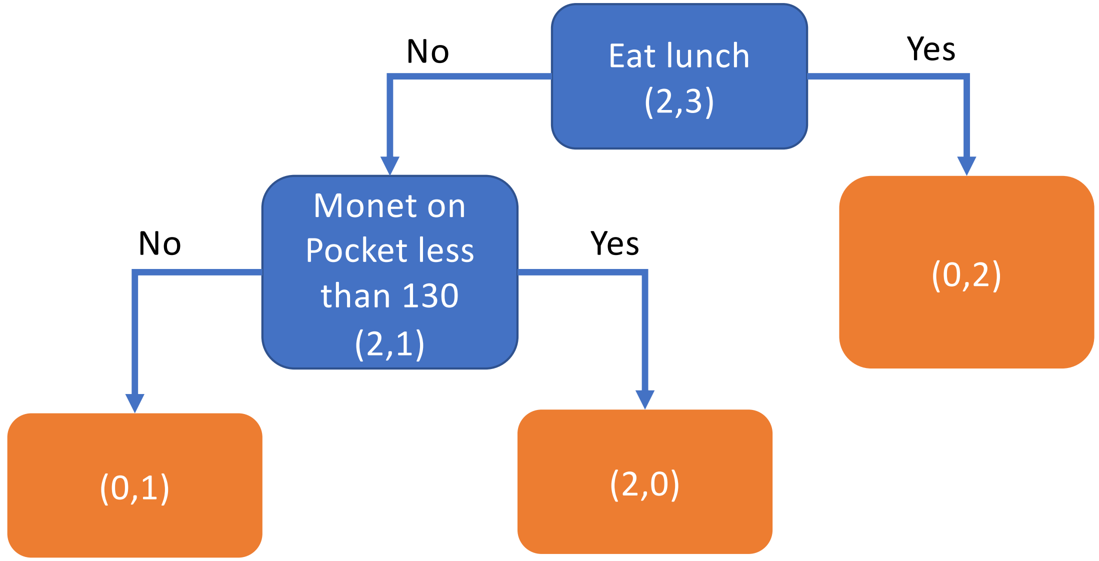
$$\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)-\left(\frac{1}{3}\,(0)+\frac{2}{3}\,(1)\right)=0.2516$$

      - less than 22

$$\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)-\left(\frac{1}{3}\,(0)+\frac{2}{3}\,(1)\right)=0.2516$$

Problem 06 - Solution

- Information gain
  - the second level (the left child):
    - we split on money on pocket less than 150



Problem 07 (Linear regression)

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

- Determine the best predicted model when $\hat{y}(x, w_0, w_1) = w_0 + w_1 x$

Problem 07 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 01: Forming Φ, namely $\boldsymbol{X}$ in the simplest linear Regression Model

$$\Phi = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{bmatrix} = \boldsymbol{X}$$

Problem 07 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 02: Forming $\boldsymbol{y}$

$$\boldsymbol{y} = \begin{bmatrix} 2 \\ 5 \\ 4 \\ 4 \\ 8 \end{bmatrix}$$

Problem 07 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$, namely $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ in the simplest linear regression model

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 25 \\ 25 & 165 \end{bmatrix} \qquad \boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{bmatrix}^T \begin{bmatrix} 2 \\ 5 \\ 4 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 23 \\ 137 \end{bmatrix}$$

Problem 07 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$, namely $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ in the simplest linear regression model

$$\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 5 & 25 \\ 25 & 165 \end{bmatrix}^{-1} \begin{bmatrix} 23 \\ 137 \end{bmatrix} = \begin{bmatrix} 1.85 \\ 0.55 \end{bmatrix}$$

$$\Downarrow$$

$$\hat{y}(x, w_0, w_1) = 1.85 + 0.55x$$

Problem 08 (Linear regression)

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

- Determine the best predicted model when

$$\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$$

Problem 08 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 01: Forming Φ, namely $\boldsymbol{X}$ in the simplest linear Regression Model

$$\Phi = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 7 & 49 \\ 1 & 9 & 81 \end{bmatrix} = \boldsymbol{X}$$

Problem 08 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 02: Forming $\boldsymbol{y}$

$$\boldsymbol{y} = \begin{bmatrix} 2 \\ 5 \\ 4 \\ 4 \\ 8 \end{bmatrix}$$

Problem 08 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	3	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$, namely $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ in the simplest linear regression model

$$\Phi^T\Phi = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 7 & 49 \\ 1 & 9 & 81 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 7 & 49 \\ 1 & 9 & 81 \end{bmatrix} = \begin{bmatrix} 5 & 25 & 165 \\ 25 & 165 & 1225 \\ 165 & 1225 & 9669 \end{bmatrix}$$

Problem 08 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$, namely $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ in the simplest linear regression model

$$\Phi^T\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 7 & 49 \\ 1 & 9 & 81 \end{bmatrix}^T \begin{bmatrix} 2 \\ 5 \\ 4 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 23 \\ 137 \\ 991 \end{bmatrix}$$

Problem 08 (Linear regression) - Solution

- Suppose we have the following dataset

x	1	3	5	7	9
y	2	5	3	4	8

The simplest linear regression model suggests $\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$, namely $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ in the simplest linear regression model

$$\boldsymbol{w}_{\text{opt}} = (\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y} = \begin{bmatrix} 5 & 25 & 165 \\ 25 & 165 & 1225 \\ 165 & 1225 & 9669 \end{bmatrix}^{-1} \begin{bmatrix} 23 \\ 137 \\ 991 \end{bmatrix} = \begin{bmatrix} 2.7607 \\ 0.0143 \\ 0.0536 \end{bmatrix}$$

$$\Downarrow$$

$$\hat{y}(x, w_0, w_1) = 2.7607 + 0.0143x + 0.0536x^2$$

Problem 09 (Linear regression)

- Suppose we have the following dataset

$x = (x_1, x_2)$	(1,1)	(1,3)	(5,2)	(7,4)	(9,2)
y	2	5	4	4	8

- Determine the best predicted model when

$$\hat{y}(x, w_0, w_1) = w_0 + w_1 x_1 + w_2 x_2$$

Problem 09 (Linear regression) - Solution

- Suppose we have the following dataset

$x = (x_1, x_2)$	(1,1)	(1,3)	(5,2)	(7,4)	(9,2)
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(\boldsymbol{x}, w_0, w_1) = w_0 + w_1 x_1 + w_2 x_2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$

Step 01: Forming Φ, namely $\boldsymbol{X}$ in the simplest linear Regression Model

$$\Phi = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 5 & 2 \\ 1 & 7 & 4 \\ 1 & 9 & 2 \end{bmatrix} = \boldsymbol{X}$$

Problem 09 (Linear regression) - Solution

- Suppose we have the following dataset

$x = (x_1, x_2)$	(1,1)	(1,3)	(5,2)	(7,4)	(9,2)
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(\boldsymbol{x}, w_0, w_1) = w_0 + w_1x_1 + w_2x_2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 02: Forming $\boldsymbol{y}$

$$\boldsymbol{y} = \begin{bmatrix} 2 \\ 5 \\ 4 \\ 4 \\ 8 \end{bmatrix}$$

Problem 09 (Linear regression) - Solution

- Suppose we have the following dataset

$x = (x_1, x_2)$	(1,1)	(1,3)	(5,2)	(7,4)	(9,2)
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(\boldsymbol{x}, w_0, w_1) = w_0 + w_1 x_1 + w_2 x_2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T\Phi)^{-1}\Phi^T\boldsymbol{y}$, namely $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ in the simplest linear regression model

$$\Phi^T\Phi = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 5 & 2 \\ 1 & 7 & 4 \\ 1 & 9 & 2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 5 & 2 \\ 1 & 7 & 4 \\ 1 & 9 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 23 & 12 \\ 23 & 157 & 60 \\ 12 & 60 & 34 \end{bmatrix}$$

Problem 09(Linear regression) - Solution

- Suppose we have the following dataset

$x = (x_1, x_2)$	(1,1)	(1,3)	(5,2)	(7,4)	(9,2)
y	2	5	4	4	8

The simplest linear regression model suggests $\hat{y}(\boldsymbol{x}, w_0, w_1) = w_0 + w_1 x_1 + w_2 x_2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T \Phi)^{-1} \Phi^T \boldsymbol{y}$, namely $(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$ in the simplest linear regression model

$$\Phi^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 5 & 2 \\ 1 & 7 & 4 \\ 1 & 9 & 2 \end{bmatrix}^T \begin{bmatrix} 2 \\ 5 \\ 4 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 23 \\ 127 \\ 57 \end{bmatrix}$$

Problem 09 (Linear regression) - Solution

- Suppose we have the following dataset

$x = (x_1, x_2)$	(1,1)	(1,3)	(5,2)	(7,4)	(9,2)
y	2	5	3	4	8

The simplest linear regression model suggests $\hat{y}(\boldsymbol{x}, w_0, w_1) = w_0 + w_1 x_1 + w_2 x_2$

Recall $\boldsymbol{w}_{\text{opt}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$

Step 03: Calculate $\Phi(\Phi^T \Phi)^{-1} \Phi^T \boldsymbol{y}$, namely $(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$ in the simplest linear regression model

$$\boldsymbol{w}_{\text{opt}} = (\Phi^T \Phi)^{-1} \Phi^T \boldsymbol{y} = \begin{bmatrix} 5 & 23 & 12 \\ 23 & 157 & 60 \\ 12 & 60 & 34 \end{bmatrix}^{-1} \begin{bmatrix} 23 \\ 127 \\ 57 \end{bmatrix} = \begin{bmatrix} 2.7730 \\ 0.4178 \\ -0.0395 \end{bmatrix}$$

$$\hat{y}(x, w_0, w_1) = 2.7730 + 0.4178 x_1 - 0.0395 x_2$$

Problem 10 (Linear regression)

- This is a programming problem
- Suppose $y = 3 + 2x + 0.2x^2 + n$ in which n is a zero-mean Gaussian noise with variance σ_N^2
- Suppose the predicted model is $\hat{y}(x, w_0, w_1) = w_0 + w_1 x + w_2 x^2$
- Determine the optimal weights when
  1. $\sigma_N = 0.01$
  2. $\sigma_N = 0.1$
  3. $\sigma_N = 1$
  4. $\sigma_N = 2$

Problem 10 (Linear regression)

- Procedure to generate your own dataset
  - step 1: uniformly generate a random variable between 0 and 8 and let this be x
  - step 2: generate a zero-mean Gaussian noise with variance σ_N^2 and let this be n
  - step 3: let $y = 3 + 2x + 0.2x^2 + n$
  - step 4: add (x, y) to your own dataset
  - use this procedure to have a dataset containing 40 pairs of x and y

Note that
1. Please plot the regression function along with the data points on the same graph
2. You should also submit your code