Task 1

Yilong Wang 483728
Linda Li 458029

1.
$$W = \frac{\sum_{i=1}^{n=3} x_i y_i}{\sum_{i=1}^{n=3} x_i x_i} = \frac{x_1y_1 + x_2y_2 + x_3y_3}{x_1x_1 + x_2x_2 + x_3x_3} = \frac{0+1+0}{0+1+4} = \frac{1}{5}$$

2.
$$W = (XX^T)^{-1} X y^T = \left(\begin{bmatrix} x_1 & x_2 & x_3 \\ (x_1)^2 & (x_2)^2 & (x_3)^2 \end{bmatrix} \begin{bmatrix} x_1 & (x_1)^2 \\ x_2 & (x_2)^2 \\ x_3 & (x_3)^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} x_1 & x_2 & x_3 \\ (x_1)^2 & (x_2)^2 & (x_3)^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 2 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 5 & 9 \\ 9 & 17 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{17}{4} & \frac{-9}{4} \\ \frac{-9}{4} & \frac{5}{4} \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 2 & -\frac{1}{2} \\ 0 & -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

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

$$g(x) = W^T \phi(x) = [2, -1] \begin{bmatrix} x \\ x^2 \end{bmatrix} = 2x - x^2$$

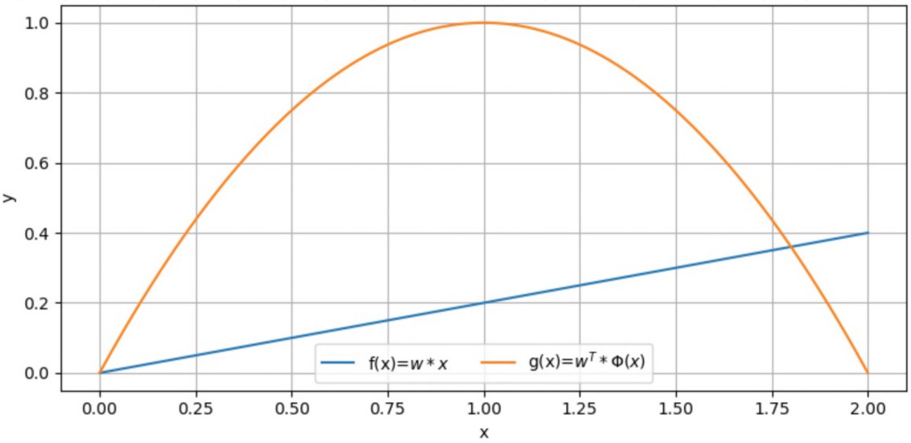
3.

```python
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace(0, 2, 400)

def g_x(x):
    return 2*x-x**2
def f_x(x):
    return 0.2*x


plt.figure(figsize=(8, 4))
plt.xlabel("x")
plt.ylabel('y')
#plt.scatter(datapoints, f_x(datapoints), label='Data Points')
plt.plot(x, f_x(x),'-', label='f(x)=$w*x$')
plt.plot(x, g_x(x),"-", label='g(x)=$w^T*\Phi(x)$')
plt.legend(fontsize=10, ncol=2)
plt.grid(True)
plt.tight_layout ()
plt.show()
```

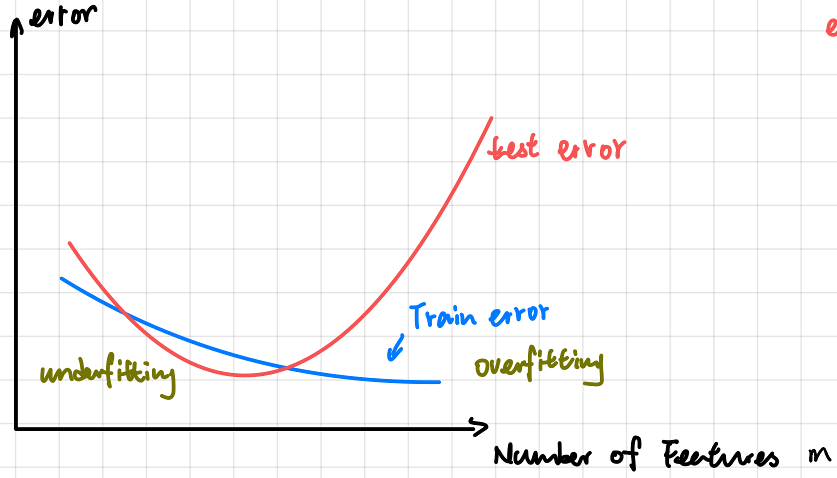


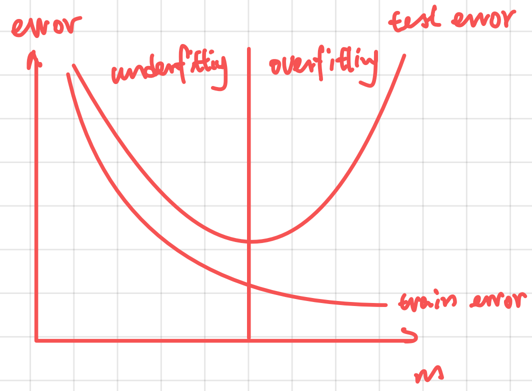
Task 2

1. a 2. b 3. a 4. c

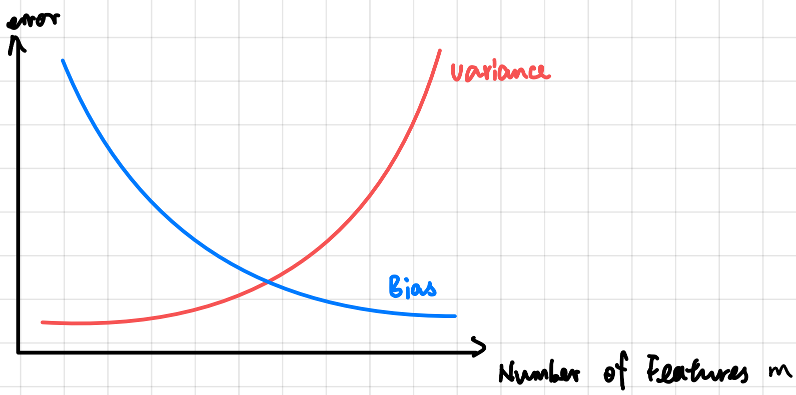
Task 3

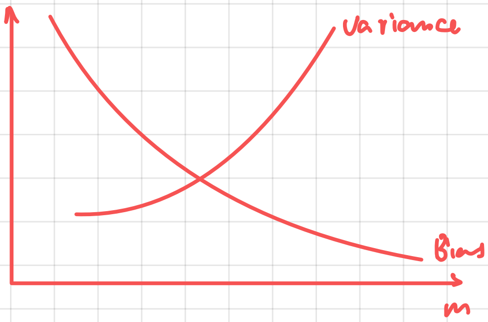
1.2.





3.





4. a) the train error will increase.

Because $E_{RR} = \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 + \lambda \sum_{i=1}^{m} w_i^2$ is greater than $\mathcal{E} = \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2$,

which means $\lambda \sum_{i=1}^{m} w_i^2$ constrains the flexibility of the model

b) the test error will decrease

Because regularization will help to prevent overfitting durch $\lambda \sum_{i=1}^{m} w_i^2$

c) the bias of $\hat{f}$ will increase

d) the variance of $\hat{f}$ will decrease

Task 4

1. $X' = AX$

$$\hat{w}_{OLS} = (XX^T)^{-1} X y^T$$

$$\hat{w}'_{OLS} = (X'X'^T)^{-1} X' y^T = (AXX^T A^T)^{-1} A X y^T$$

$$= (A^T)^{-1} (XX^T)^{-1} A^{-1} A X y^T \qquad \Big| \text{ A ist invertible}$$

$$= (A^T)^{-1} (XX^T)^{-1} X y^T = (A^T)^{-1} \hat{w}_{OLS}$$

$$\hat{Y}_{OLS} = \hat{w}^T_{OLS} X = ((XX^T)^{-1} X y^T)^T X = y X^T (XX^T)^{-1} X$$

$$\hat{Y}'_{OLS} = \hat{w}'^T_{OLS} X' = [(A^T)^{-1} \cdot (XX^T)^{-1} X y^T]^T A X$$

$$= y X^T (XX^T)^{-1} ((A^T)^{-1})^T \cdot AX$$

$$= y X^T \cdot (XX^T)^{-1} \cdot ((A^{-1})^T)^T \cdot AX$$

$$= y X^T \cdot (XX^T)^{-1} \cdot A^{-1} A X$$

$$= y X^T \cdot (XX^T)^{-1} \cdot X = \hat{Y}_{OLS}$$

2. A ist orthogonal Matrix $\Rightarrow A^T = A^{-1}$

$$\hat{y}_{RR} = \hat{w}^T_{RR} X = [(XX^T + \lambda I)^{-1} X y^T]^T X = y X^T ((XX^T + \lambda I)^{-1})^T X$$

$$\hat{y}'_{RR} = y (AX)^T ((AXX^T A^T + \lambda I)^{-1})^T AX$$

$$= y (AX)^T ((AXX^T A^T + \lambda \cdot A \cdot A^T)^{-1})^T AX$$

$$= y X^T A^T \cdot [(A (XX^T + \lambda I) A^T)^{-1}]^T AX$$

$$= y X^T A^T [(A^T)^{-1} (XX^T + \lambda I)^{-1} \cdot A^{-1}]^T AX$$

$$= y X^T A^T \cdot (A^{-1})^T [(XX^T + \lambda I)^{-1}]^T ((A^T)^{-1})^T AX$$

$$= y X^T A^T \cdot A [(XX^T + \lambda I)^{-1}]^T \cdot A^{-1} A X$$

$$= y X^T [(XX^T + \lambda I)^{-1}]^T X = \hat{Y}_{RR}$$

Task 5

1. reject

This paper does model evaluation and model selection at the same time.

They could be too optimistic

2. for exsample, we could regard x_1, x_2, x_3, x_4, x_5 as training set, and x_6 as test set

i.e. label: $+ \quad - \quad - \quad + \quad + \quad -$

$$\underbrace{x_1, x_2, x_3, x_4, x_5}_{\text{train set}}, x_6$$

predict: + , aginst -

the numbers of positive and negative exsamples are same size. In this case the train set has more exsamples with + label, therefore it predict x_6 would be + , meanwhile the test set has more exsamples with - label because positive and negative exsamples both have 100 exsamples.