1. (a)

$$f(w) = w^T Xb = w^T \begin{bmatrix} x_{11}b_1 + \cdots + x_{1p}b_p \\ \vdots \\ x_{n1}b_1 + \cdots + x_{np}b_p \end{bmatrix} = \sum_{j=1}^{n} w_j \sum_{i=1}^{p} (x_{j1}b_1 + \cdots + x_{ip}b_p)$$

$$\nabla f(w) = \begin{bmatrix} x_{11}b_1 + x_{12}b_2 + \cdots + x_{1p}b_p \\ x_{21}b_1 + x_{22}b_2 + \cdots + x_{2p}b_p \\ \vdots \\ x_{n1}b_1 + x_{n2}b_2 + \cdots + x_{np}b_p \end{bmatrix} = Xb$$

(b)

$$\begin{aligned} tr(Bww^T A) &= tr(ABww^T) \\ &=< (AB)^T, ww^T >_F \\ &= AB_{11}w_1^2 + \cdots + AB_{nn}w_n^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}((AB)^T) \cdot (ww^T))_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(AB)_{ij}^T \times w_i w_j \end{aligned}$$

$$\nabla f(w) = \begin{bmatrix} 2w_1(AB)_{11}^T + \sum_{i=2}^{n}((AB)_{1i}^T + AB_{i1}^T)w_i \\ w_1((AB)_{12}^T + (AB)_{21}^T) + 2w_2(AB)_{22}^T + \sum_{i=3}^{n}((AB)_{2i}^T + AB_{i2}^T)w_i \\ \vdots \\ 2w_n(AB)_{nn}^T + \sum_{i=1}^{n-1}((AB)_{ni}^T + (AB)_{in}^T)w_i \end{bmatrix}$$

(c)

$$H = \begin{bmatrix} 2AB_{11}^T & AB_{12}^T + AB_{21}^T & \dots & AB_{1n}^T + AB_{n1}^T \\ AB_{12}^T + AB_{21}^T & 2AB_{22}^T & \dots & AB_{2n}^T + AB_{n2}^T \\ AB_{n1}^T + AB_{1n}^T & AB_{2n}^T + AB_{n2}^T & & 2AB_{nn}^T \end{bmatrix}$$

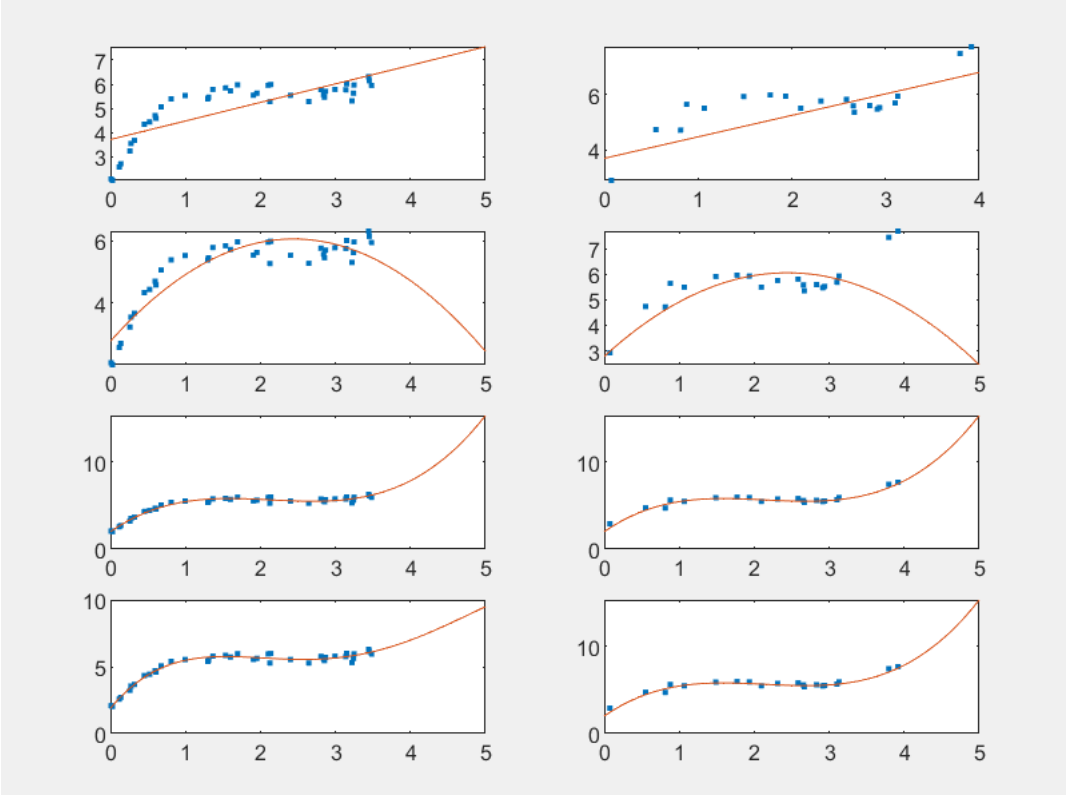
(d) With A = $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and B = $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, H = $\begin{bmatrix} 2 & 4 \\ 4 & -6 \end{bmatrix}$ which is not a positive definite matrix and therefore, f(x) is not a convex function

(e)

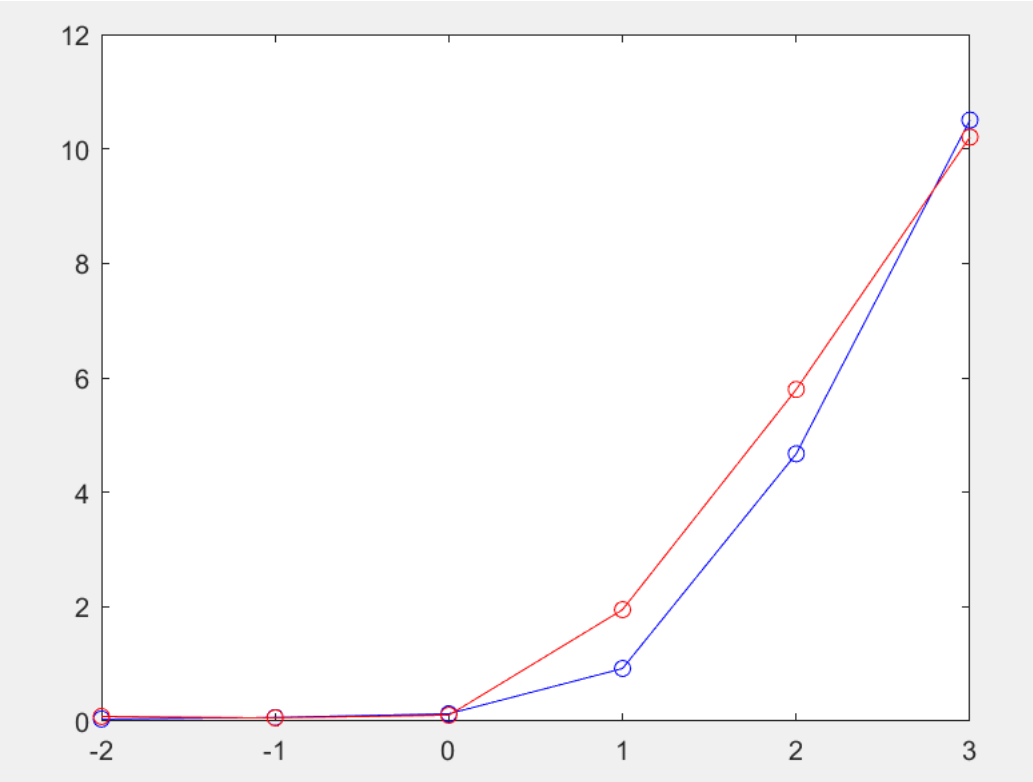
$$\begin{aligned} \log \sigma(w^T x) &= \log \frac{1}{1 + e^{-(\sum_{i=1}^{n} w_i x_i)}} \\ &= -\log 1 + e^{-(\sum_{i=1}^{n} w_i x_i)} \\ &= -ln(1 + e^{-(\sum_{i=1}^{n} w_i x_i)}) \\ &= -ln(1 + e^{-w^T x}) \end{aligned} \tag{1}$$

$$\nabla f(w) = \begin{bmatrix} \frac{x_1}{1+e^{-w^T x}} \\ \vdots \\ \frac{x_n}{1+e^{-w^T x}} \end{bmatrix}$$

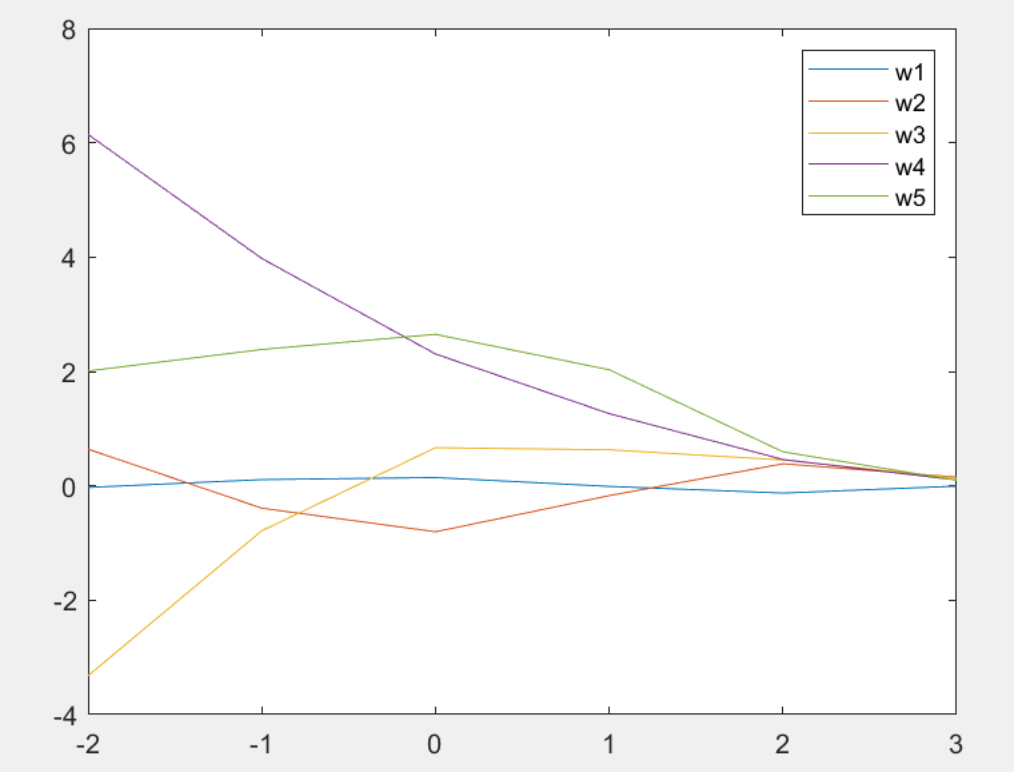
2. The graph shown by the program is shown below
   row1: left: 1st order polynomial regression curve and training data; right: 1st order polynomial regression curve and test data;
   row2: left: 2nd order polynomial regression curve and training data; right: 2nd order polynomial regression curve and test data;
   row3: left: 3rd order polynomial regression curve and training data; right: 3rd order polynomial regression curve and test data;
   row4: left: 4th order polynomial regression curve and training data; right: 4th order polynomial regression curve and test data;



3. (a) $\lambda = 0.1$ is the best for fitting data since here both squared error of training and test data is the lowest
   blue curve: error function on training data; red curve: error function on test data



(b) Graph for different w in b


8
6
4
2
0
-2
-4
-2
-1
0
1
2
3
w1
w2
w3
w4
w5