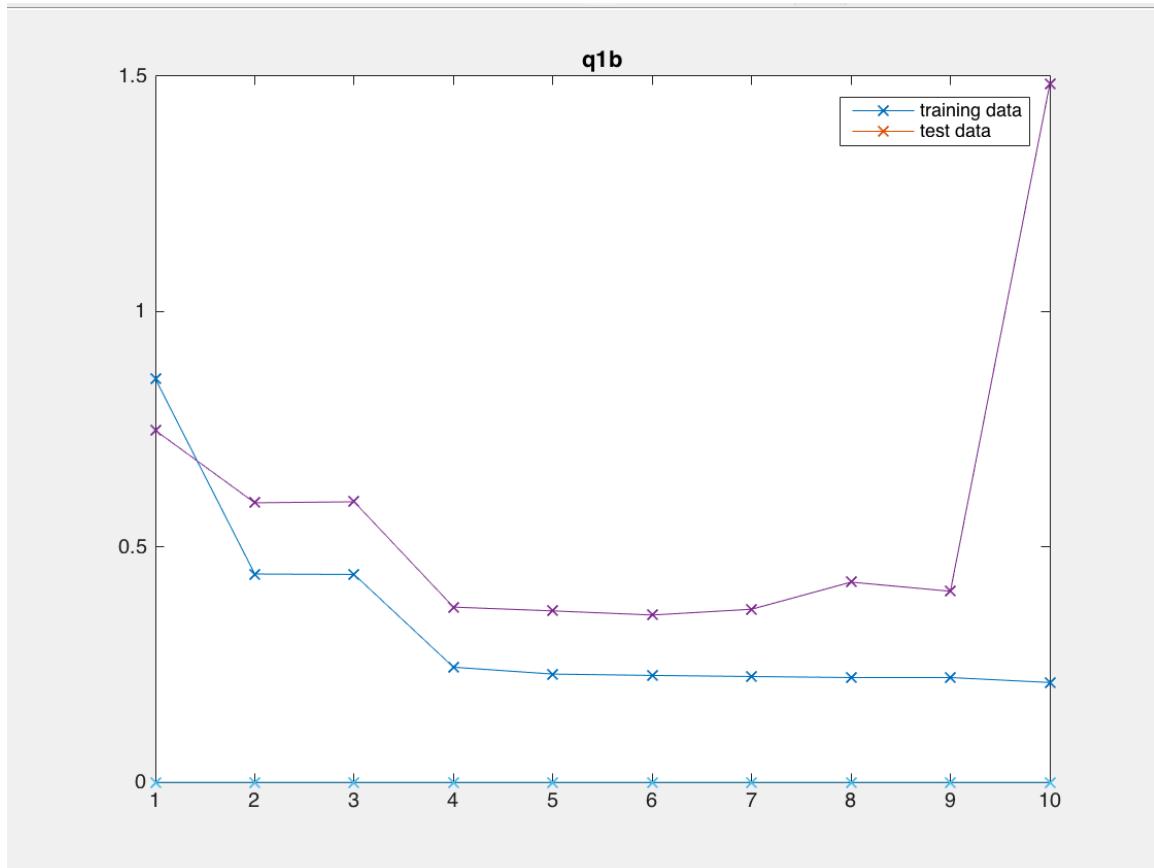


EECS 445 HW1

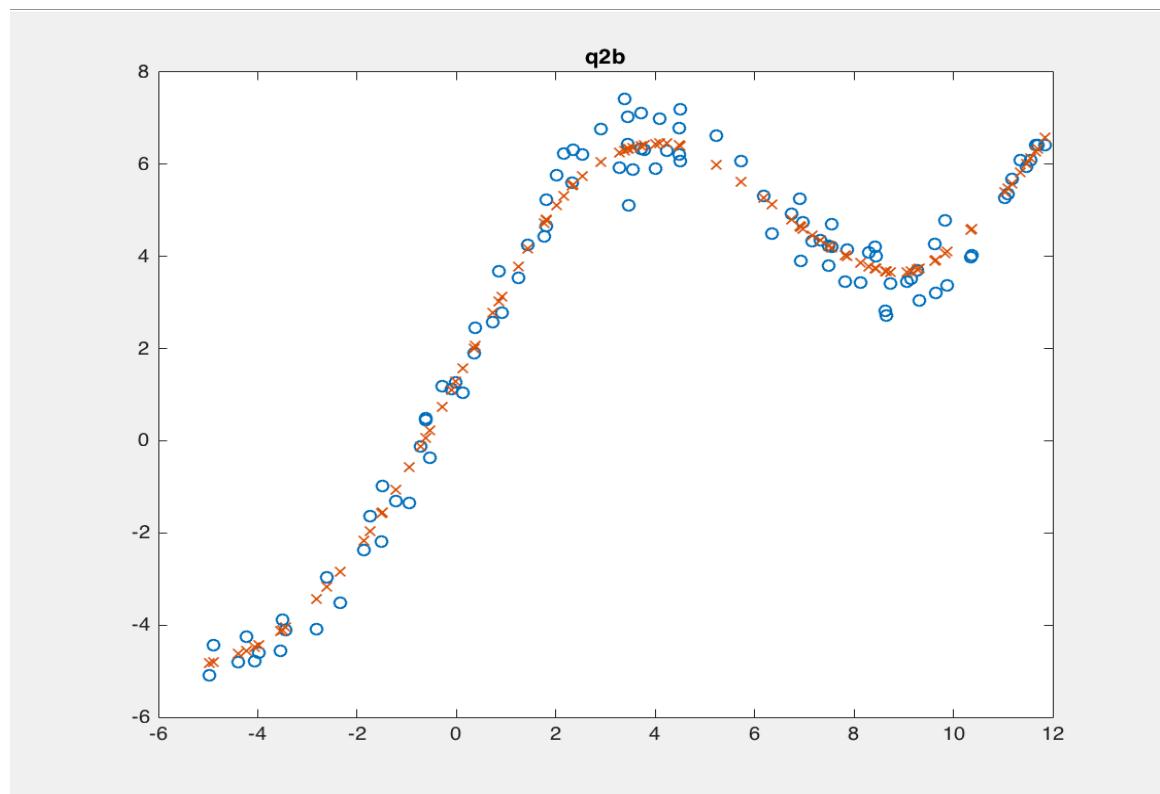
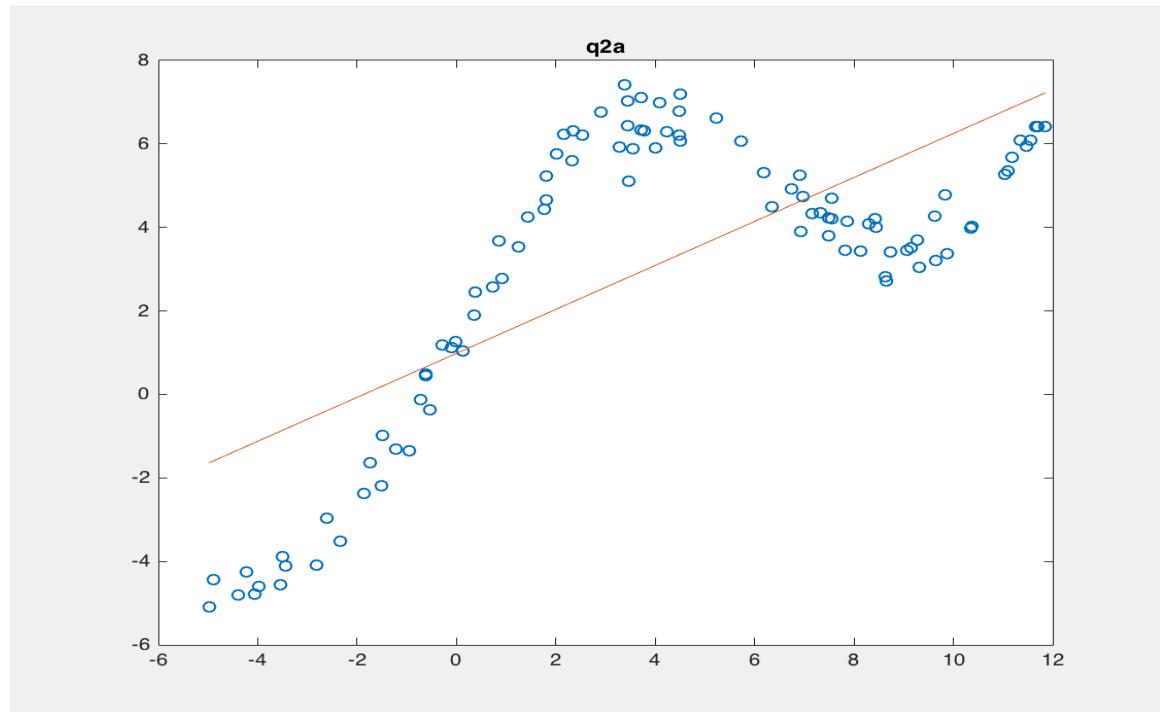
Deepak Kumar – kumarde
Kevin Pitt – kpitt
Stephen Kovacinski - kova

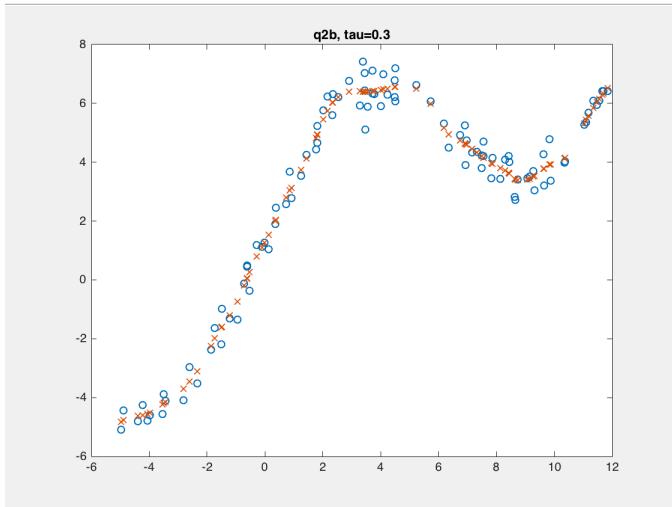
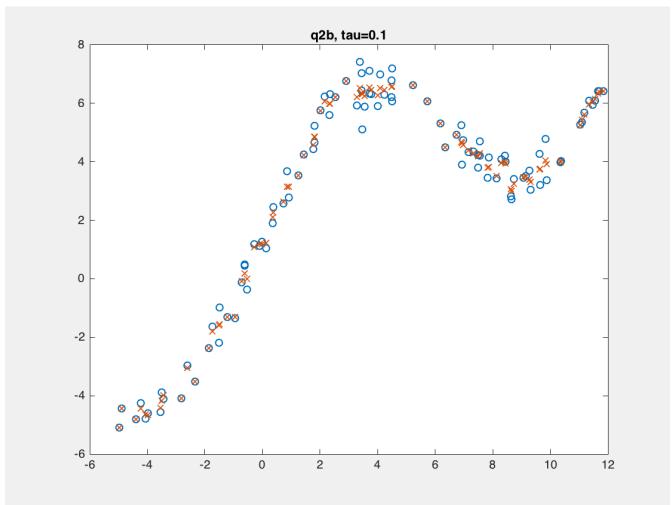
1(b)

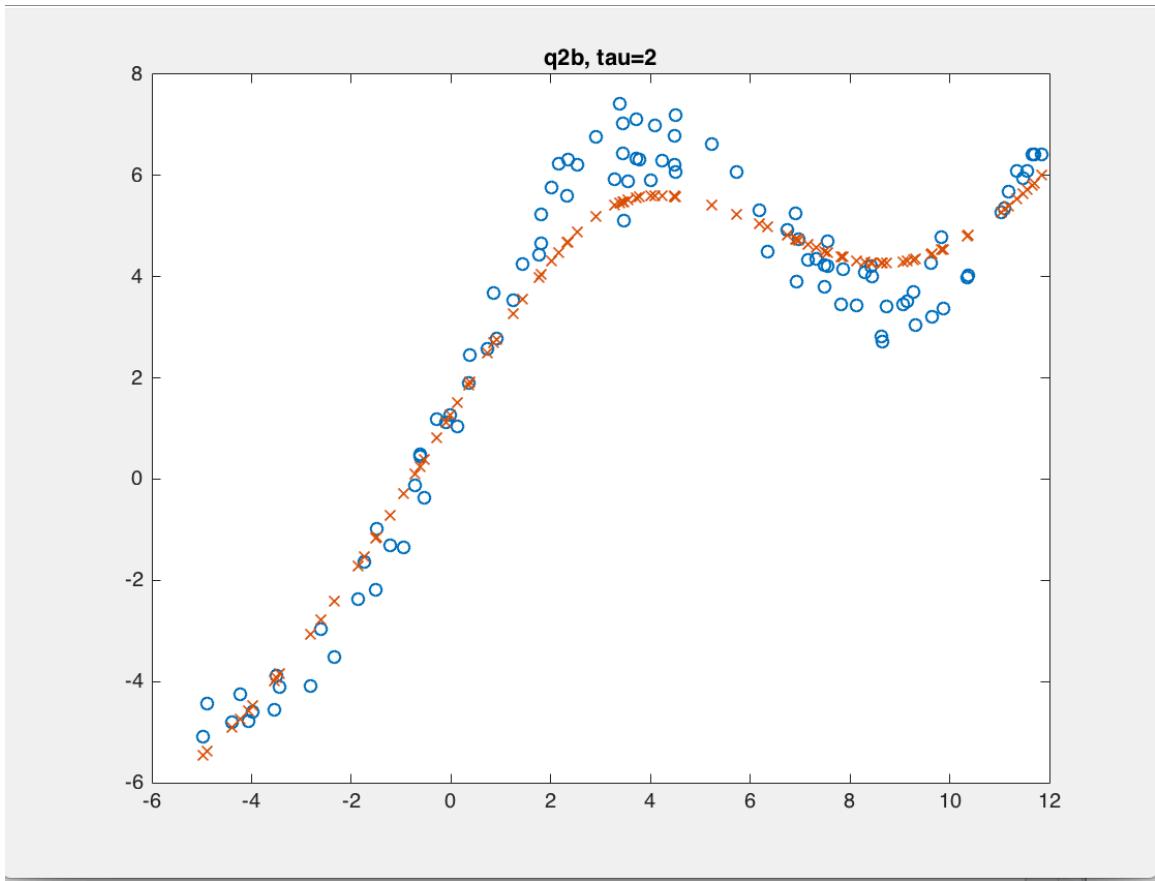


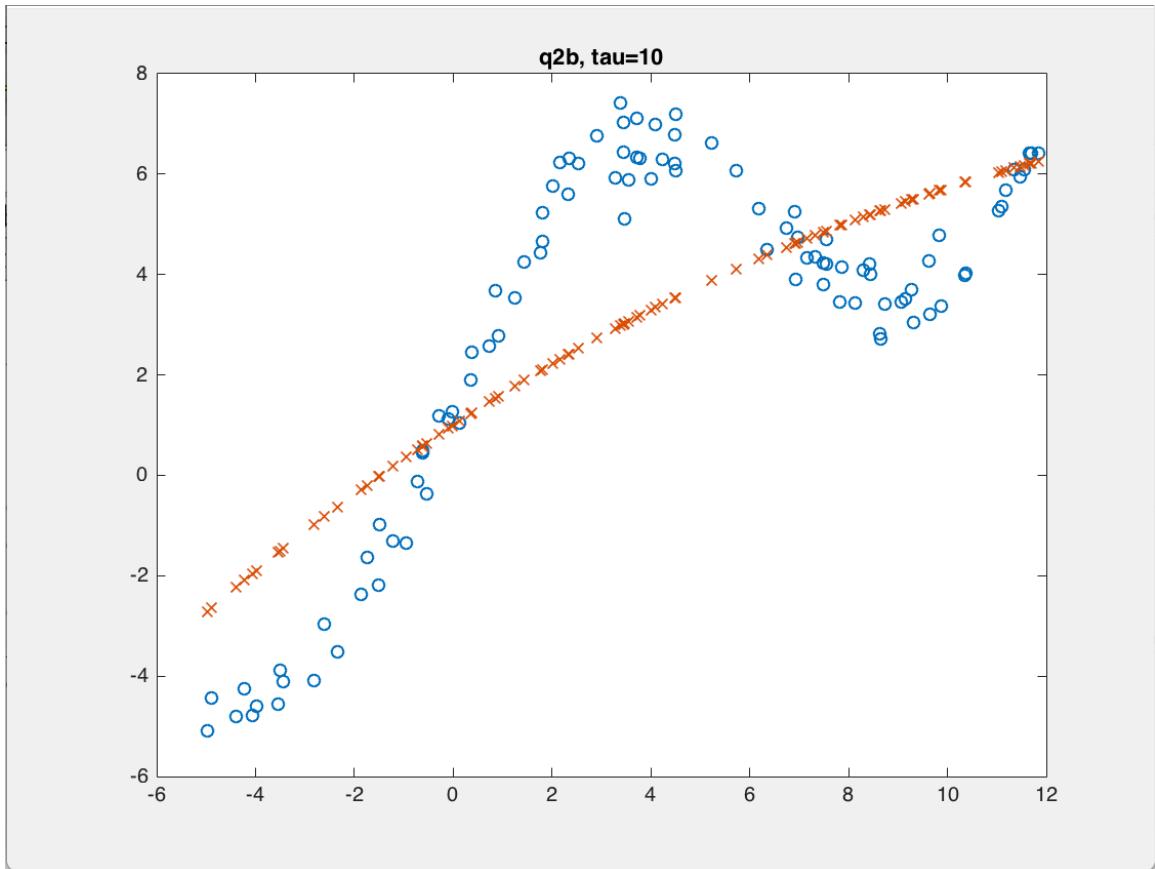
2(d) i., ii.

(i). $w = [0.9844, 0.5267]$



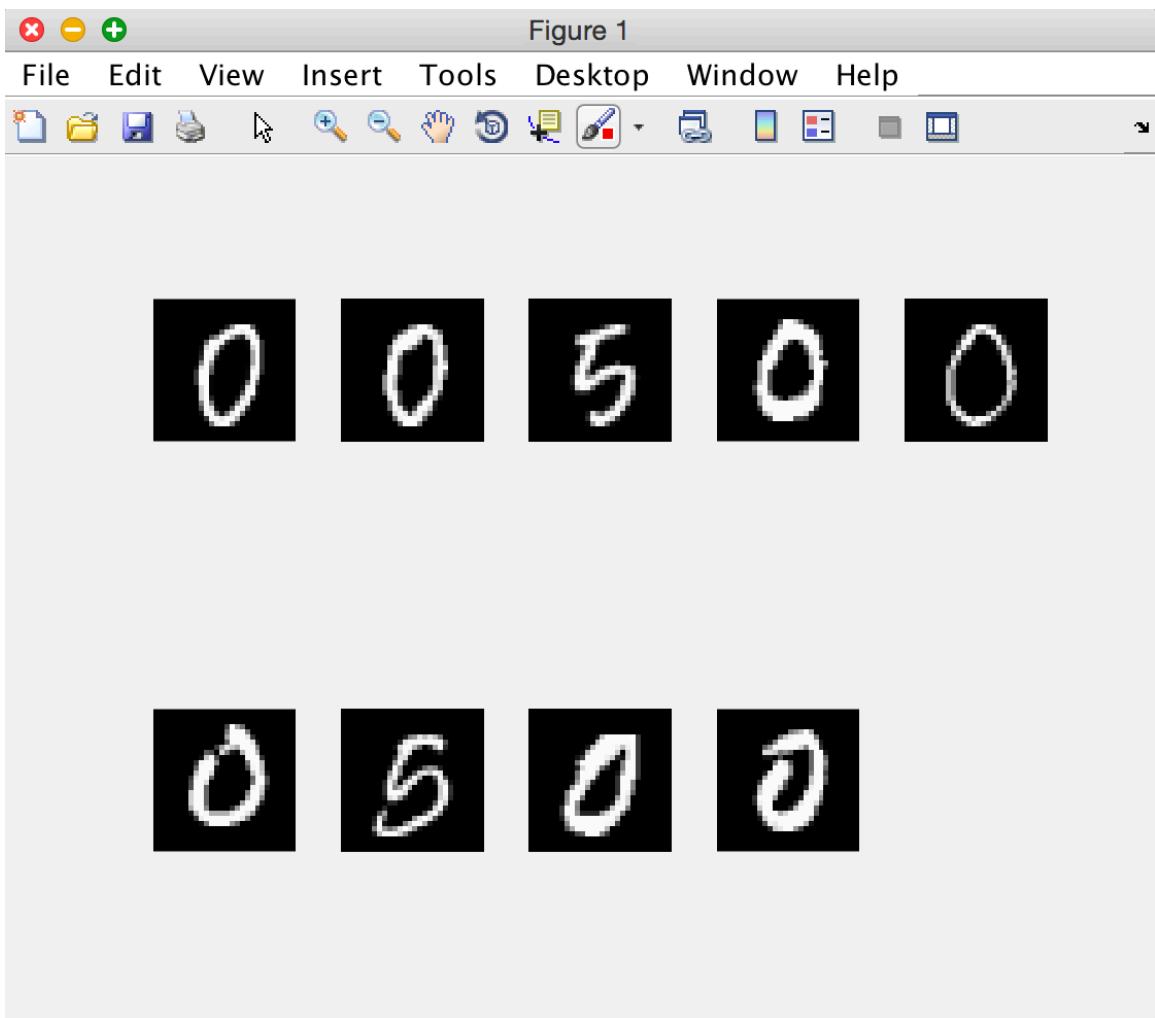




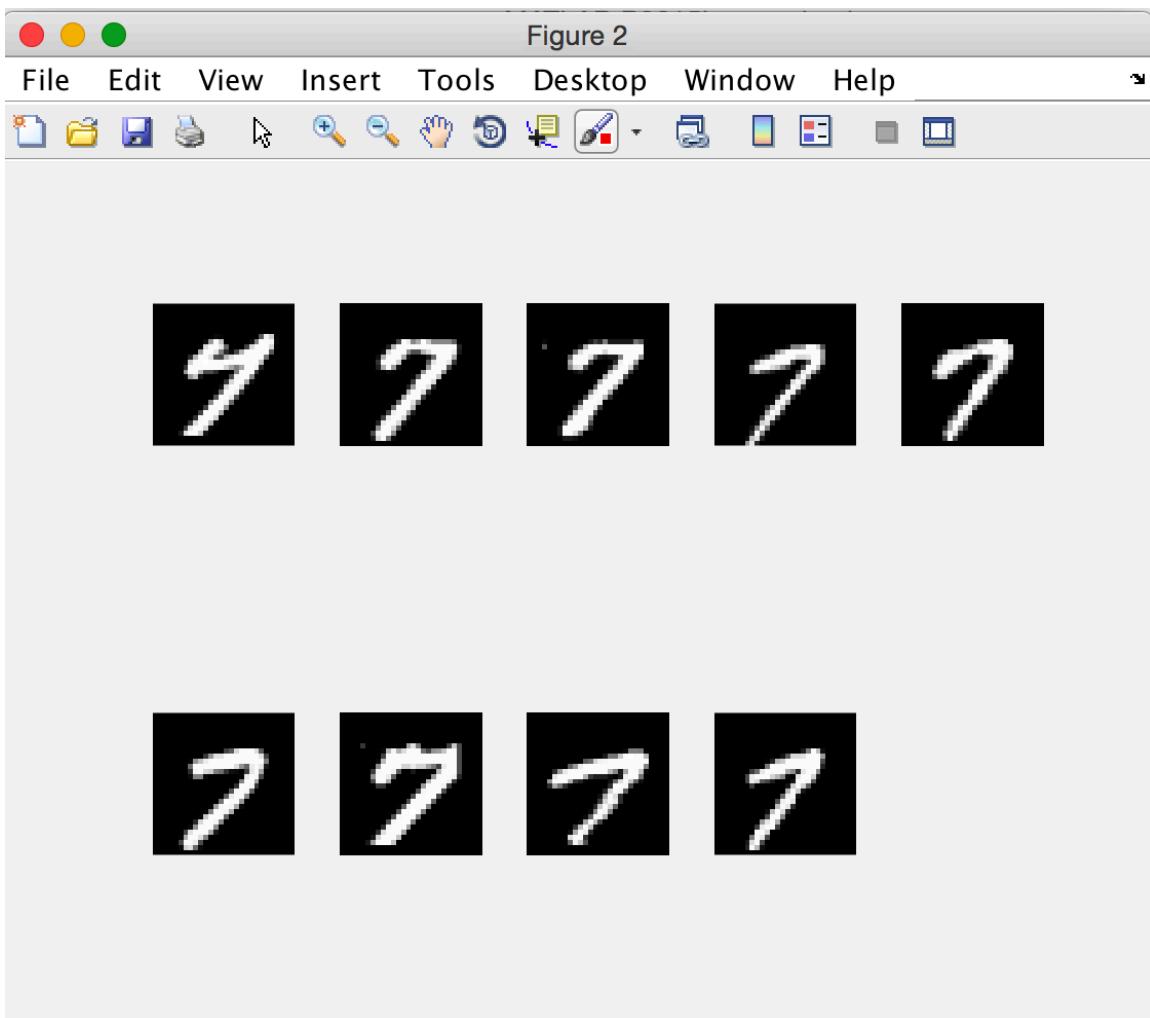


3(a).

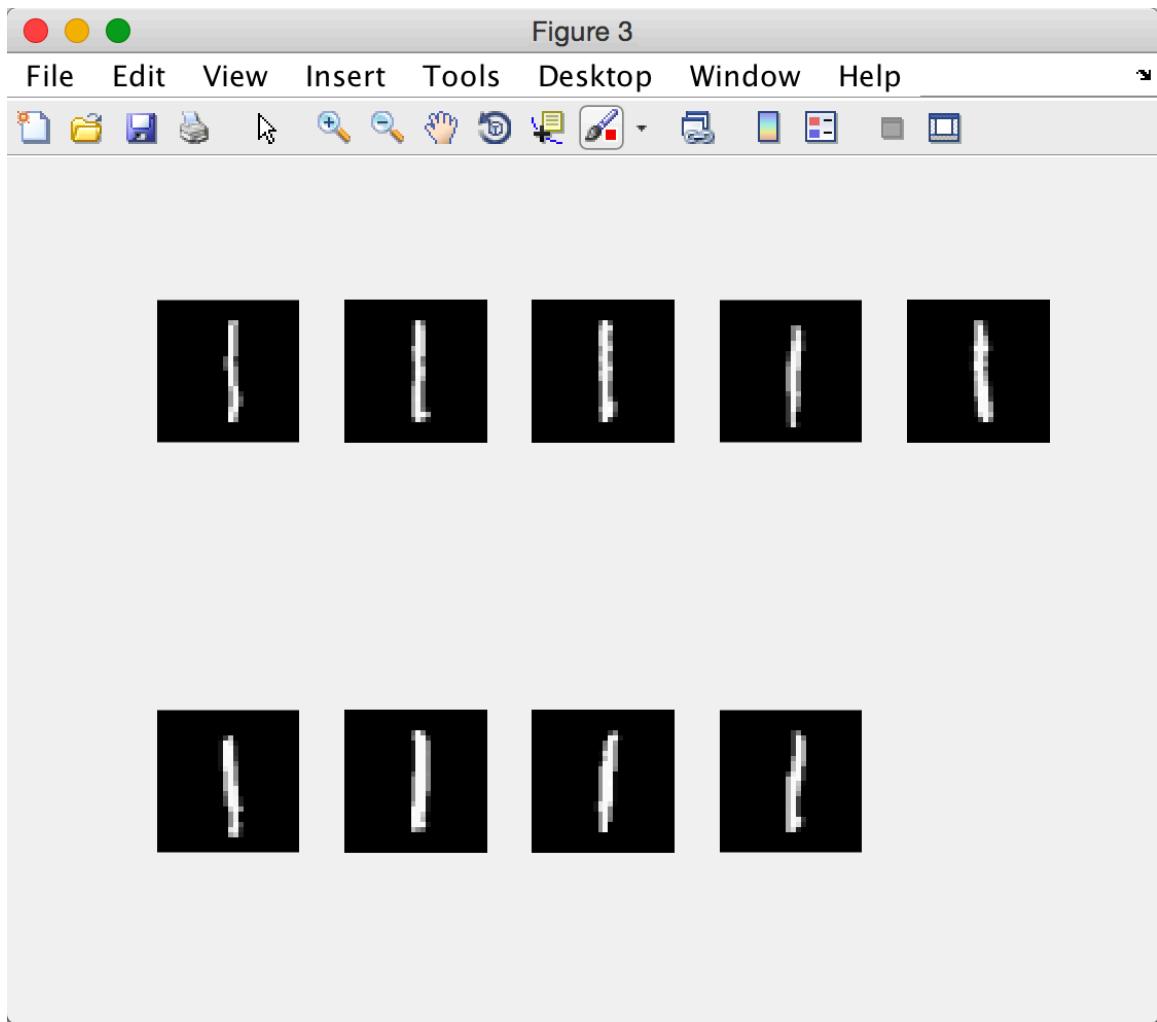
```
nearest index 1 for test point 1 is 953
nearest index 2 for test point 1 is 691
nearest index 3 for test point 1 is 321
nearest index 4 for test point 1 is 211
nearest index 5 for test point 1 is 193
nearest index 6 for test point 1 is 837
nearest index 7 for test point 1 is 82
nearest index 8 for test point 1 is 543
class label for test point 1 is 0
```



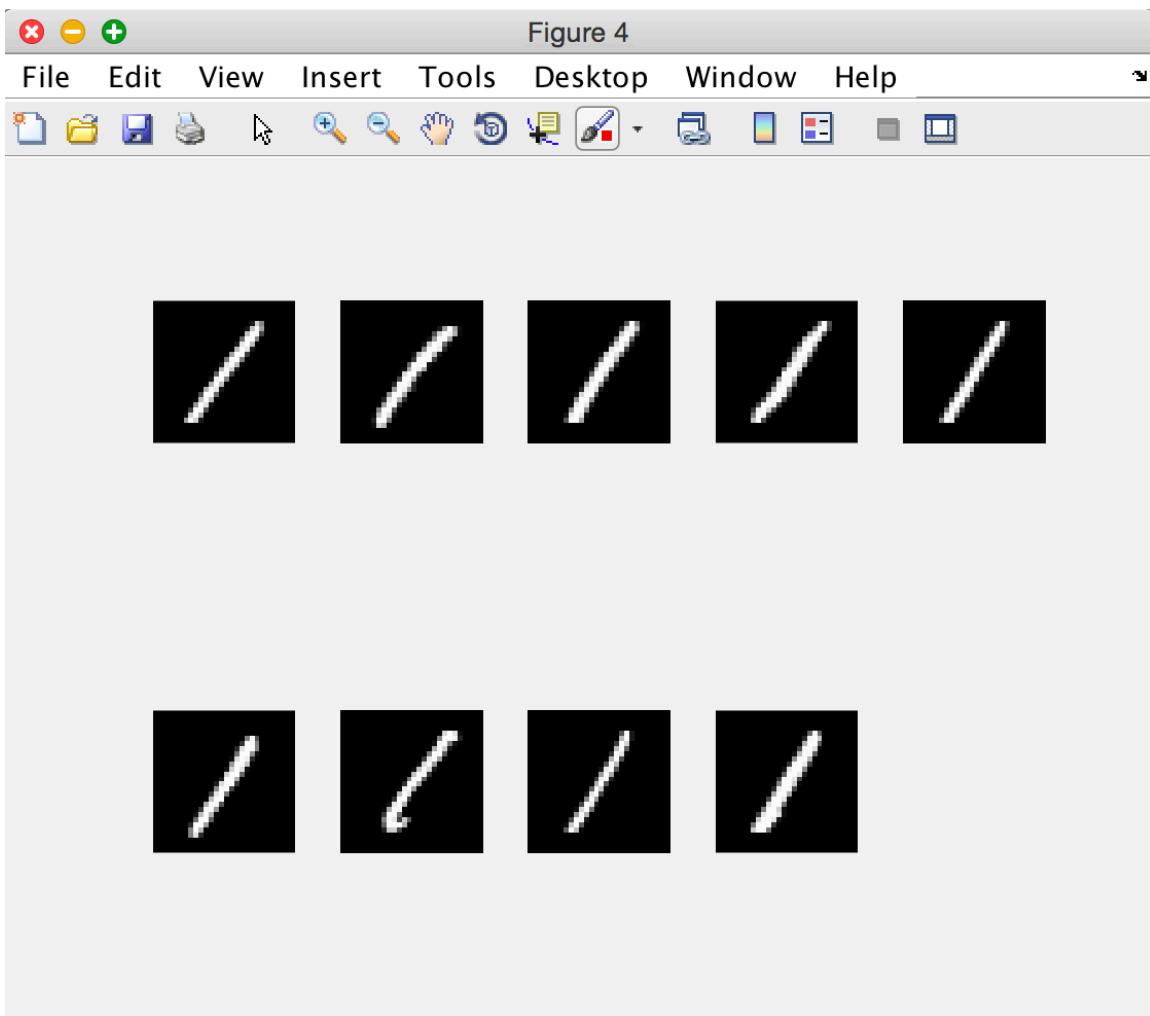
```
nearest index 1 for test point 2 is 724
nearest index 2 for test point 2 is 754
nearest index 3 for test point 2 is 72
nearest index 4 for test point 2 is 16
nearest index 5 for test point 2 is 912
nearest index 6 for test point 2 is 982
nearest index 7 for test point 2 is 194
nearest index 8 for test point 2 is 80
class label for test point 2 is 7
```



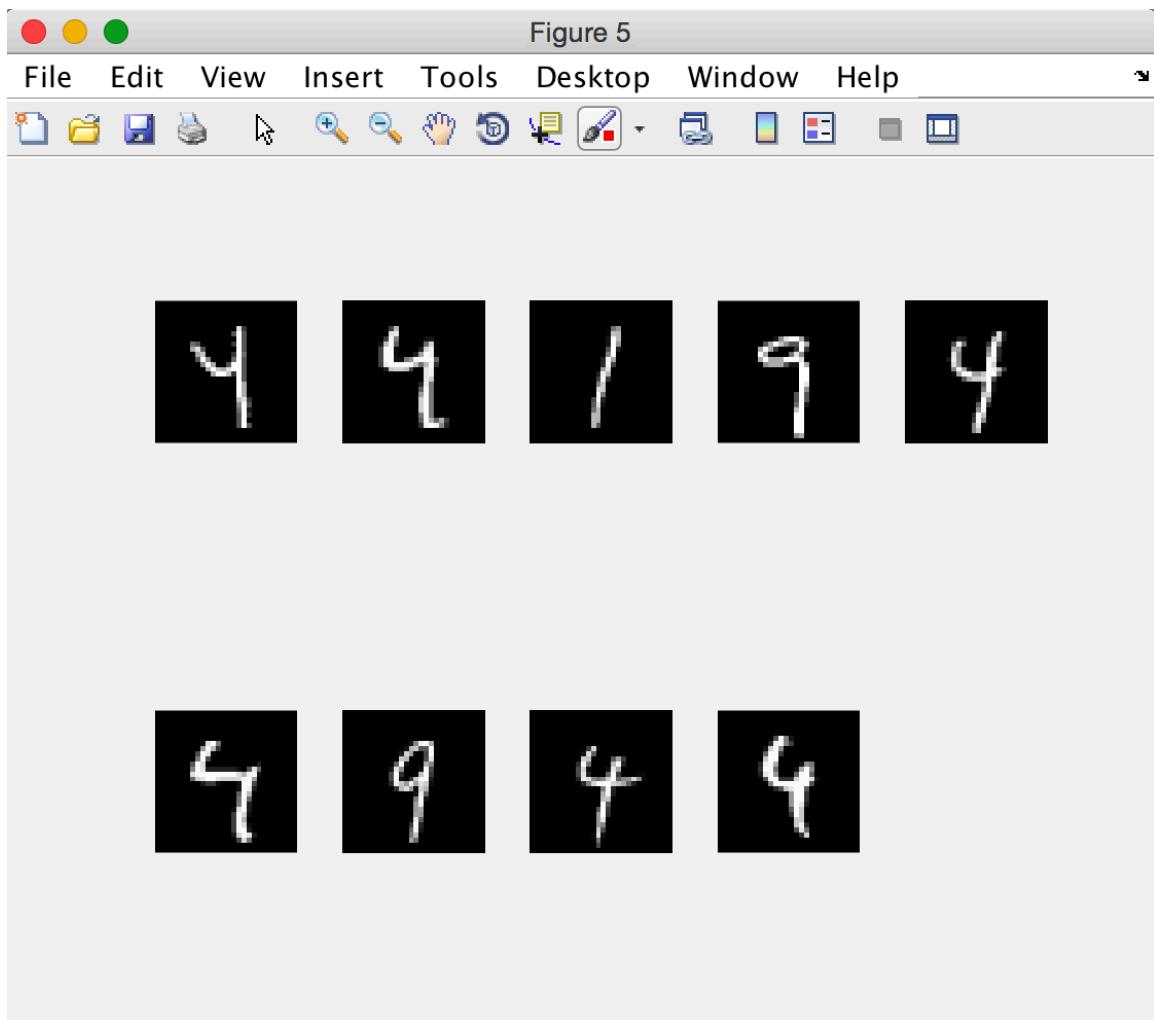
```
nearest index 1 for test point 3 is 485
nearest index 2 for test point 3 is 399
nearest index 3 for test point 3 is 455
nearest index 4 for test point 3 is 639
nearest index 5 for test point 3 is 9
nearest index 6 for test point 3 is 679
nearest index 7 for test point 3 is 113
nearest index 8 for test point 3 is 311
class label for test point 3 is 1
```



```
.  
nearest index 1 for test point 4 is 588  
nearest index 2 for test point 4 is 952  
nearest index 3 for test point 4 is 594  
nearest index 4 for test point 4 is 962  
nearest index 5 for test point 4 is 476  
nearest index 6 for test point 4 is 638  
nearest index 7 for test point 4 is 994  
nearest index 8 for test point 4 is 920  
class label for test point 4 is 1
```



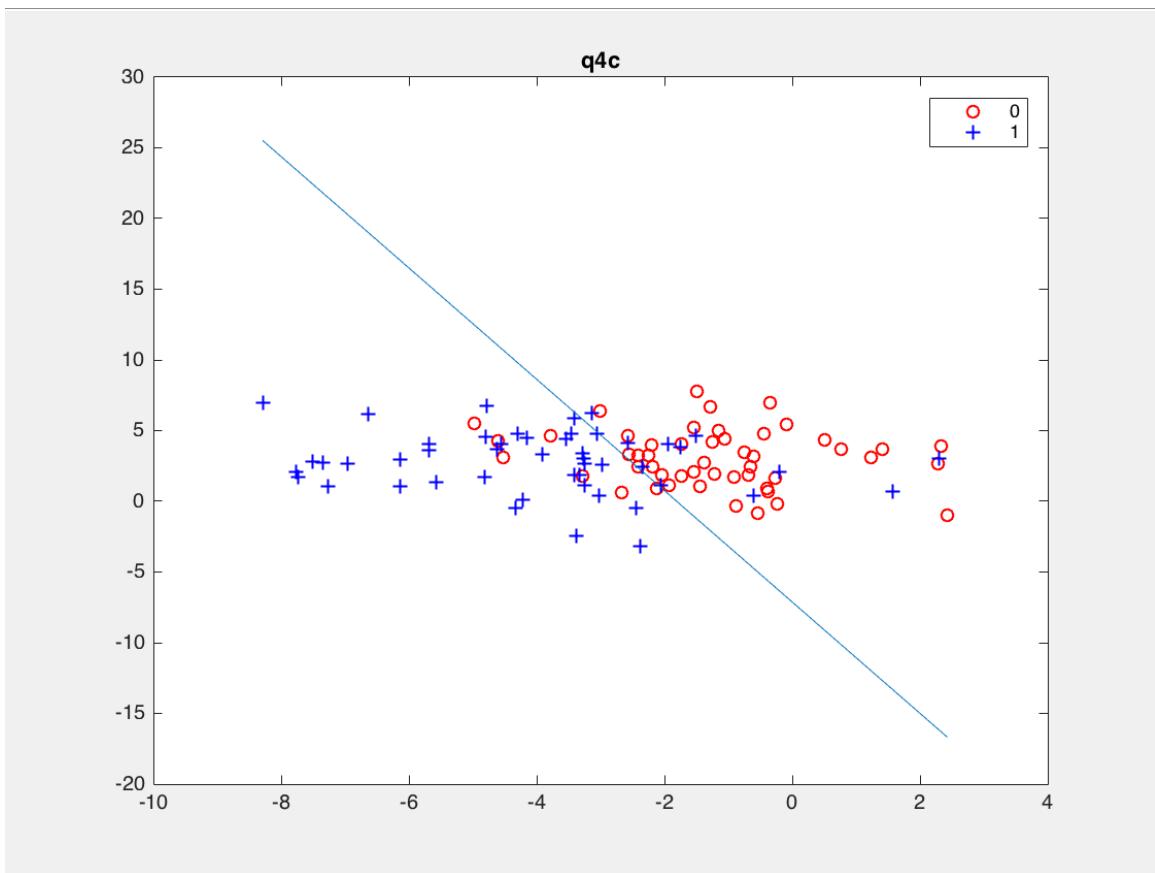
```
nearest index 1 for test point 5 is 577
nearest index 2 for test point 5 is 605
nearest index 3 for test point 5 is 305
nearest index 4 for test point 5 is 27
nearest index 5 for test point 5 is 719
nearest index 6 for test point 5 is 901
nearest index 7 for test point 5 is 413
nearest index 8 for test point 5 is 355
class label for test point 5 is 4
```



3(b)

Classification accuracy is 85.700000

4(b)



Deepak kumar

EECS 495 HW

i. (a) Batch gradient descent

$$w = [2.4420, -2.8074] \quad 531$$

(b) Stochastic gradient descent

$$w = [2.4375, -2.8119] \quad 519$$

(c) Newton's method

$$w = [2.4464, -2.8164] \quad 1$$

ii) In order of speed, the methods converged as follows:

Batch < Stochastic < Newton,

Batch - 531 iterations

Stochastic - 519 iterations

Newton - No iterations, 1 computation

(b) Plot given

ii) I would say a polynomial of degree 3-6 would best fit the data. There was evidence of overfitting for a polynomial of degree 10, when the E_{RMS} spiked up.

$$2(z) E_D(\vec{w}) = \frac{1}{2} \sum_{t=1}^N r^{(i)} (w^T x^{(i)} - t^{(i)})^2$$

$$\text{Substitute } Z = w^T x^{(i)} - t^{(i)}$$

$$= E_D(\vec{w}) = \frac{1}{2} \sum_{i=1}^N r^{(i)} Z^{(i)2}$$

Transform $r^{(i)} \Rightarrow R$, diagonal matrix with $r^{(i)}$ on diagonal

$$E_D(\vec{w}) = \sum_{i=1}^N Z^{(i)} R_{ii} Z^{(i)} = Z^T R Z$$

$$= (w^T x - t)^T R (w^T x - t)$$

$$= (\vec{w} - t)^T R (\vec{w} - t), \text{ by matrix property rules.}$$

R is the diagonal matrix with values

$$R = \begin{cases} R_{ii} = \frac{1}{2} r^{(i)} \\ R_{ij} = 0 \end{cases}$$

$$(b) (\Phi w - t)^T R (\Phi w - t)$$

Before taking gradient, rewrite

$$= (w^T \Phi^T - t^T)(R \Phi w - Rt)$$

$$E_p(\vec{w}) = w^T \Phi^T R \Phi w - t^T R \Phi w - w^T \Phi^T R t + t^T R t$$

Now, take gradient wrt w^T

$$\nabla_{w^T} E_p(\vec{w}) = \Phi^T R \Phi w - \Phi^T R t$$

and set = 0, to maximize.

$$\Phi^T R \Phi w - \Phi^T R t = 0$$

$$\Phi^T R \Phi w = \Phi^T R t$$

$$w = (\Phi^T R \Phi)^{-1} \Phi^T R t$$

$$(c) P(t^{(i)} | x_i, w) = \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{(t^{(i)} - w^T x^{(i)})^2}{2(\sigma^{(i)})^2}\right)$$

We have

$$\log P(t^{(1)}, \dots, t^{(n)} | x, w),$$

$$= \log \prod_{i=1}^n N(t^{(i)}; w^T \phi(x^{(i)}), (\sigma^{(i)})^2)$$

$$= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{1}{2(\sigma^{(i)})^2} \{t^{(i)} - w^T \phi(x^{(i)})\}^2\right)\right)$$

$$= \sum_{i=1}^n \left(\frac{1}{2} \log(\sigma^{(i)})^2 - \frac{1}{2} \log 2\pi - \frac{1}{2(\sigma^{(i)})^2} \{t^{(i)} - w^T \phi(x^{(i)})\}^2 \right)$$

2(c) Observe that the first two terms remain constants, so, simply put, we solve for $r^{(i)}$'s in terms of $\sigma^{(i)}$'s to be

$$r^{(i)} = \frac{1}{2(\sigma^{(i)})^2} \quad (18)$$

(d) (iii)

When T is too small or large, you have the problems of over/underfitting, where the curve it fits becomes too small/large or unreasonably tries to fit the data.

3. (c). KNN is a reasonable algorithm to use. Its advantages are that it is simple and a reasonable approximation of a good classifier. It also uses local information which yields good behavior at optimal values of k .

The disadvantages involve large storage requirements w/ more features and data. Additionally, if k is poorly chosen, the classifier accuracy decreases drastically, as adding too many neighbors ruins the probabilities.

4. (a) Start w/

$$\nabla_w \ell(w) = \sum_{n=1}^N (t^{(n)} - \sigma^{(n)}) \phi(x^{(n)})$$

We know $\sigma^{(n)} = \sigma(w^T \phi(x^{(n)}))$, and that $\frac{\partial}{\partial s} \sigma(s) = \sigma(s)(1-\sigma(s))$
 $= h(x^{(n)})$.

Re-write gradient

$$\nabla_w \ell(w) = \sum_{n=1}^N (t^{(n)} - \sigma(w^T \phi(x^{(n)})) \phi(x^{(n)})$$

Take derivative wrt w

$$= \sum_{n=1}^N -\phi(x^{(n)}) \sigma(w^T \phi(x^{(n)}))(1-\sigma(w^T \phi(x^{(n)}))) \phi(x^{(n)})$$

$$= -\sum_{n=1}^N \Phi_n h(x^{(n)}) (1-h(x^{(n)})) \Phi_n^T$$

Thus, the hessian is

$$H = -\sum_{n=1}^N \Phi_n h(x^{(n)}) (1-h(x^{(n)})) \Phi_n^T, \text{ which is the desired form.}$$

Now, we wts

$$z^T H z \leq 0$$

Start w/

$$\begin{aligned} \sum_i \sum_j z_i x_i x_j z_j &= \sum_i z_i x_i \sum_j x_j z_j = (z^T x)(x^T z) \\ &= (x^T z)^2 \geq 0 \quad \square \end{aligned}$$

4(a) Now, look @ the form of the hessian.

$$H = - \sum_n \mathbb{I}_{n_i} h(x^{(n)}) (1-h(x^{(n)})) \mathbb{I}_{n_j}$$

$$\therefore z^T H z = - z^T \sum_n \mathbb{I}_{n_i} h(x^{(n)}) (1-h(x^{(n)})) \mathbb{I}_{n_j} z$$

By our previous result, we thus see it is in the same form, so

$$- z^T \sum_n \mathbb{I}_{n_i} h(x^{(n)}) (1-h(x^{(n)})) \mathbb{I}_{n_j} z \geq 0$$

or $z^T H z \leq 0$



(b) $w = [-1.4196, -0.7849, -0.1993]$

(c) Given plot.