Exercise 3.4:
Part A:
From pg85 and 86 of textbook we know:

$$\mathrm{H} = \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}.$$

And $\hat{\mathbf{y}} = \mathrm{H}\mathbf{y}$.
From question we know:

$$y = \mathbf{w}^{*\mathrm{T}}\mathbf{x} + \epsilon$$

So $\hat{y}$ = X$(X^T X)^{-1}X^T$ *($W^{*T}x + \epsilon$)
= X$(X^T X)^{-1}X^T w^{*T}x$+X$(X^T X)^{-1}X^T\epsilon$
=xw^* + Hϵ

Part b:
From last part we know $\hat{y}$ = xw^* + Hϵ and y = $w^{*T}x + \epsilon$
$\hat{y}$ - y = xw^* + Hϵ - $w^{*T}x - \epsilon$
= (H-I)ϵ
Where (H - I) is a matrix.

Part c:
E_{in} = $\frac{1}{N}(\hat{y}$ - y$)^2$
= $\frac{1}{N}$((H-I)$\epsilon)^2$
=$\frac{1}{N}$((H-I)$\epsilon)^T$(H-I)ϵ
=$\frac{1}{N}\epsilon^T$*ϵ*(H-I$)^2$
(H-I$)^2$ = (I - H$)^2$
On exercise 3.3 part c, we know:

(c) If I is the identity matrix of size N, show that $(\mathrm{I} - \mathrm{H})^K = \mathrm{I} - \mathrm{H}$ for any positive integer K.

(I - H$)^2$ = I - H
E_{in} = $\frac{1}{N}\epsilon^T$*ϵ(I - H)
Part d:
From part c we know $E_{in}(w_{lin})$ = $\frac{1}{N}\epsilon^T$*ϵ(I - H)
$E_D(E_{in}(w_{lin}))$ = $E_D(\frac{1}{N}\epsilon^T$*$\epsilon$(I - H))

$= \frac{1}{N} E_D(\epsilon^T * \epsilon(I - H))$

$= \frac{1}{N} E_D(\epsilon^T * \epsilon - \epsilon^T * \epsilon H)$

From the question we know ϵ has 0 mean and α^2 variance. $E_D(\epsilon^T * \epsilon) = N\alpha^2$.

$E_D(\epsilon^T * \epsilon H)$ = trace(H)α^2 from exercise 3.3(d):

(d) Show that $\text{trace}(H) = d + 1$, where the trace is the sum of diagonal elements. *[Hint:* $\text{trace}(AB) = \text{trace}(BA)$.*]*

We know trace(H) = d +1

$E_D(E_{in}(w_{lin})) = \frac{1}{N}(N\alpha^2 - \alpha^2(d+1)) = \frac{1}{N}\alpha^2(N-(d+1)) = \alpha^2(1-\frac{d+1}{N})$

Part e:

Since $y = w^{*T}x + \epsilon$

$y' = w^T x + \epsilon'$

$E_{test} = \frac{1}{N}(\hat{y} - y')^2$

$= \frac{1}{N}(Xw^* + H\epsilon - (w^{*T}X + \epsilon'))^2$

$= \frac{1}{N}(H\epsilon - \epsilon')^2$

$= \frac{1}{N}(HH\epsilon\epsilon + \epsilon'\epsilon' - 2H\epsilon\epsilon')$

From part b of exercise 3.3:

(b) Show that $H^K = H$ for any positive integer K.

HH = H

$E_{test} = \frac{1}{N}(H\epsilon\epsilon + \epsilon'\epsilon' - 2H\epsilon\epsilon')$

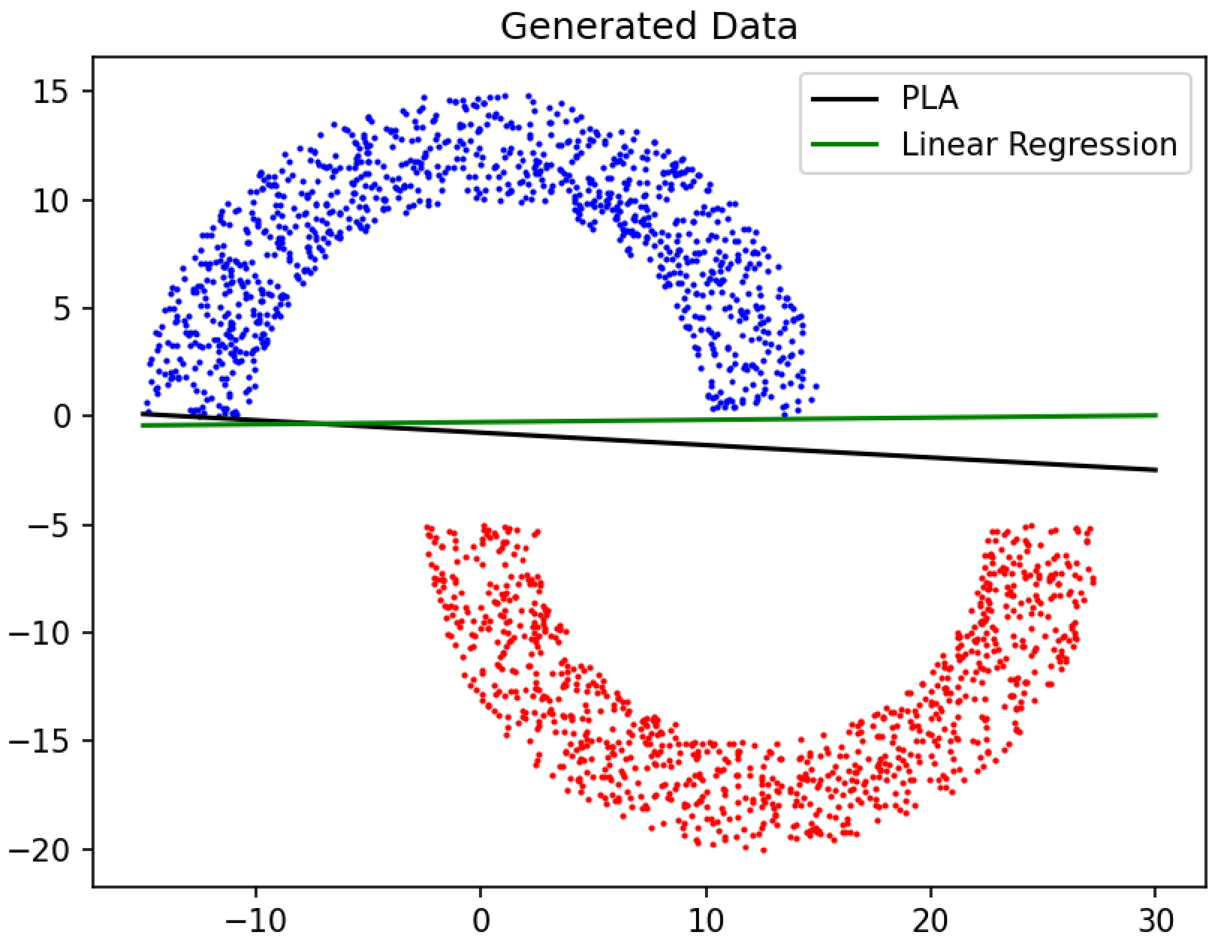
$E_{D,\epsilon'}(E_{test}) = E_{D,\epsilon'}(\frac{1}{N}(H\epsilon\epsilon + \epsilon'\epsilon' - 2H\epsilon\epsilon'))$

$= \frac{1}{N}(E_{D,\epsilon'}(H\epsilon\epsilon) + E_{D,\epsilon'}(\epsilon'\epsilon') - E_{D,\epsilon'}(2H\epsilon\epsilon'))$

$= \frac{1}{N}((d+1)\alpha^2 + N\alpha^2 + 0)$

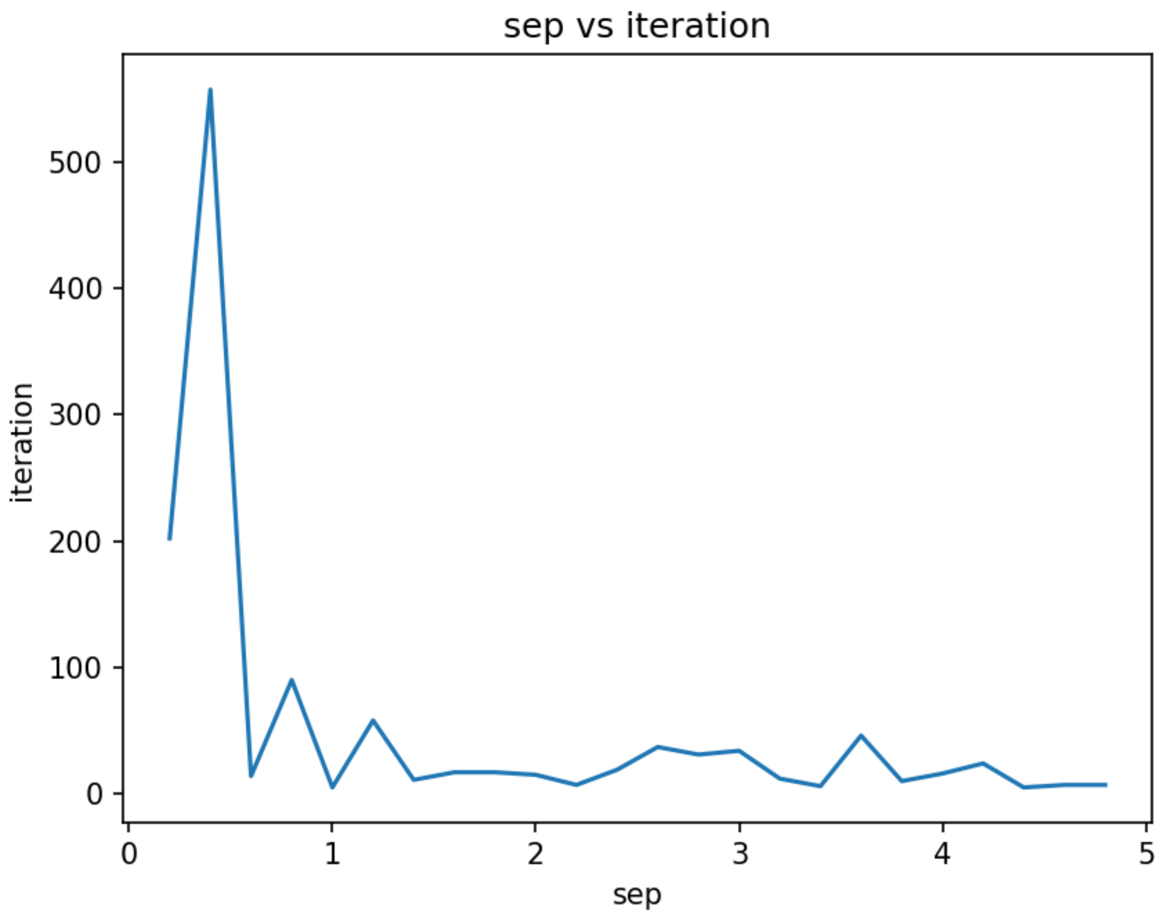
$= \alpha^2(1+\frac{d+1}{N})$

Problem 3.1



As shown in the graph, both methods successfully separated data.

Problem 3.2



From the graph we can see that when sep is very small it takes a lot of iteration, but when sep gets bigger than 1 the number of iteration becomes much less.

Problem 3.8:

From the question we know E_{out}(h) = E$\left[(h(x) - y)^2\right]$

= E$\left[((h(x) - h^*(x)) + (h^*(x) - y))^2\right]$

=E$\left[(h(x) - h^*(x))^2 + (h^*(x) - y)^2 + 2((h(x) - h^*(x) * (h^*(x) - y))\right]$

= E[$(h(x) - h^*(x))^2$] + E[$(h^*(x)$ - y)^2 + 2E[$((h(x) - h^*(x) * (h^*(x) - y))$]

2E[$((h(x) - h^*(x) * (h^*(x) - y))$] = 2 * E[$(h(x) - h^*(x)$] * E[$(h^*(x) - y)$]

From the question we know E[y|x] = $h^*(x)$

Since E[y] = E[E[y|x]], E[y] = E[$h^*(x)$]

E[$(h^*(x) - y)$] = E[$h^*(x)$] - E[y] = 0

So 2E[$((h(x) - h^*(x) * (h^*(x) - y))$] = 0

E_{out}(h) = E[$(h(x) - h^*(x))^2$] + E[$(h^*(x)$ - y$)^2$]

To make E_{out}(h) be min, we want E[$(h^*(x)$] = E[y]

y = $h^*(x)$ + ϵ(x)

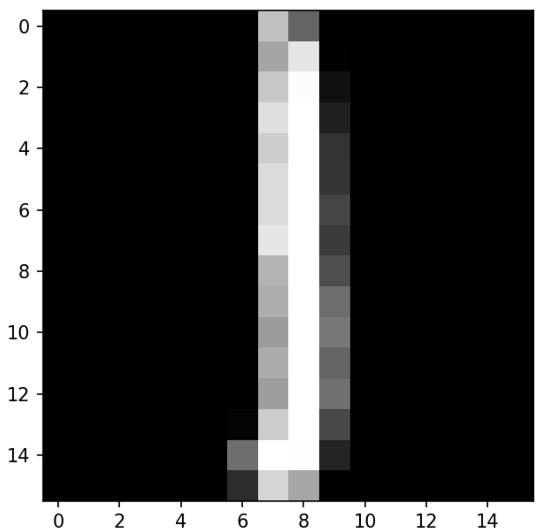
E[y] = E[$h^*(x)$ + ϵ(x)]

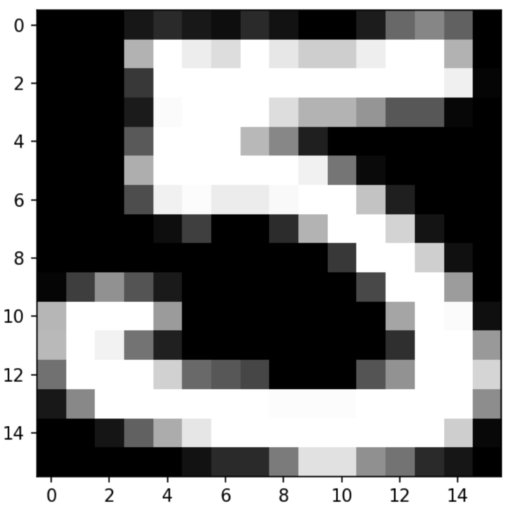
E[y] = E[$h^*(x)$] + E[ϵ(x)]

E[ϵ(x)] = 0

Hand Writing:

1:





2:

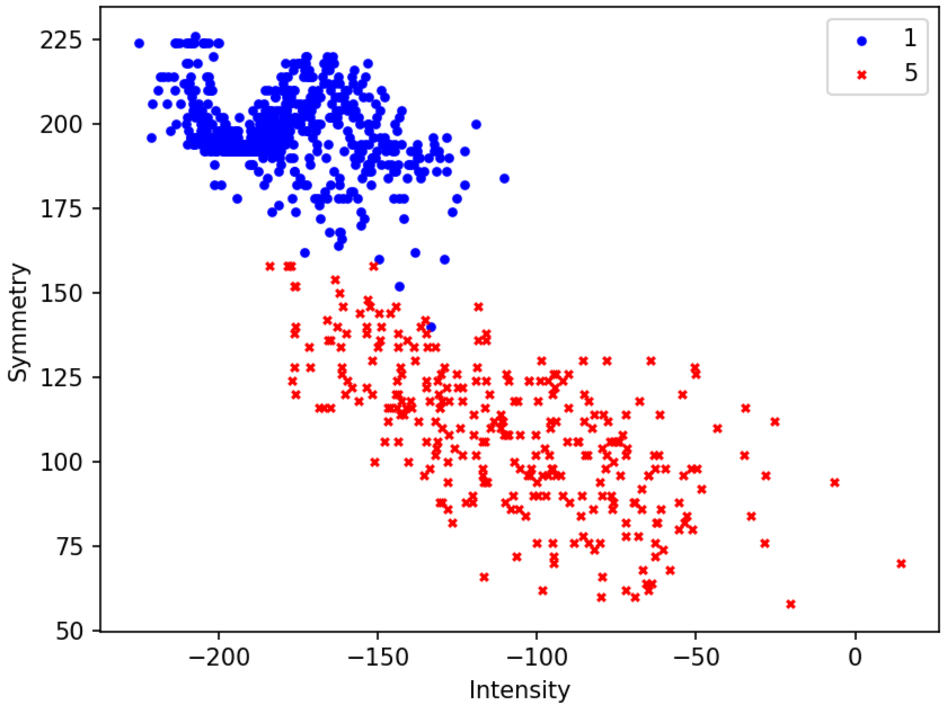
Intensity: the sum of all data points. Assume D is a 16 by 16 matrix.

Intensity = $\sum_{i=1}^{16}\sum_{j=1}^{16} D_{i,j}$

Symmetry: the number of pixels that is symmetry. Reshape D in to a 16 * 16

With $\sum_{i=1}^{16}\sum_{j=1}^{16}$, Symmetry +=1 for any($D_{i,j}$ = $D_{i,\,16-j}$)

3:


225
200
175
150
125
100
75
50
Symmetry
−200
−150
−100
−50
0
Intensity
1
5