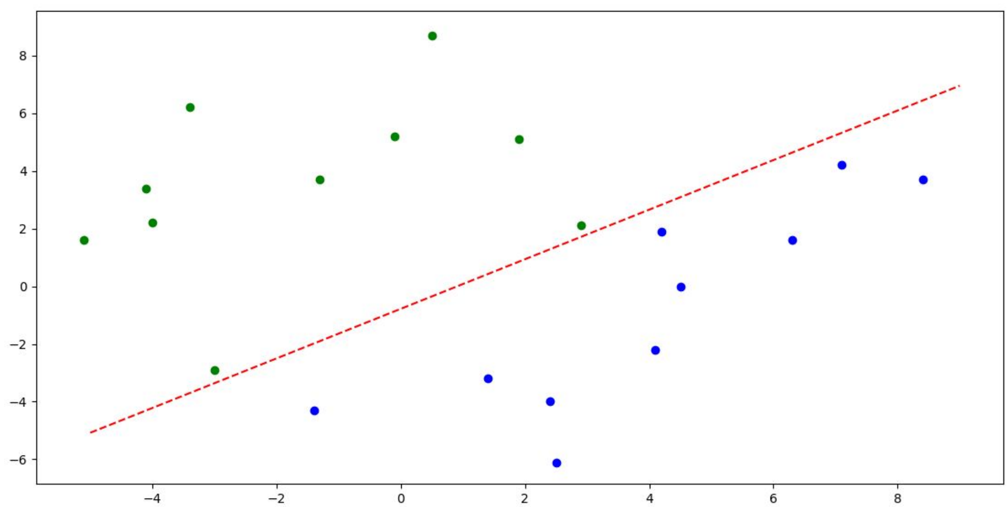
SM in AI

Assignment 1

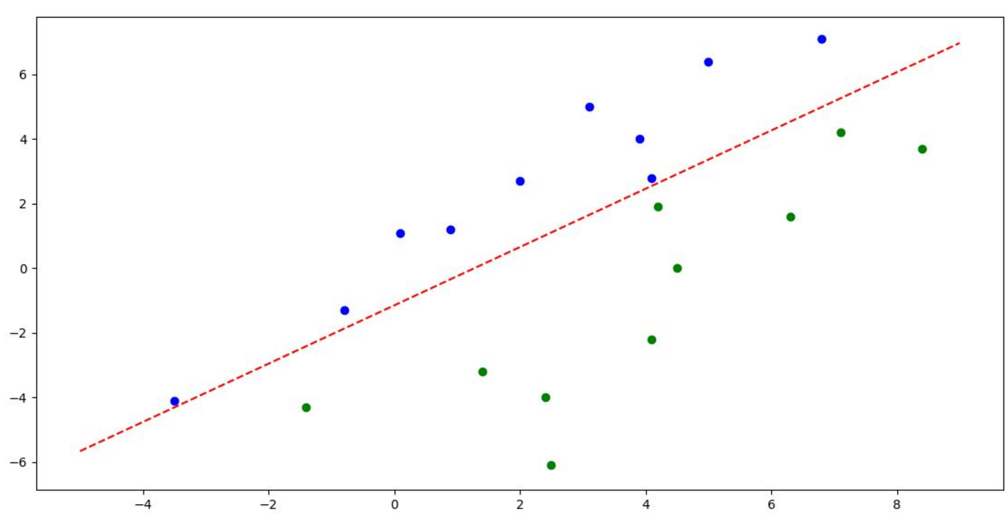
Hemanth Kumar Veeranki
201401145

Question1

Dataset 2



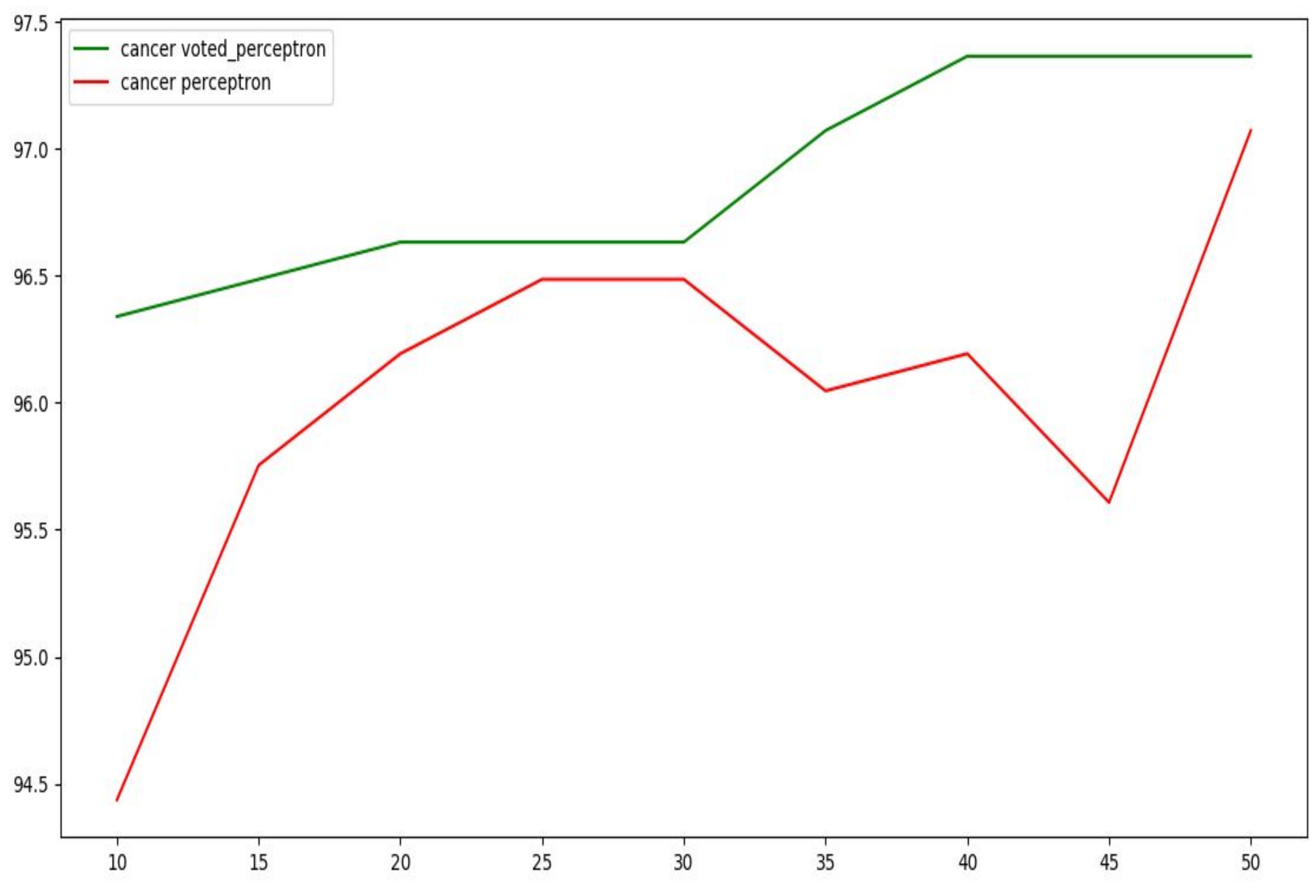
Dataset 1

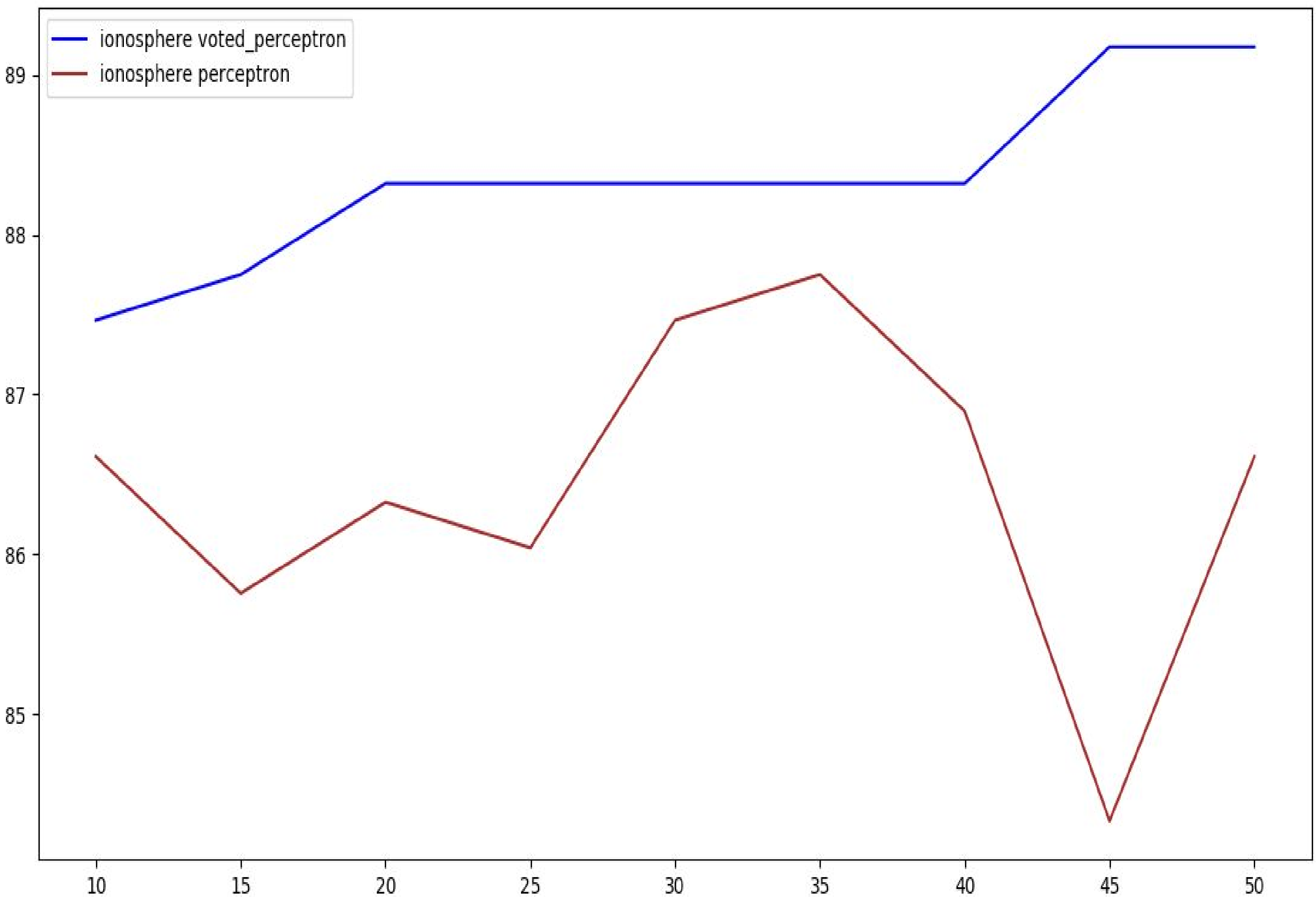


First data set takes 163 iterations where 2nd data set takes 93 iterations. This is because in 2nd set data is distant from each other so converges easily.

Question2

This is because vanilla perceptron gives more importance to last points than the beginning points.What we would like is for weight vectors that “survive” a long time to get more say than weight vectors that are overthrown quickly. Voted .erceptron performs far better than vanilla perceptron on large sets.

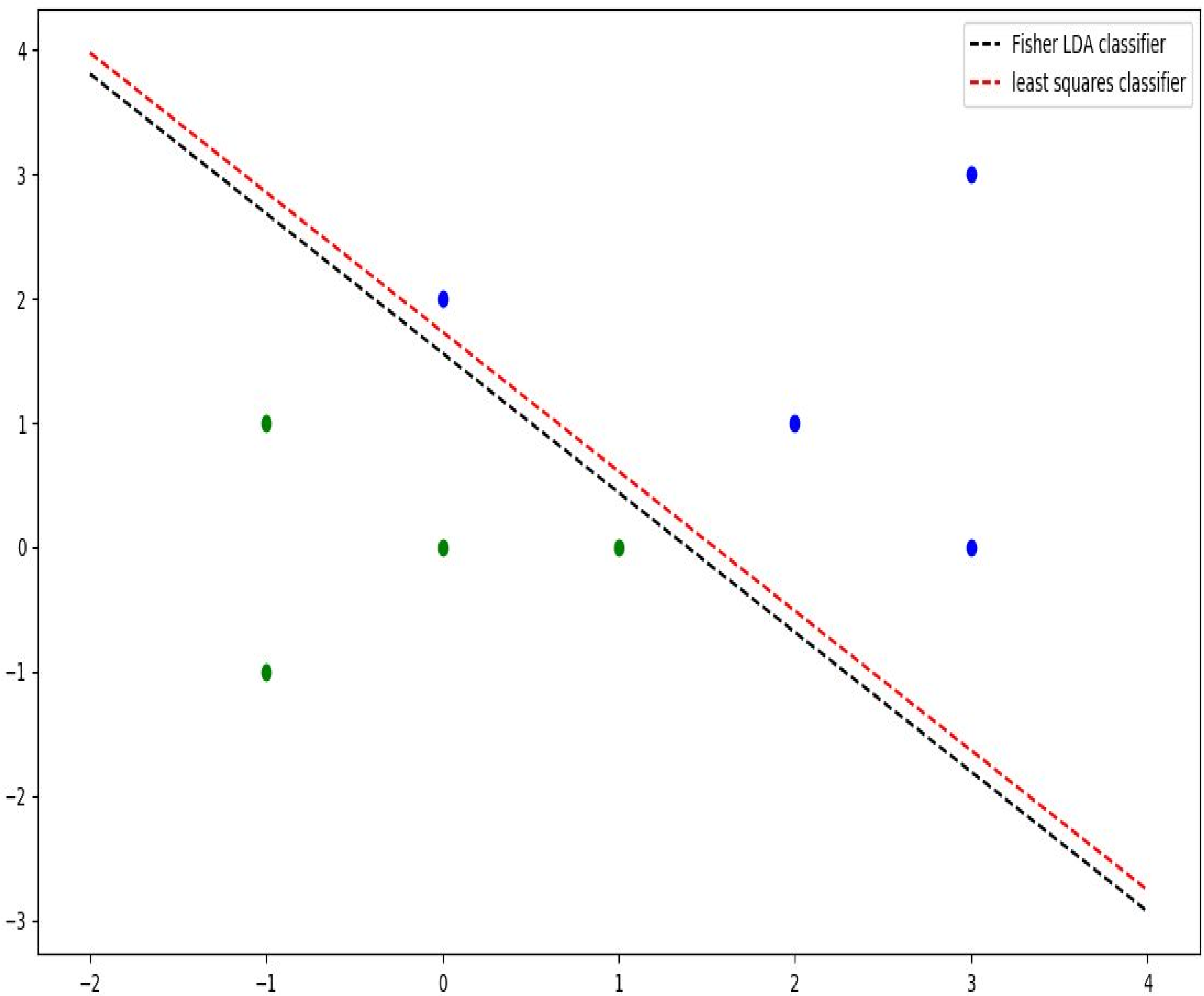




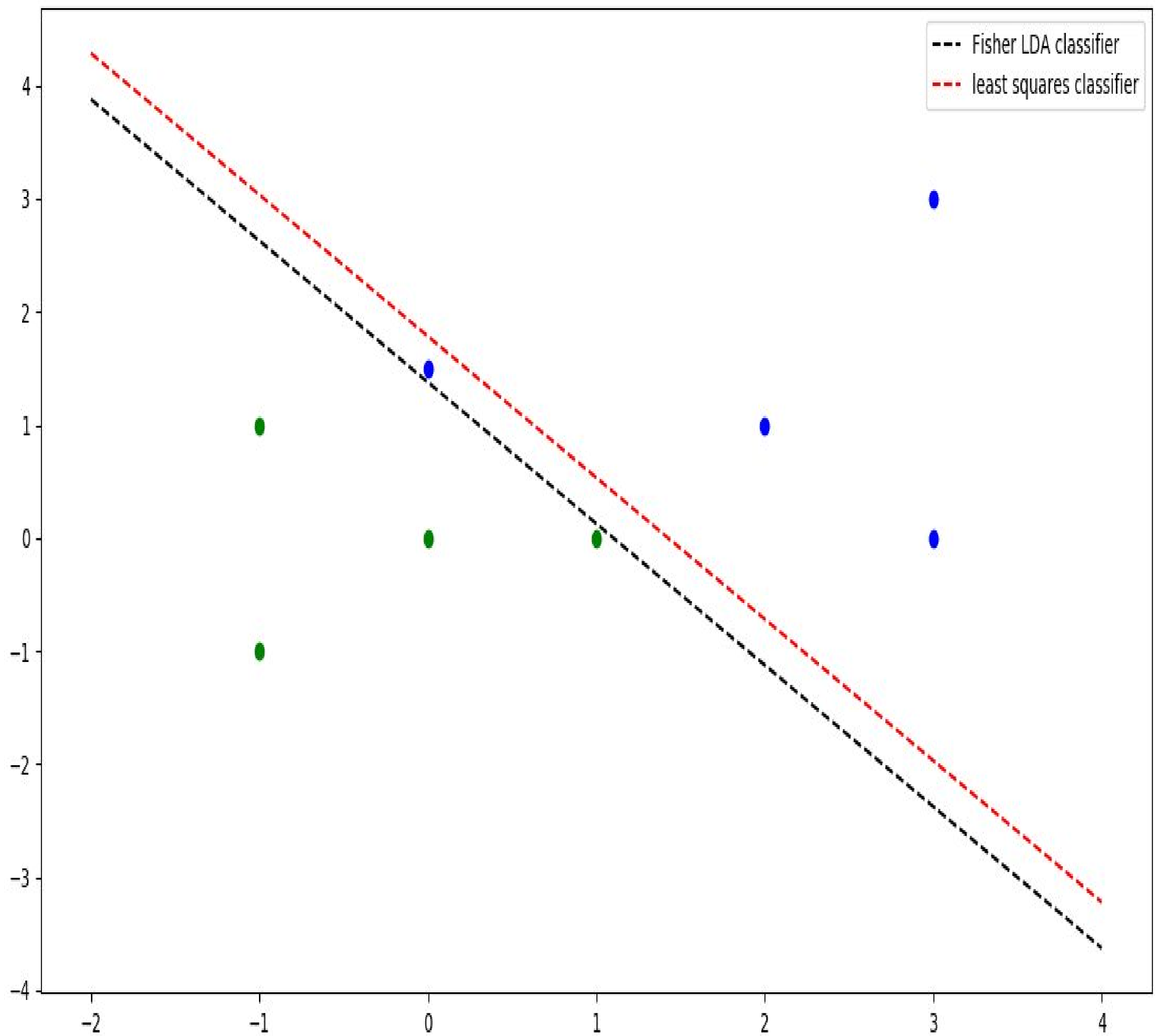
In both the cases we can see voted perceptron having far better accuracy in both datasets. These are 10 fold accuracies for corresponding epochs.

Question3

Fishers LDA and Least squares approach both classify the data properly in dataset1 for dataset2 we can see least squares will not properly classify it. This is because we need to choose the margin efficiently. For Fishers LDA we will chose it so it will be able to classify properly in both the cases



For the below dataset we can see that the Least squares approach classifier will classify one point wrongly.



Question 4

Below is the written solution for Question4 without loss of generality we will assume the data is centered $\Rightarrow$ $\sum x_i = 0$

So $m_1 \mu_1 + m_2 \mu_2 = 0$

Now $A = \begin{bmatrix} X_1 & 1 \\ X_2 & 1 \end{bmatrix}$ where $X_1 = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_{m_1}^T \end{bmatrix}$, $X_2 = \begin{bmatrix} x_{m_1}^T \\ x_{m_1+1}^T \\ \vdots \\ x_{m_2}^T \end{bmatrix}$

$X_1 \in C_1 \quad X_2 \in C_2$;

$Y = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} \quad M_1 = \begin{bmatrix} \frac{m}{m_1} \\ \frac{m}{m_1} \\ \vdots \end{bmatrix} m_1 \text{ times} \quad M_2 = \begin{bmatrix} -\frac{m}{m_2} \\ \vdots \\ -\frac{m}{m_2} \end{bmatrix} (m_2 \text{ times})$

for least squares approach.

$A^T A \tilde{w} = A^T Y \quad ; \quad \tilde{w} = \begin{bmatrix} w \\ b \end{bmatrix}$

$A^T A = \begin{bmatrix} X_1^T & X_2^T \\ 1^T & 1^T \end{bmatrix} \begin{bmatrix} X_1 & 1 \\ X_2 & 1 \end{bmatrix} \quad ; \quad A^T Y = \begin{bmatrix} X_1^T & X_2^T \\ 1^T & 1^T \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$

$[X_1^T \; X_2^T] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \sum_i^{\text{class 1}} x_i x_i^T + \sum_i^{\text{class 2}} x_2 \cdot x_2^T = MS$

$[X_1^T \; X_2^T] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sum^{\text{class 1}} x_i + \sum^{\text{class 2}} x_i$

$= (m_1 \mu_1 + m_2 \mu_2)^T$

$[1^T \; 1^T] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = (m_1 \mu + m_2 \mu_2)$

$[1^T \; 1^T] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \underline{m}$

$$[x_1^T \; x_2^T] \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \frac{m}{m_1} \sum^{class 1} x_i - \frac{m}{m_2} \sum_i^{class 2} x_i$$

$$= m(\mu_1 - \mu_2)$$

$$[1^T \; 1^T] \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \frac{m}{m_1} \times m_1 - \frac{m}{m_2} \times m_2$$

$$= 0$$

$$\Rightarrow \begin{bmatrix} m S_w & (m_1\mu_1 + m_2\mu_2)^T \\ (m_1\mu_1 + m_2\mu_2) & m \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} m(\mu_1 - \mu_2) \\ 0 \end{bmatrix}$$

We know data is centered $\mu = 0 \Rightarrow m_1\mu_1 + m_2\mu_2$

$$\Rightarrow m S_w w + 0 = m(\mu_1 - \mu_2)$$

$$\Rightarrow w = \cancel{m} S_w^{-1} (\mu_1 - \mu_2)$$

$$\boxed{w = S_w^{-1}(\mu_1 - \mu_2)}$$