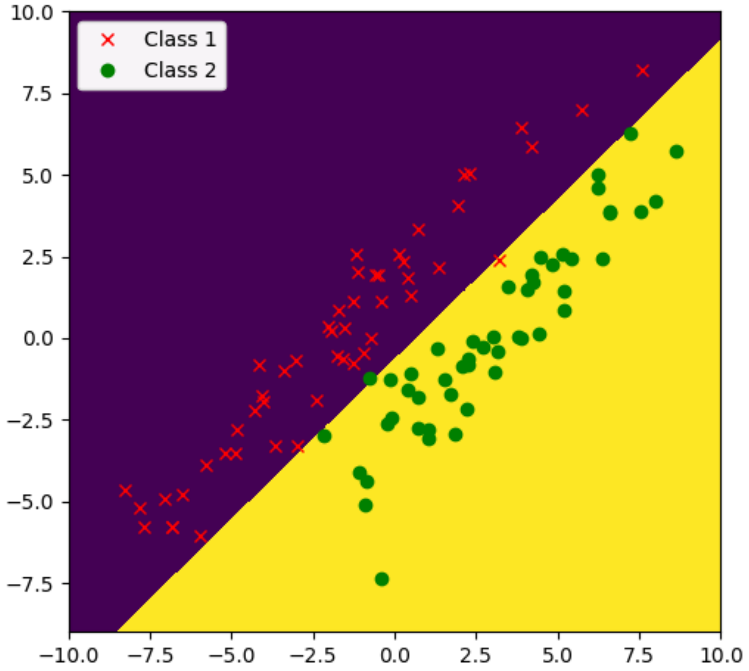
EE559   Homework Week 5   1139093874   Pin-Hsuan Lee

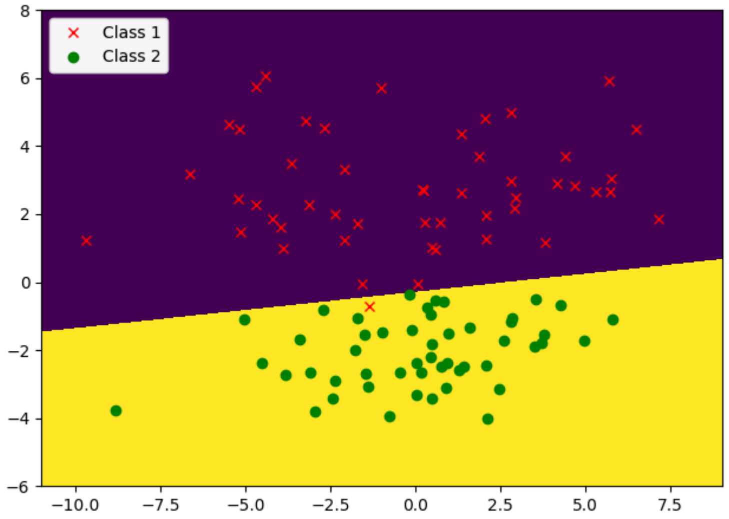
1.(a)

	Testing data error%	Training data error%	Final weight vector
Synthetic 1	2%	2%	[33.1 , -51.68856 , 52.76636]
Synthetic 2	1%	1%	[4.1 , 1.189405 , 17.949958]
Synthetic 3	0%	0%	[4.1 , -11.51453 , 9.285]

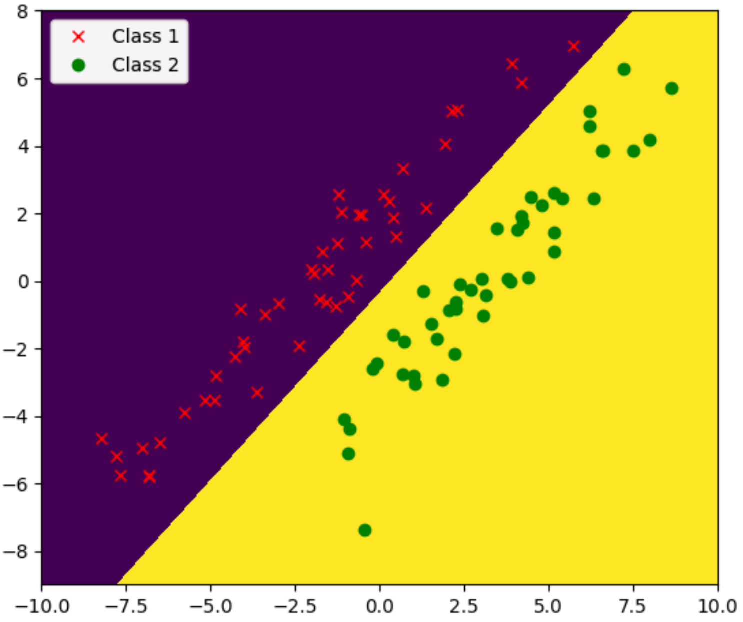
(b)



Synthetic 1_training data



Synthetic 2_training data



Synthetic 3_training data

(c)

	HW2		HW5	
	Training error%	Testing error%	Training error%	Testing error%
Synthetic1	21%	24%	2%	2%
Synthetic2	3%	4%	1%	1%

When we tried to get the classification line on HW2, we summed all of the points to get the sample mean and used the sample means of the two classes to get the line. However, when we did the classification on HW5, we modify our weight vector every time when we imported a misclassified point. In addition, we repeated many times epochs to optimize our final weight vector. Therefore, the error rate of synthetic1 and synthetic 2 in HW5 is lower than HW2.

2. (a)

$$\begin{cases} \Delta \underline{W}(i) = \underline{W}(i+1) - \underline{W}(i) \quad , \quad J(\underline{w}) = \sum_{n=1}^{N} J_n(\underline{w}) \\ \underline{W}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J_n(\underline{w}) \end{cases}$$

$$\Rightarrow \Delta \underline{W}(i) = \underline{w}(i+1) - \underline{w}(i) = -\eta(i) \nabla_{\underline{w}} J_n(\underline{w})$$

$$\Rightarrow E[\Delta \underline{W}(i)] = \sum_{n=1}^{N} P(i=n)(-\eta(i) \nabla_{\underline{w}} J_n(\underline{w})) = \frac{1}{N} \sum_{n=1}^{N} (-\eta(i) \nabla_{\underline{w}} J_n(\underline{w}))$$

(b)

$$E\left\{ \sum_{i=0}^{N-1} \Delta \underline{w}(i) \right\} = \sum_{i=0}^{N-1} E\{\Delta \underline{w}(i)\} \underset{IID}{=} N\left(\frac{1}{N} \sum_{n=1}^{N} (-\eta(i) \nabla_{\underline{w}} J_n(\underline{w})) \right) = \sum_{n=1}^{N} (-\eta(i) \nabla_{\underline{w}} J_n(\underline{w}))$$

(c)

Batch gradient descent : $\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J(\underline{w})$, $\eta(i) \geq 0$

$$\Delta \underline{W}(i) = \underline{W}(i+1) - \underline{w}(i) = -\eta(i) \nabla_{\underline{w}} J(\underline{w}) = (-\eta(i)) \nabla_{\underline{w}} \left(\sum_{n=1}^{N} J_n \right)(\underline{w}) \quad , \quad \eta(i) \geq 0$$

The result of (b) is the same with $\Delta \underline{w}(i)$ for batch gradient descent.

In batch gradient descent, we sum up all of the J_n and then differentiate the sum of J_n. And in (b), we do one single update for every points. Although they are different in the meaning of the means in the process, they have the same value.