



# 314337 類神經網路 Assignment #1 感知器測試 - 鳶尾花

---

班級：創新AI碩一

學號：111C71008

姓名：何哲平

# Outline 大綱

- A. 特徵選取 ( Select Attribute )
- B. 切分資料 ( Split Data )
- C. PLA 感知器
- D. 評估模型

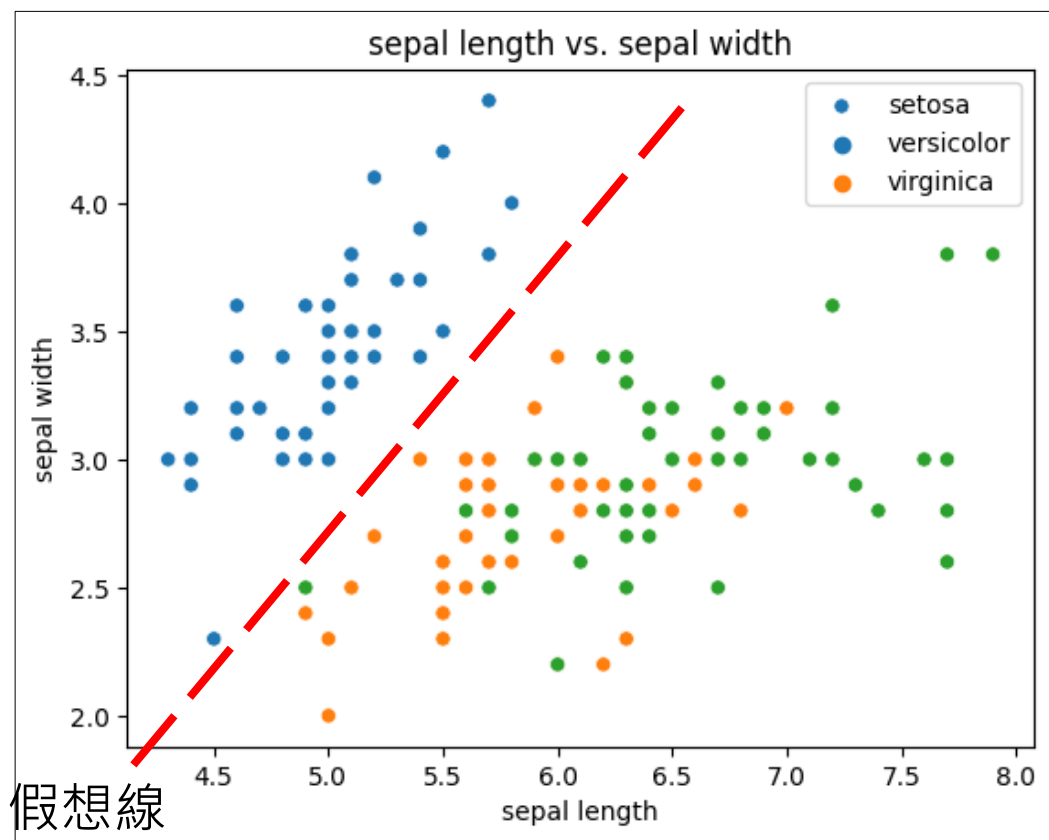


# 選取的特徵？原因？

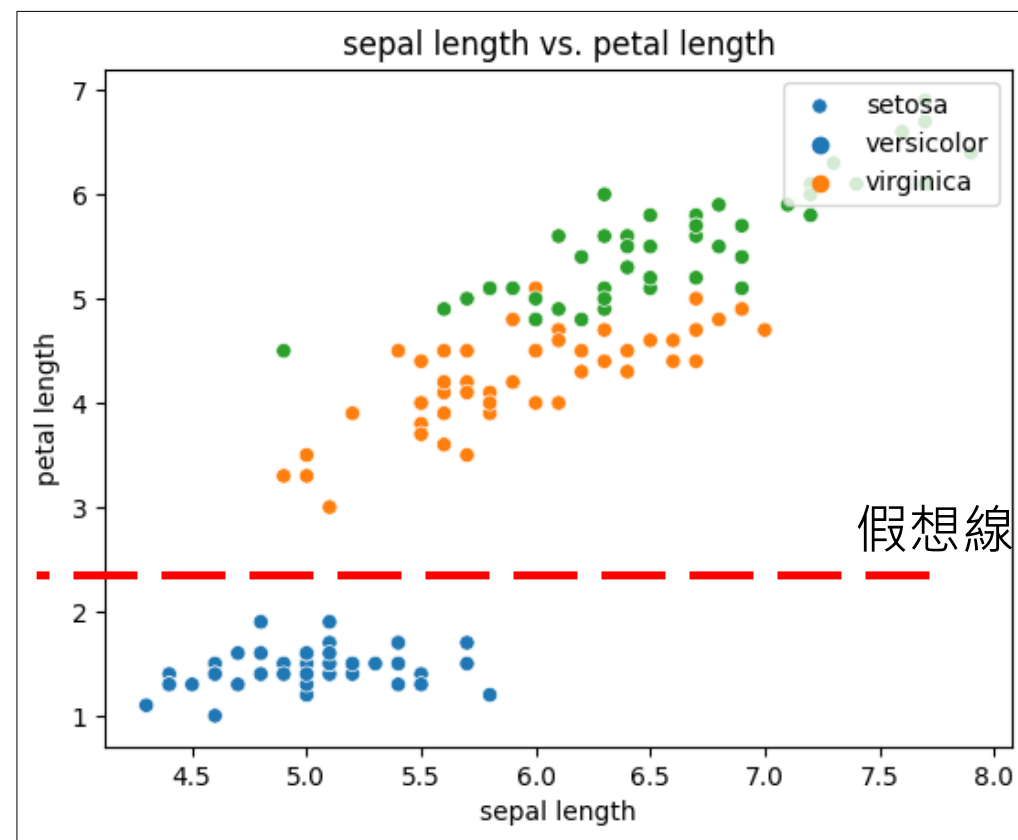
---

- ① 視覺化：  
Scatter Plot，觀察特徵之間的關聯。
- ② 量化：  
計算相關係數（Pearson Correlation），數值呈現。

# Scatter Plot 散佈圖

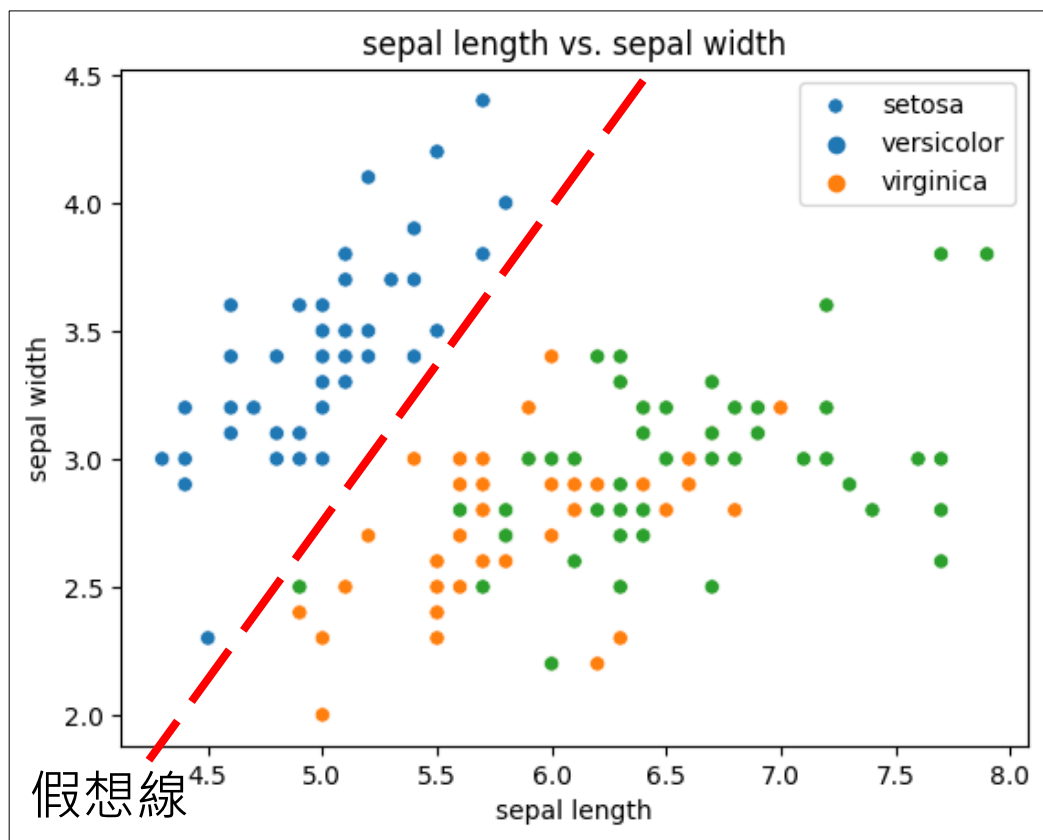


sepal length vs. sepal width

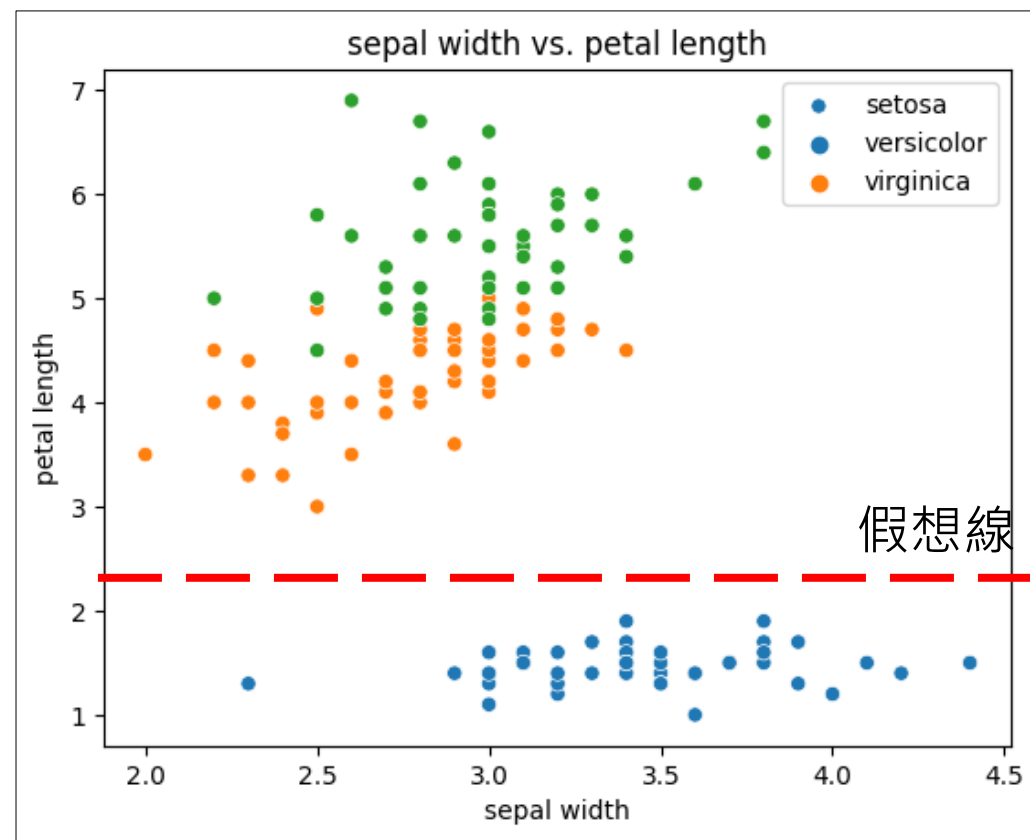


sepal length vs. petal length

# Scatter Plot 散佈圖

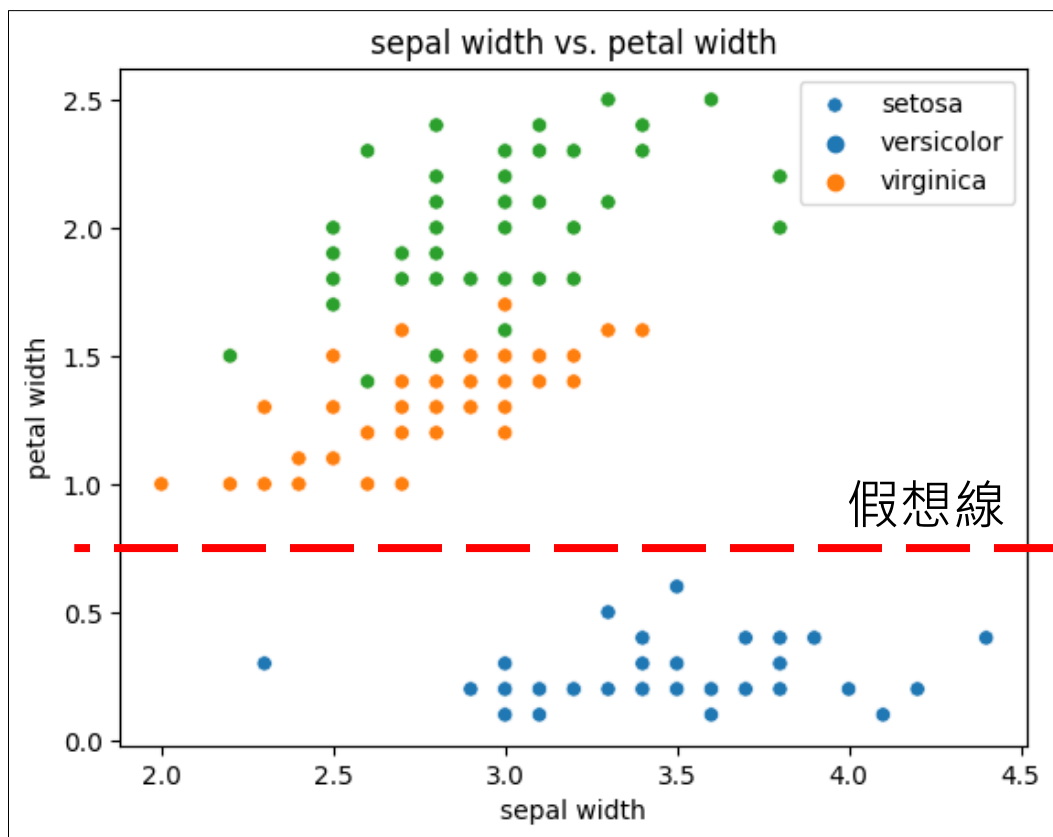


sepal length vs. petal width

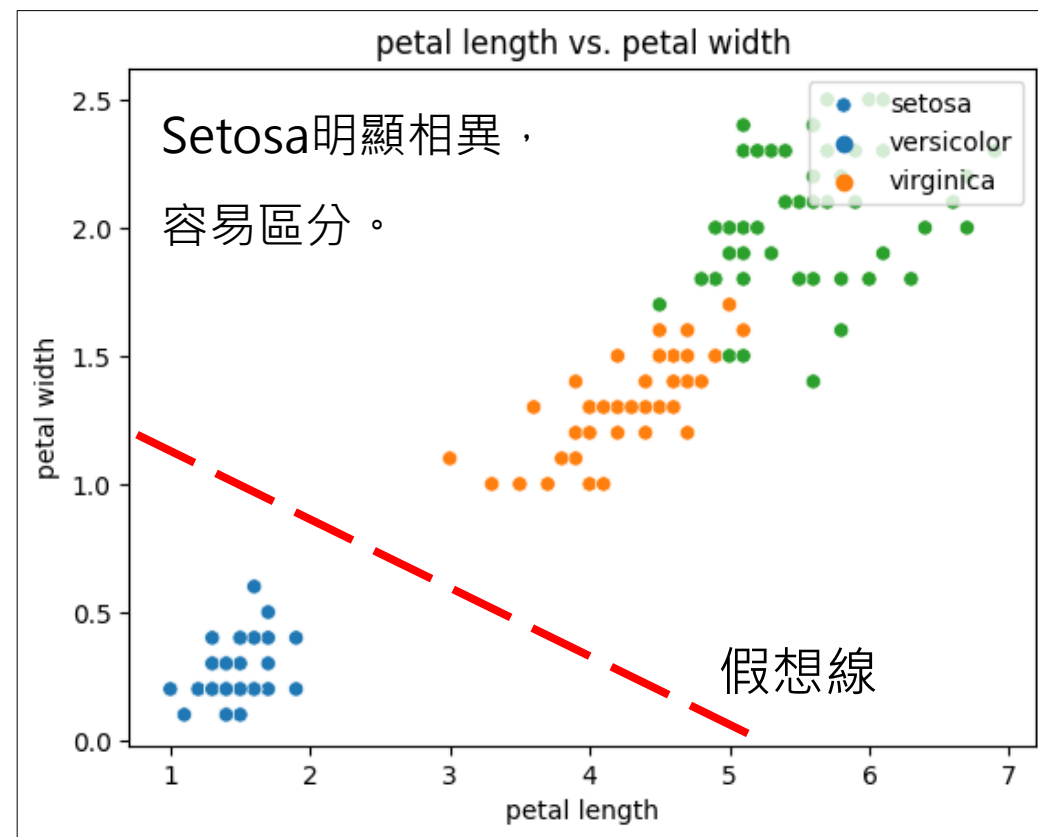


sepal width vs. petal length

# Scatter Plot 散佈圖



sepal width vs. petal width



petal length vs. petal width

# Thinking About Scatter Plot

觀察後感想：

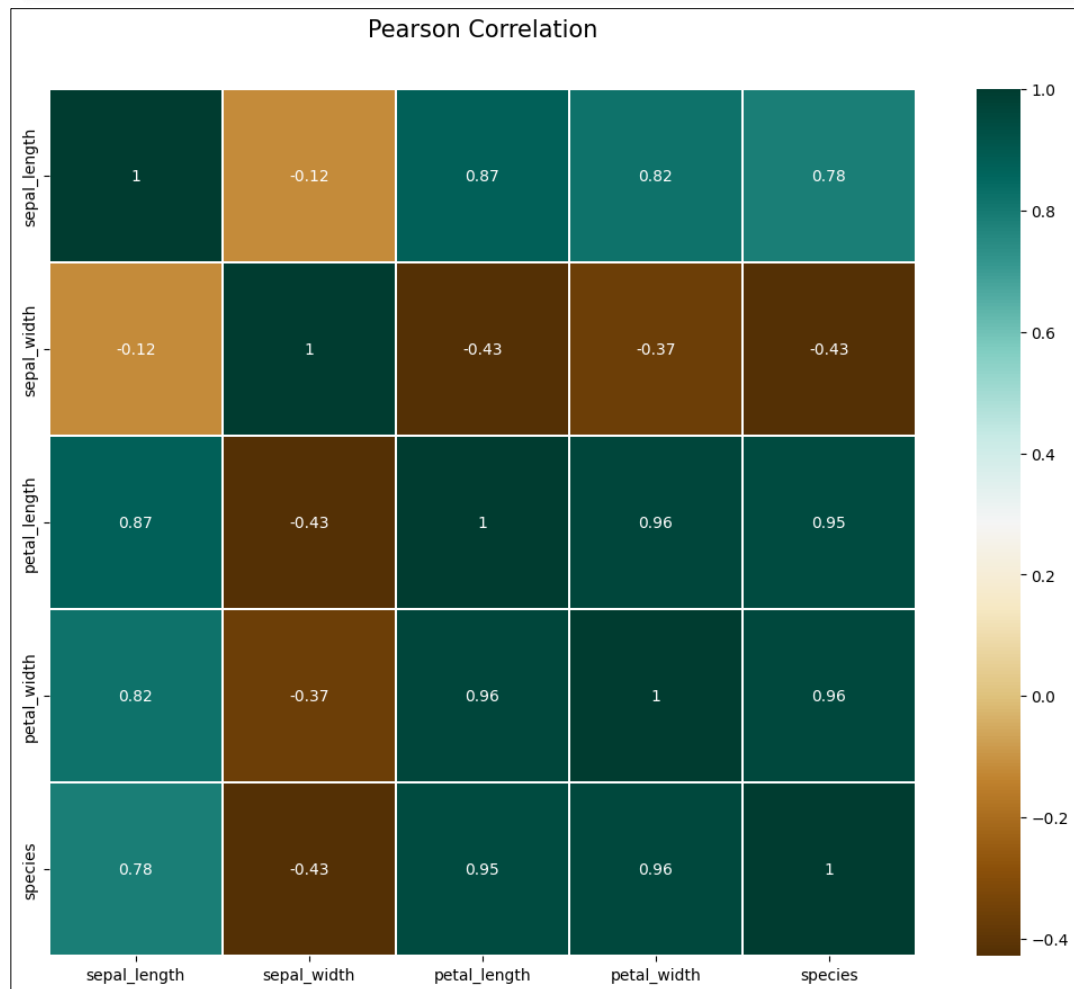
A. Setosa皆能很快分出

B. Versicolor及Virgina不容易分辨

→ 那應當如何選取？

→ 計算相關係數 ( Correlation )

# HeatMap

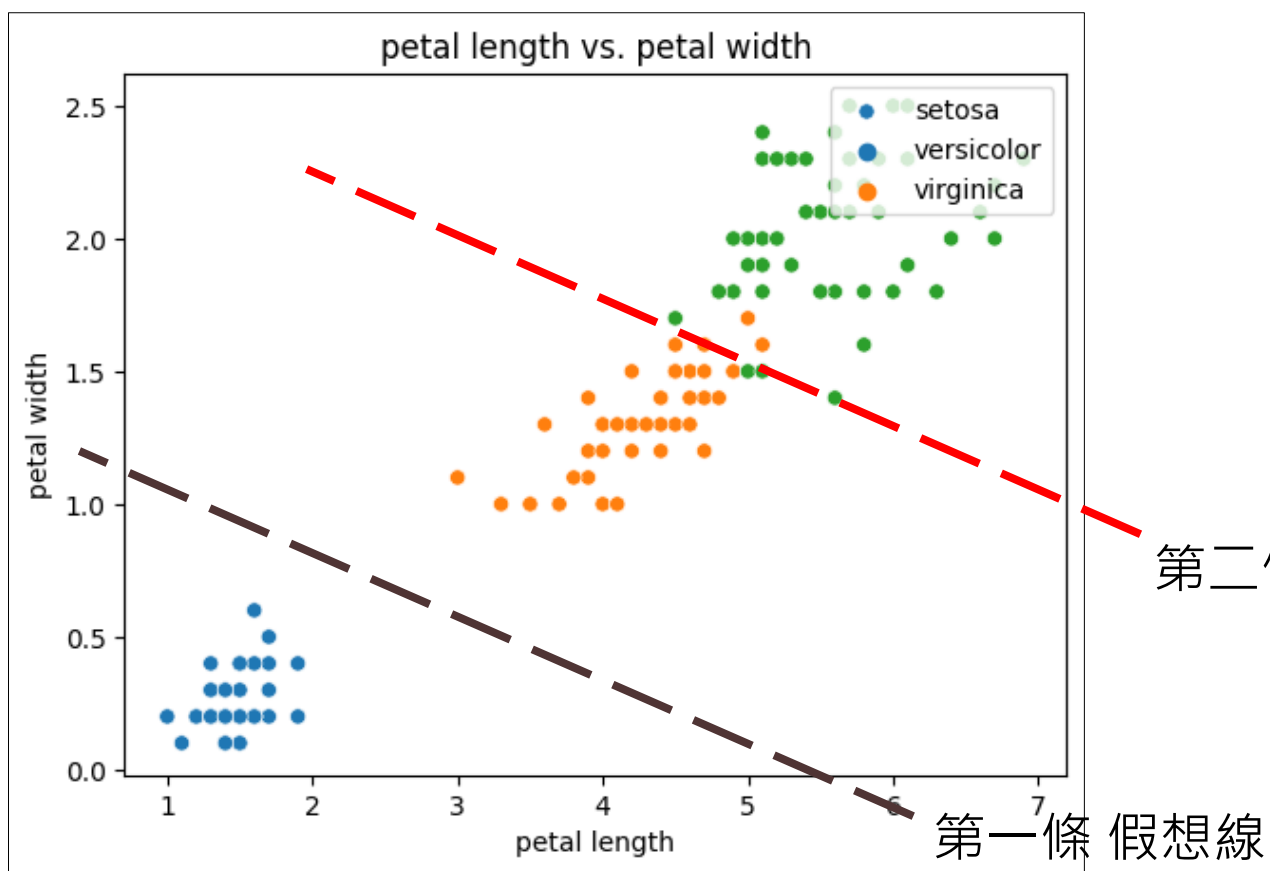


Pearson Correlation	species
sepal_length	0.783
sepal_width	-0.427
petal_length	0.949
petal_width	0.957
species	1.000

選擇Petal Length & Petal Width



# 再次檢視 petal length vs. petal width





# 切分資料

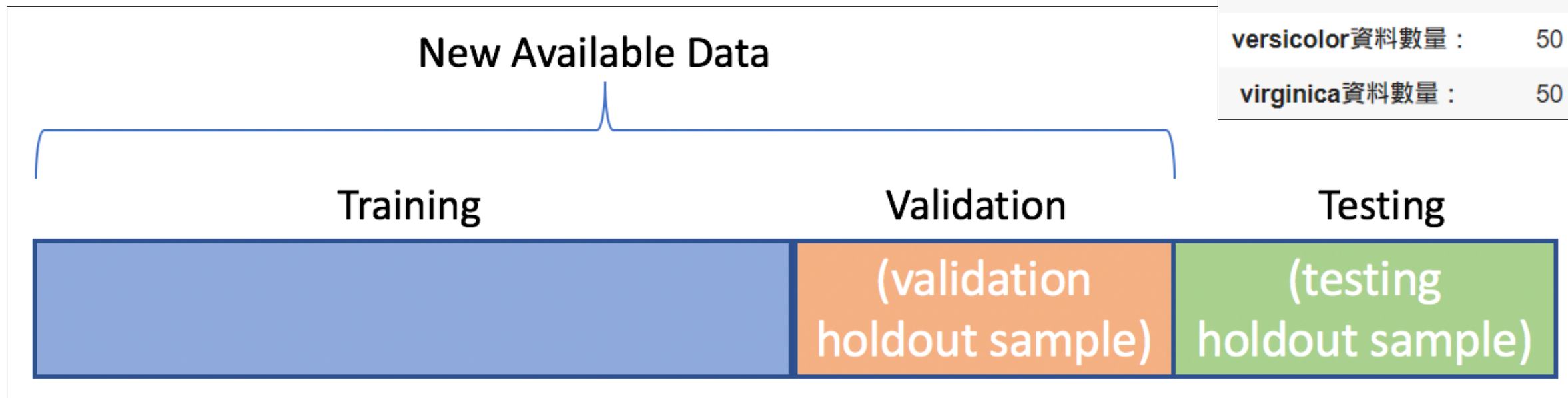
---

Train-Test-Split

# Train-Test-Split

比例

	Count
setosa資料數量 :	50
versicolor資料數量 :	50
virginica資料數量 :	50



Train\_Size : 70%  
105筆

Val : 10%  
15筆

Test\_Size : 20%  
30筆

# Train-Test-Split

## Parameters

- Train\_Size : 用多少資料訓練，這邊設定 *0.7*
- Test\_Size : 用多少資料測試，這邊設定 *0.2*
- **shuffle** : 是否要隨機抽樣 → 此資料集必須要設定 True
- **stratify** : 如何抽樣 → 依照原始 'species' 分布

```
from sklearn.model_selection import train_test_split
train, test = train_test_split(iris_data, train_size=0.7, test_size = 0.2, shuffle=True, stratify=iris_data['species'])
```

# Train-Test-Split

- **shuffle**：必須要設定True
- 若使用False，則僅會選到前面的資料

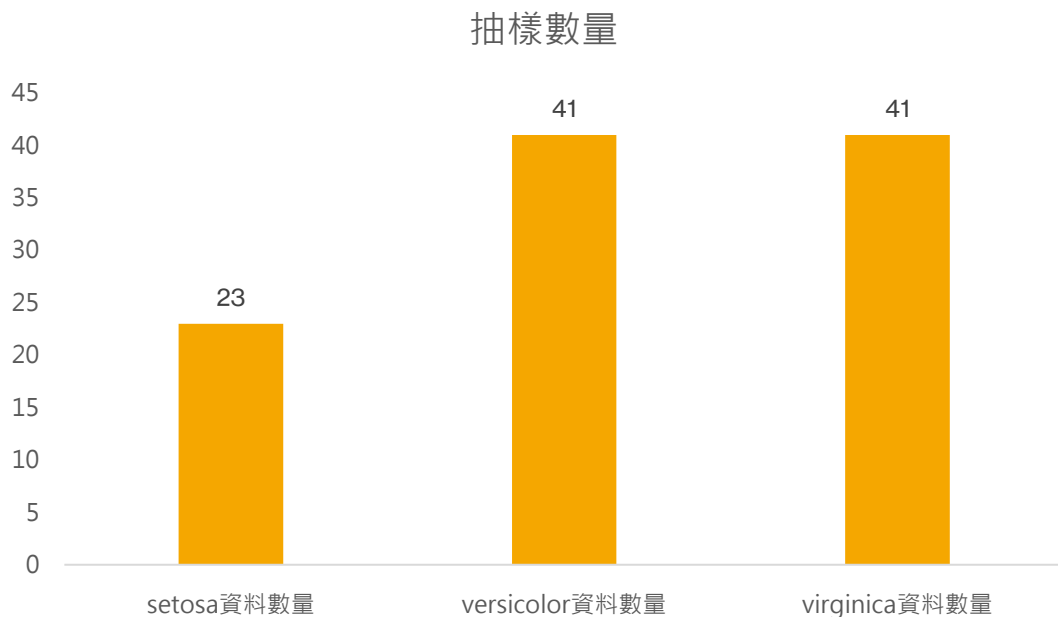
	Count
<b>setosa</b> 資料數量：	50
<b>versicolor</b> 資料數量：	50
<b>virginica</b> 資料數量：	5

	petal_length	petal_width	species
0	1.4	0.2	setosa
1	1.4	0.2	setosa
2	1.3	0.2	setosa
3	1.5	0.2	setosa
4	1.4	0.2	setosa
...	...	...	...
100	6.0	2.5	virginica
101	5.1	1.9	virginica
102	5.9	2.1	virginica
103	5.6	1.8	virginica
104	5.8	2.2	virginica

105 rows × 3 columns

# Train-Test-Split

- stratify = species 分布
- Otherwise : 抽樣不均勻，靠運氣



	Count
setosa資料數量 :	23
versicolor資料數量 :	41
virginica資料數量 :	41

# resample

或者可以利用resample，取得訓練資料

- 仍會是均勻的資料分布

```
from sklearn.utils import resample
train = resample(iris_data, n_samples=105, replace=False, stratify=iris_data['species'])
```

resample

- 處理類別不平衡問題 ( By. [CSDN](#) )
- 但因為資料集分布均勻，所以沒有選擇使用及深入研究。



# PLA 二元分類

$(W_1, W_2, W_0)$   
 $(X_1, X_2, X_0)$

---

- ① Setosa
- ② Versicolor
- ③ Virginica

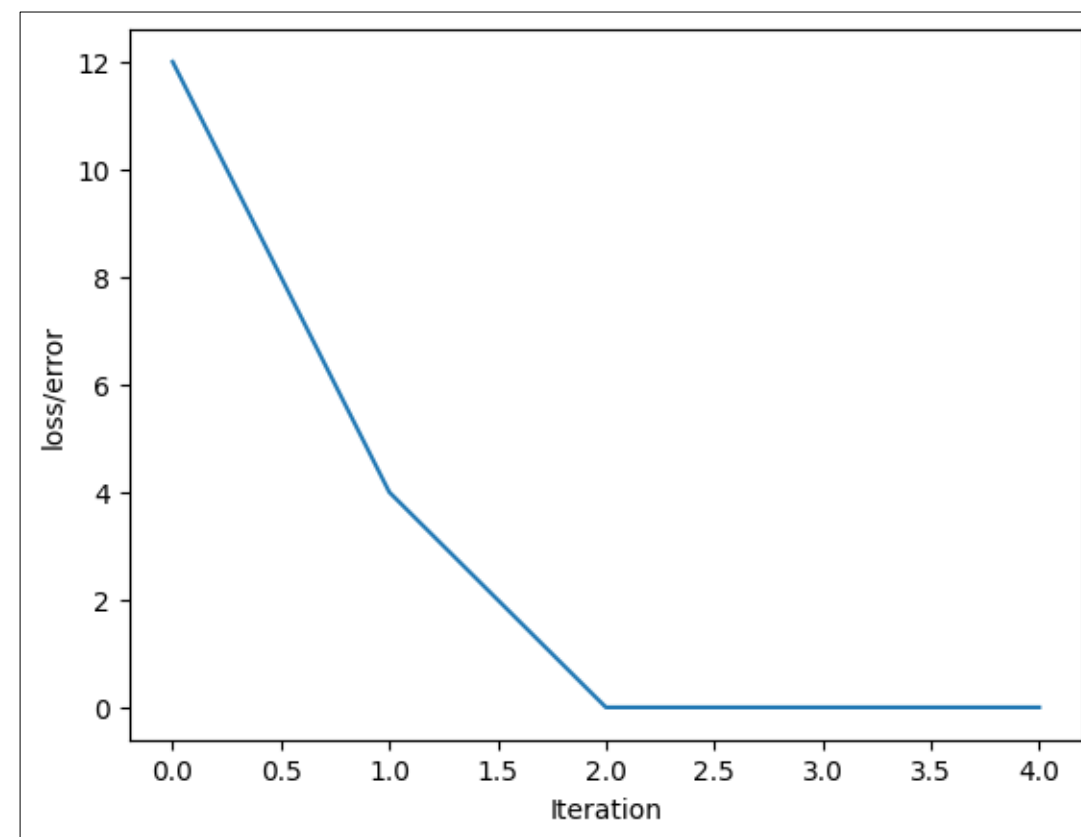


# 1. 區分Setosa

- Setosa最容易區分  
初始權重  $w_0 = \text{np.array}([0., 0., 0.])$
- Learning Rate = 0.95
- Iters = 5

結果：很快就找到PLA

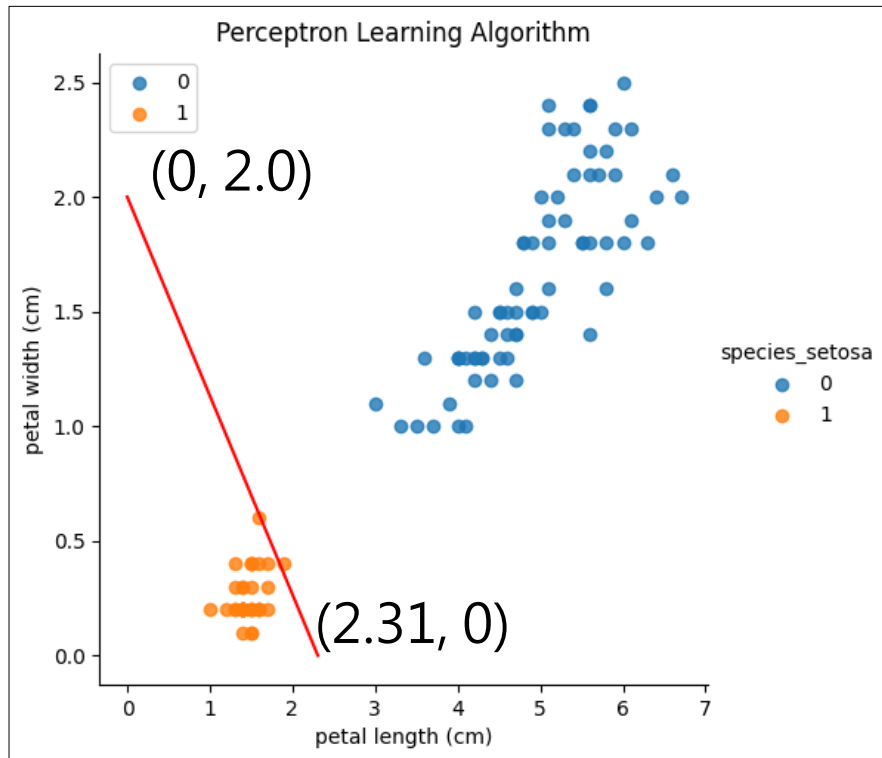
- Trained  $w_4 = \text{np.array}[-1.71, -2.945, 7.6]$



# 1. 區分Setosa

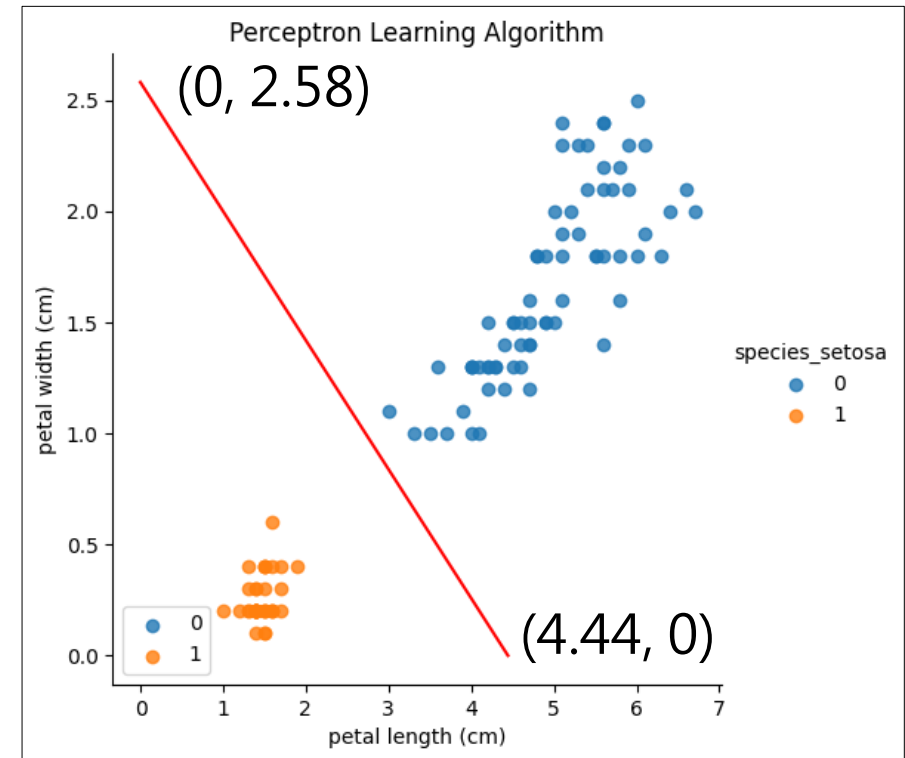
Setosa為1；其他為0

- Iters = 1 ;  $w = [-2.47 \ -2.85 \ 5.7]$



$$Y = (-2.47x_1) + (-2.85x_2) + 5.7$$

- Iters = 2 ;  $w = [-1.71 \ -2.945 \ 7.6]$



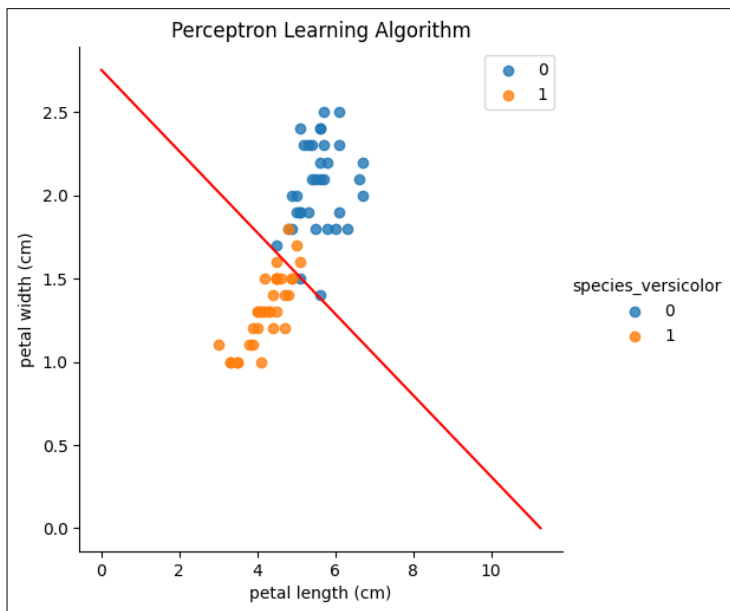
$$Y = (-1.71x_1) + (-2.945x_2) + 7.6$$

## 2. 區分Versicolor

- Versicolor與Virginica較難區分  
→ 分為三階段調整權重
- 第一階段  
 $w_0 = \text{np.array}([0., 0., 0.])$  ; Learning Rate = 0.1 ; Iters = 300
- 第二階段  
 $w_{300} = [-1.42 \ -5.81 \ 16. \ ]$  ; Learning Rate = 0.01 ; Iters = 300
- 第三階段  
 $w_{600} = [-1.872 \ -4.205 \ 16.08 \ ]$  ; Learning Rate = 0.001 ; Iters = 300
- 最終權重  $w_{900} = [-1.8486 \ -4.1222 \ 16.074 \ ]$

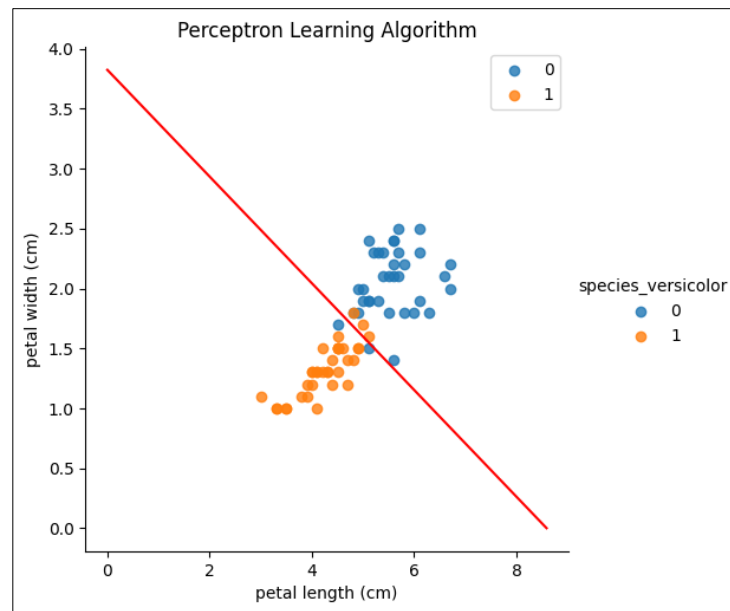
## 2. 區分Versicolor

- Iters = 300
- $w = [-1.42 \ -5.81 \ 16. ]$
- $Y = (-1.42x_1) + (-5.81x_2) + 16.0$



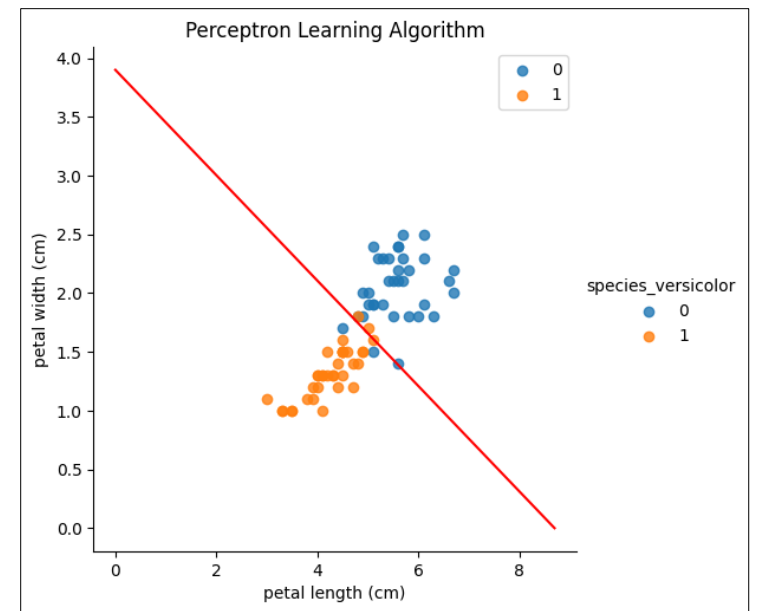
(11.27, 0) ; (0, 2.75)

- Iters = 600
- $W = [-1.872 \ -4.205 \ 16.08 ]$
- $Y = (-1.87x_1) + (-4.21x_2) + 16.08$



(8.59, 0) ; (0, 3.82)

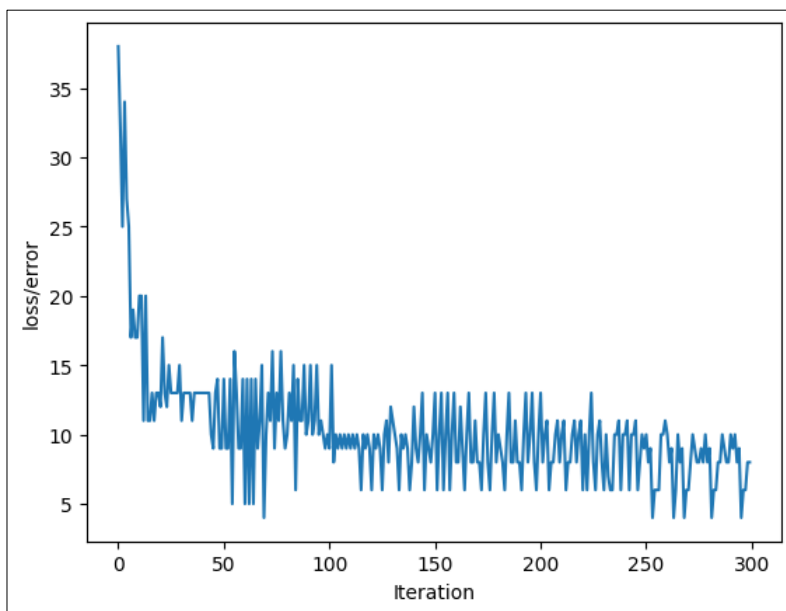
- Iters = 900
- $w = [-1.8486 \ -4.1222 \ 16.074 ]$
- $Y = (-1.85x_1) + (-4.12x_2) + 16.07$



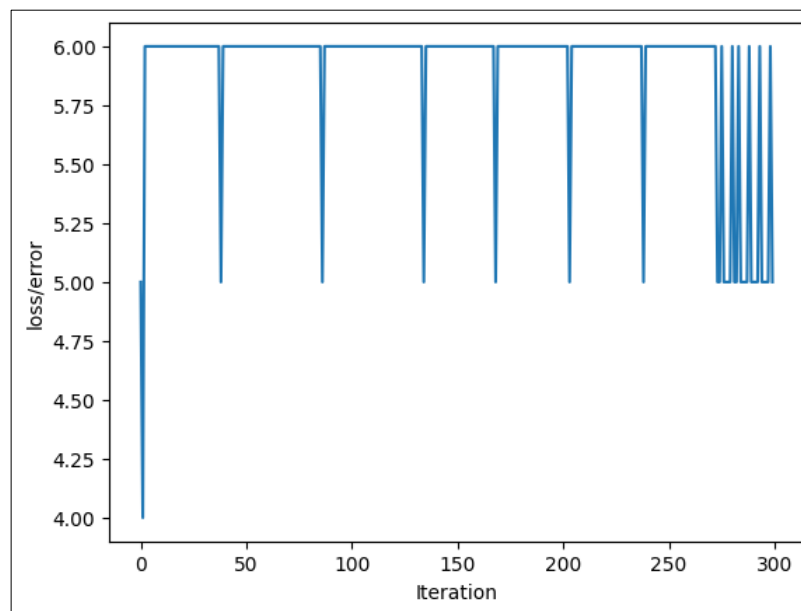
(8.7, 0) ; (0, 3.9)

## 2. 區分Versicolor

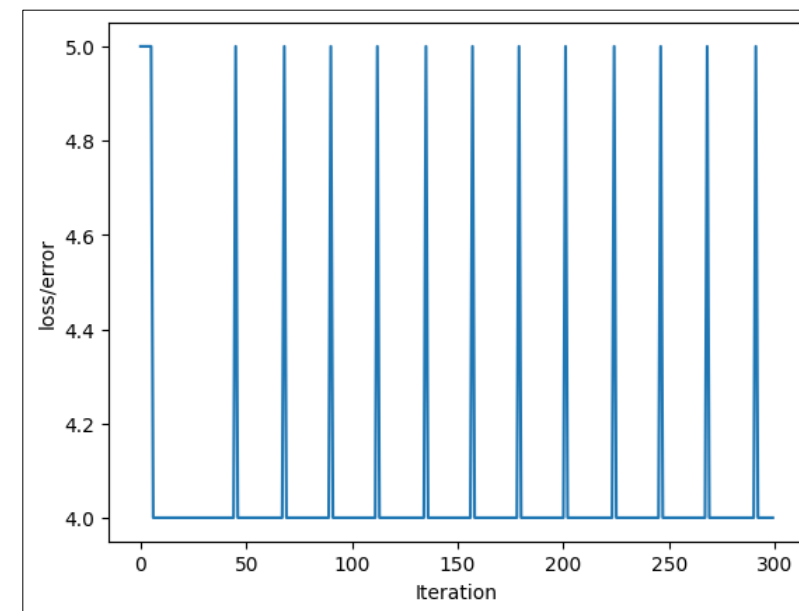
- 第一批次訓練



- 第二批次訓練

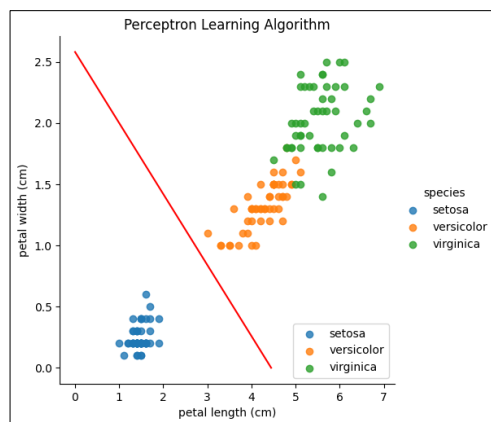


- 第三批次訓練

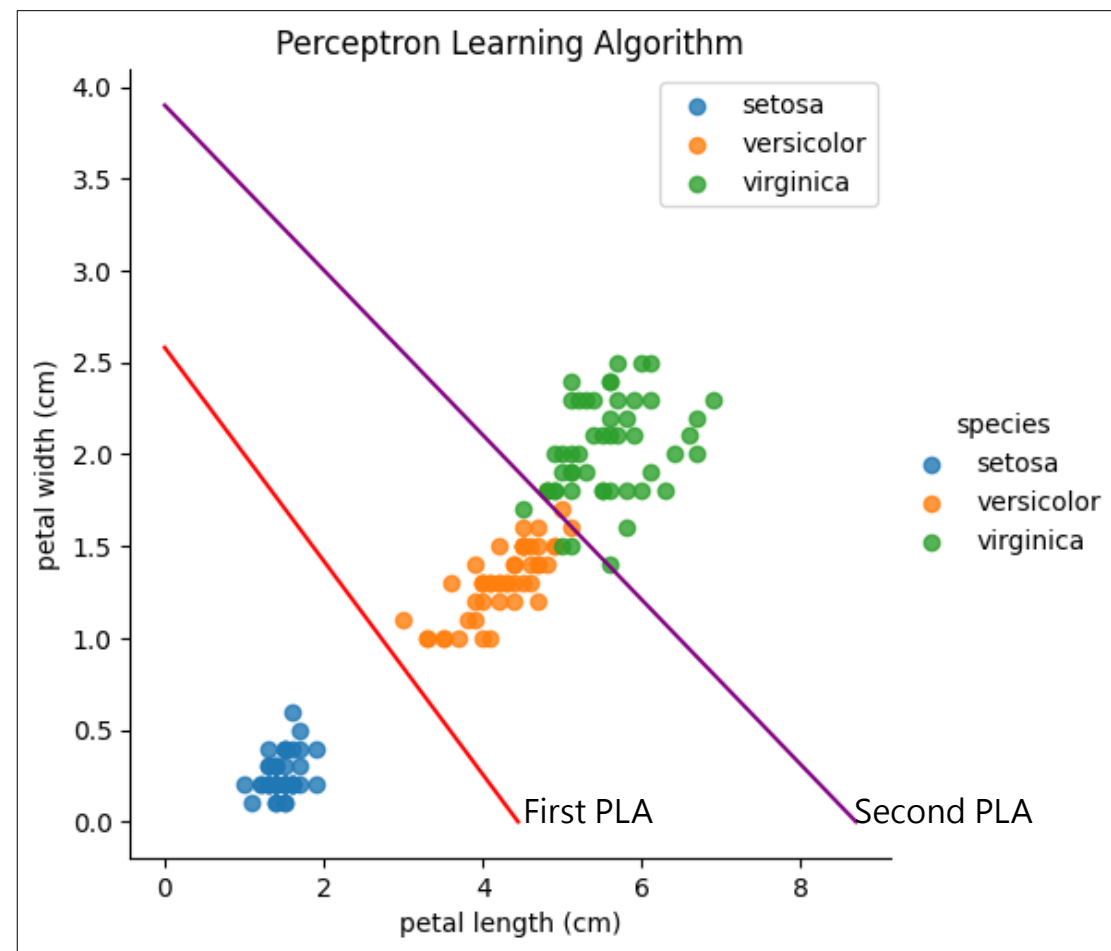
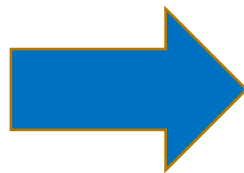
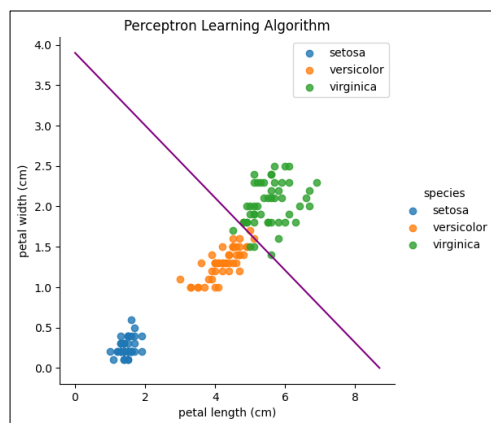


# 合併圖表

## 1. 區分Setosa



## 2. 區分Versicolor



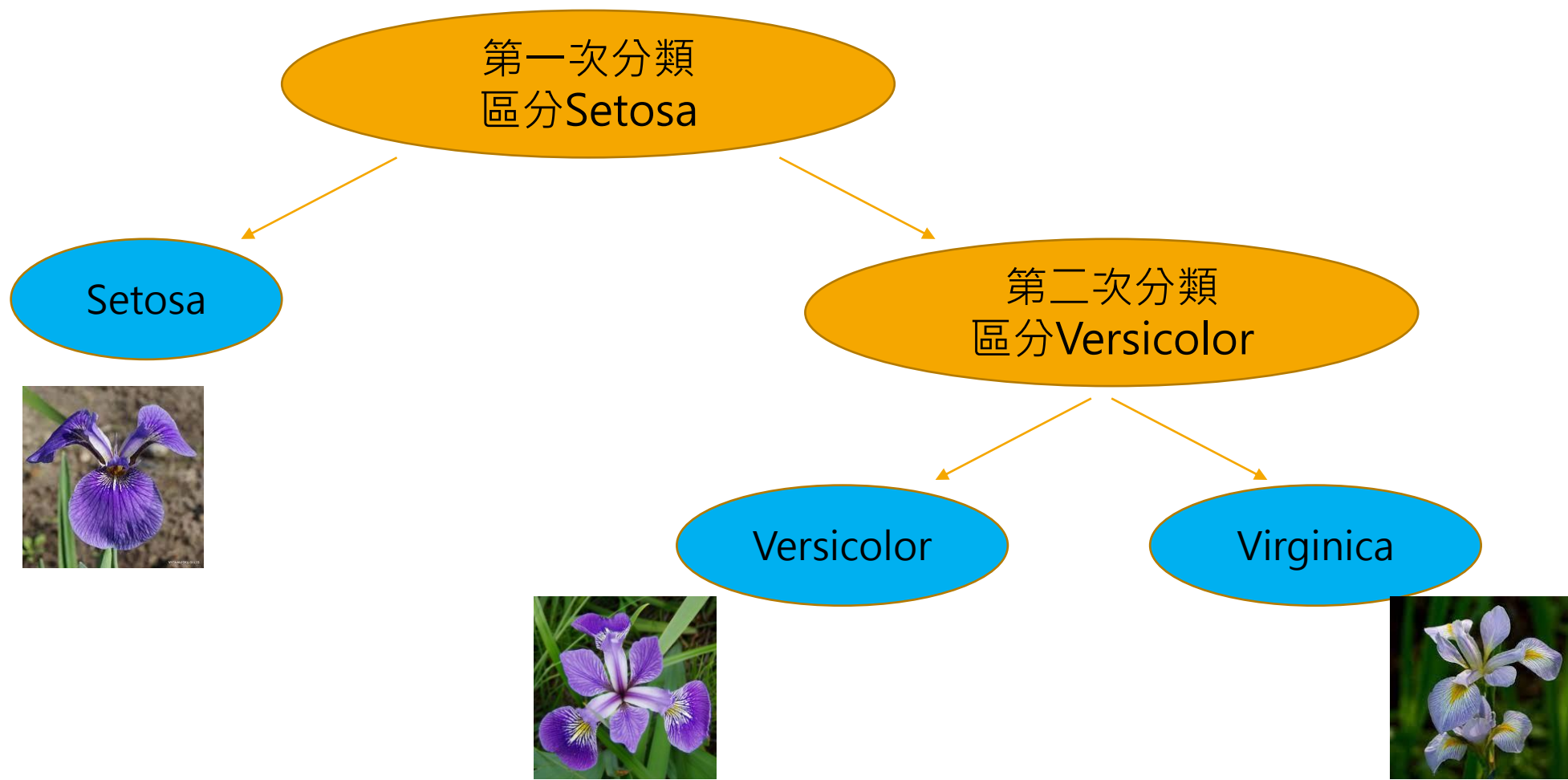


# 評估模型

---

- ① Confusion Matrix
- ② Accuracy VS. Precision VS. Recall

# 評估流程





# 評估指標

- 正確率 Accuracy：在所有情況中，正確判斷真假的比例。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- 精確率 Precision：判斷為真的情況下，有多少是真的真。

$$Precision = \frac{TP}{TP + FP}$$

- 召回率 Recall：為真的情況下，有多少被正確判斷出來。

$$Recall = \frac{TP}{TP + FN}$$

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

# First PLA Validation

區分Setosa

- 答對15個；答錯0個
  - 正確率：100.0%
  - Precision：100.0%
  - Recall：100.0%

Label  
( 真實 )

	Predict ( 預測 )	
	不是Setosa	Setosa
不是Setosa	TN 10	FP 0
Setosa	FN 0	TP 5

# Second PLA Validation

區分versicolor

- 答對9個；答錯1.0個
- 判斷錯誤的資料：

petal length (cm)	petal width (cm)	label	predict
4.5	1.7	0.0	1.0

Label  
( 真實 )

Predict ( 預測 )

	不是Versicolor	Versicolor
不是Versicolor	TN 4	FP 1
Versicolor	FN 0	TP 5

- 正確率：90.0%
- Precision：83.33%
- Recall：100.0%

# Validation

		預測		
	Validation	Setosa	Versicolor	Virginica
真實	Setosa	5	0	0
	Versicolor	X	5	0
	Virginica	X	1	4

# First PLA Test

區分Setosa

- 答對30個；答錯0個
  - 正確率：100.0%
  - Precision：100.0%
  - Recall：100.0%

Label  
( 真實 )

Predict ( 預測 )

	Predict ( 預測 )	
	不是Setosa	Setosa
不是Setosa	TN 20	FP 0
Setosa	FN 0	TP 10

# Second PLA Test

區分versicolor

- 答對19個；答錯1.0個
- 判斷錯誤的資料：

petal length (cm)	petal width (cm)	label	predict
5.0	1.7	1.0	0.0

Label  
( 真實 )

Predict ( 預測 )

	不是Versicolor	Versicolor
不是Versicolor	TN 10	FP 0
Versicolor	FN 1	TP 9

- 正確率：95.0%
- Precision：100.0%
- Recall：90.0%

# Test

		預測		
	Validation	Setosa	Versicolor	Virginica
真實	Setosa	10	0	0
	Versicolor	X	9	1
	Virginica	X	0	10

# 關於 Second PLA

區分versicolor

- 驗證所有Versicolor & Virginica資料
- 答對95個；答錯5.0個
- 判斷錯誤的資料：

	petal length (cm)	petal width (cm)	label	predict
0	4.8	1.8	1.0	0.0
1	5.0	1.7	1.0	0.0
2	4.5	1.7	0.0	1.0
3	5.0	1.5	0.0	1.0
4	5.1	1.5	0.0	1.0

Label  
( 真實 )

Predict ( 預測 )

	不是Versicolor	Versicolor
不是Versicolor	TN 47	FP 3
Versicolor	FN 2	TP 48

- 正確率：95.0%
- Precision：94.12%
- Recall：96.0%





# 簡報完畢

---

讚美感謝主

榮耀歸神