Q 1 : 在每個分類器使用下(1-NN, 3-NN, LDA)，以及採用不同特徵組合的條件下，包括(1)四個特徵皆採用、(2)PL 與 PW 兩個特徵、(3)SL 與 SW 兩個特徵，分別計算出三個(C1 vs. C2、C1 vs. C3、C2 vs. C3)二元分類(binary classification)的分類準確率。

C1 : setosa C2 : versicolor C3 : virginica
pl : petal length pw : petal width sl : sepal length sw : sepal width

1NN classifier

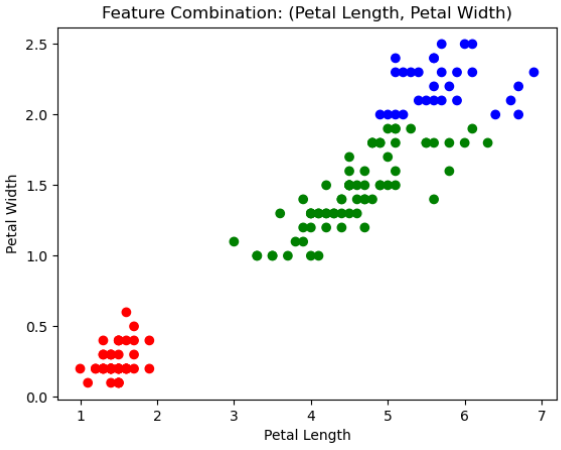
	CR1 (train from 1~25)	CR2 (train from 26~50)	2 fold cv accuracy
1-NN all features			
C1 vs. C2	1.0000	1.0000	1.0000
C1 vs. C3	1.0000	1.0000	1.0000
C2 vs. C3	0.9200	0.9200	0.9200
1-NN pl_pw			
C1 vs. C2	1.0000	1.0000	1.0000
C1 vs. C3	1.0000	1.0000	1.0000
C2 vs. C3	0.9000	0.9600	0.9300
1-NN sl_sw			
C1 vs. C2	0.9800	1.0000	0.9900
C1 vs. C3	0.9800	0.9800	0.9800
C2 vs. C3	0.5200	0.6000	0.5600

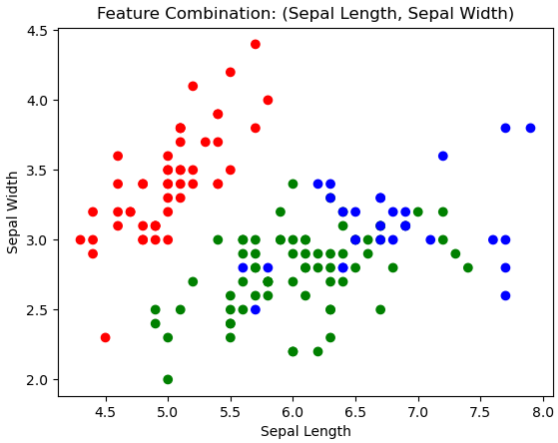
3NN classifier

	CR1 (train from 1~25)	CR2 (train from 26~50)	2 fold cv accuracy
3-NN all features			
C1 vs. C2	1.0000	1.0000	1.0000
C1 vs. C3	1.0000	1.0000	1.0000
C2 vs. C3	0.8800	0.9400	0.9100
3-NN pl_pw			
C1 vs. C2	1.0000	1.0000	1.0000
C1 vs. C3	1.0000	1.0000	1.0000
C2 vs. C3	0.9200	0.9400	0.9300
3-NN sl_sw			
C1 vs. C2	0.9800	1.0000	0.9900
C1 vs. C3	1.0000	0.9800	0.9900
C2 vs. C3	0.6200	0.6800	0.6500

LDA classifier

	CR1 (train from 1~25)	CR2 (train from 26~50)	2 fold cv accuracy
LDA all features			
C1 vs. C2	1.0000	1.0000	1.0000
C1 vs. C3	1.0000	1.0000	1.0000
C2 vs. C3	0.9400	0.9400	0.9400
LDA pl_pw			
C1 vs. C2	1.0000	1.0000	1.0000
C1 vs. C3	1.0000	1.0000	1.0000
C2 vs. C3	0.9400	0.9400	0.9400
LDA sl_sw			
C1 vs. C2	0.9800	1.0000	0.9900
C1 vs. C3	1.0000	0.9800	0.9900
C2 vs. C3	0.7400	0.7000	0.7200



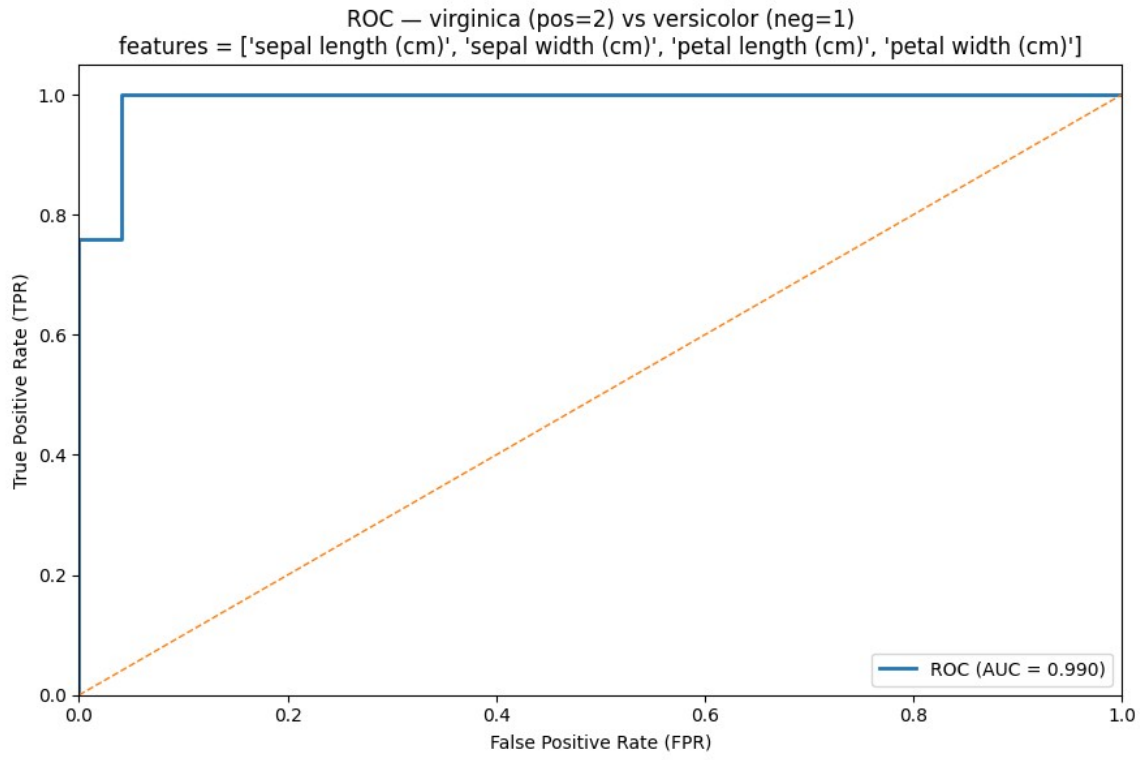


從散佈圖可以看出 petal length 與 petal width 對 3 個 class 是線性可分的, 因此在 LDA 上的表現相較於 1-NN, 3-NN 還要好, 而當特徵選擇為 sepal length 與 sepal width 時除了 setosa 與其於兩者是線性可分之外, versicolor 與 virginica 有高度重疊, 因此兩種分類結果都不盡理想

Q2. 在每個分類器的使用條件下(1-NN, 3-NN, LDA)，計算出三類別分類(3-class classification)的準確率，其中 LDA 採用 one-against-one 方法及 voting 策略來實現多類別分類。

	CR1 (train from 1~25)	CR2 (train from 26~50)	2 fold cv accuracy
1-NN			
all features	0.9467	0.9467	0.9467
petal length and width	0.9333	0.9733	0.9533
sepal length and width	0.6933	0.7333	0.7133
3NN			
all features	0.9200	0.9600	0.9200
petal length and width	0.9467	0.9600	0.9533
sepal length and width	0.7333	0.7600	0.7467
LDA (c1 = c2 = 1)			
all features	0.9600	0.9600	0.9600
petal length and width	0.9600	0.9600	0.9600
sepal length and width	0.8133	0.8000	0.8067

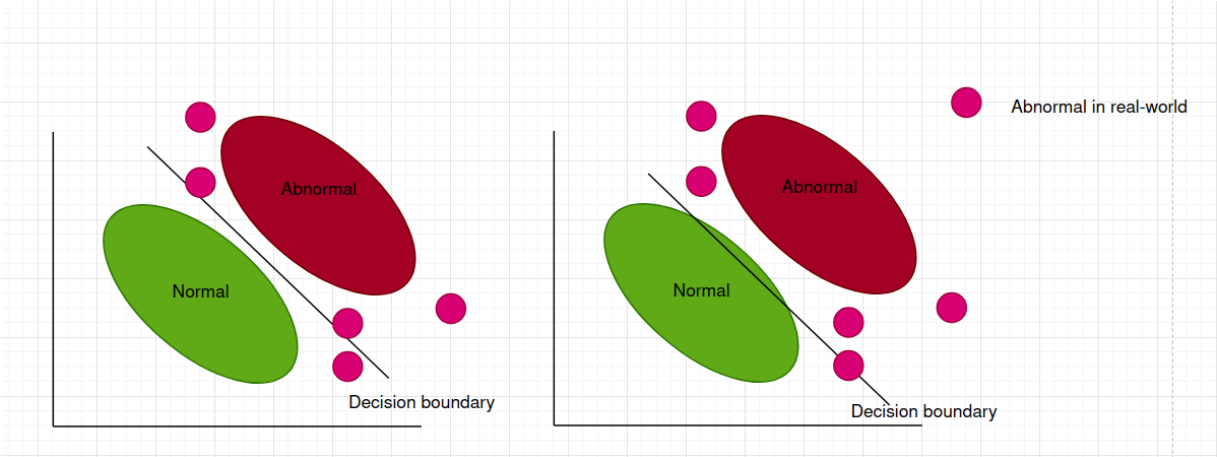
Q3. 假若 class 3 為 positive class (代表異常情況)，class 2 為 negative class (代表正常情況)，並採用 LDA classifier 及四個特徵值。以調整 LDA 參數的方式來繪出 ROC (receiver operating characteristic)曲線，並計算出 AUC (area under the ROC curve)的值。ROC 曲線的 y 軸為 TPR (true positive rate)，x 軸為 FPR (false positive rate)。



AUC = 0.990 #

Q4. 如果你是一位 AI 工程師，從(C)這個簡單的例子中，你該怎麼從 LDA 模型去設計，才能幫公司確切的開發出製程或是設備的異常診斷模組? 簡單合理的描述你的設計想法。

在實際現場當中, 我們寧願遇到假警報也不要真的有警報發生了而沒有異常通知, 根據要求, class 3 為異常狀況, 因此在訓練 LDA 時應該要把 class 3 錯誤分類時的懲罰權重設置較大的數值, 模型在訓練時才會由以下左圖變為右圖情形 (有真實警報以及類似警報但實際不是警報時也會通知現場人員)



同時此模型也可能更好的泛化於真實世界遇到的異常狀況 (如粉色點), 達到良好異常檢測模型