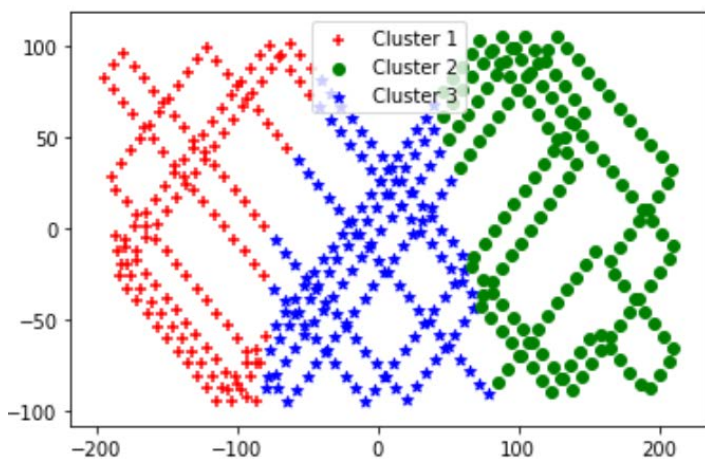


前言:

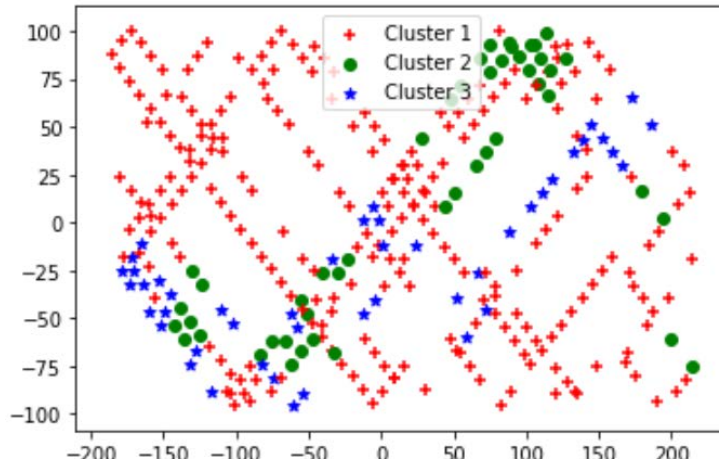
上禮拜上課時有提到，大家可以利用助教們給的code 對機器學習模型的結果做視覺化呈現，剛好這幾週教的KNN和K Means在視覺化的變化非常顯著，如果想要在寫" Discussion" 時強調這件事，可以參考我的作法，我是使用奕翔助教的code下去修改的。不過我的Demo Discussion也只是給大家參考，不是說大家都一定要照著做才能得分，但適當的使用" 量化數據" 、" 統計圖" 、" 視覺化" ，是讓解釋一件事情變得簡單且有說服力的好方法(也是會讓教授們看到各位的報告時眼睛一亮的地方)，希望大家可以學起來，以後一定用得到。

Discussion

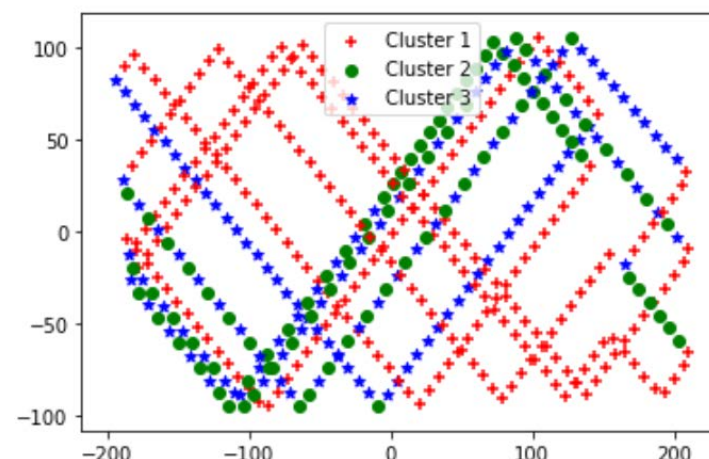
問題一: 以不同的機器學習演算法處理同一筆data會有什麼差異?



K Means



KNN(test size:0.7)

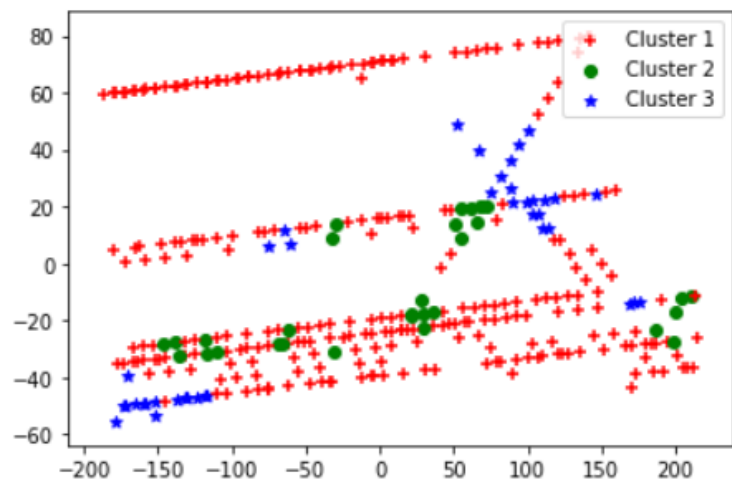


Actual Label

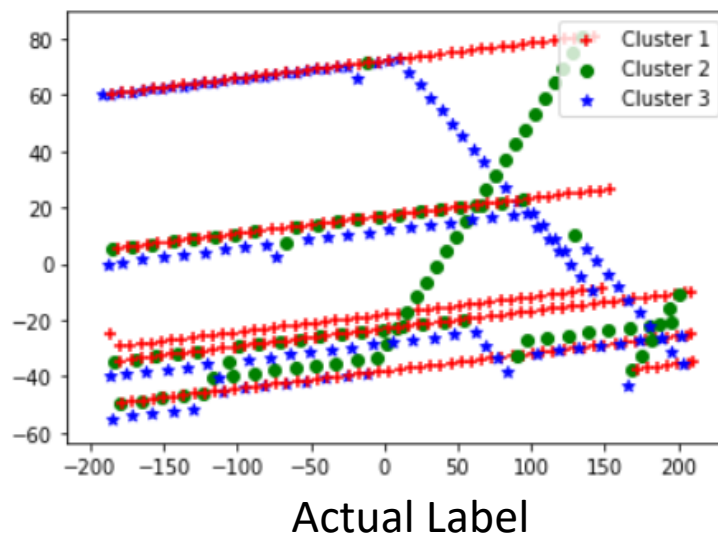
Accuracy:0.99

這三張圖，是我將所選取的特徵參數以PCA的方式投影到二維畫面所畫出的點，我分別以紅十字、綠圓圈、藍星星將KNN模型預測答案為comment = 0("NONE")、comment = 1("MOVE_LEFT")、comment = 2 ("MOVE_RIGHT" 和其他狀況)的點上色，K Means演算法為非監督式學習的分群演算法，因此只會按照分群邏輯分成三群，但與答案無關，從視覺化顯示的結果可以看出，其正確性與Actual Label 有明顯的差異，而KNN演算法為監督式學習演算法，學習的過程與我們所選的答案有關，正確率較高，但所選的特徵數量多寡、是否具有代表性，也會影響到模型學習的成效，這件事我會於問題二進行討論。

問題二: 減少特徵或置換成更有代表性的特徵之後之後，表現上有什麼改變呢？

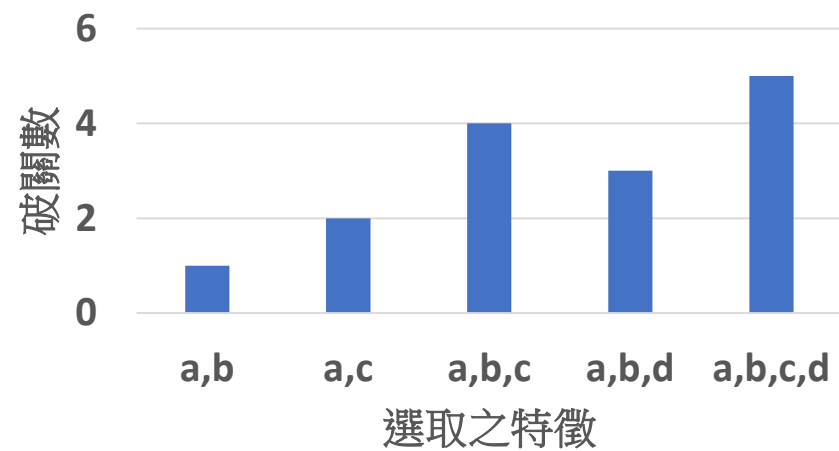


KNN 僅以球Y座標和板子X座標
作為特徵(test size:0.7)
Accuracy:0.52



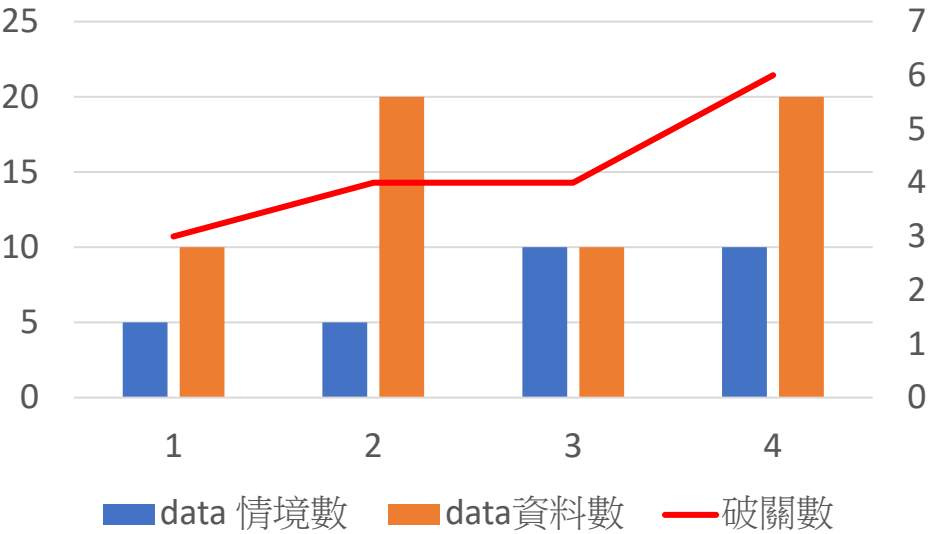
上一頁KNN模型選取的特徵有球的X座標、球的Y座標、板子的X座標三個，Accuracy為0.99，而本頁的KNN模型只選取球Y座標和板子X座標作為特徵，可以看到Accuracy顯著下降，從視覺化呈現的圖形中，也可看到，與Actual Label相比，預測錯誤的點數激增，可見被捨去的特徵(球X座標)在此是一個極為重要的特徵。

問題三: 分別選取a,b,c,d 作為特徵，實際投入遊戲之後的效果



文字說明:

問題三: 訓練data的多寡、data多元性是否足夠，是否影響訓練模型的成效?



文字說明:

data情境: 以向一側發球破一個關卡定義為一個情境 ex: 向左發_破EASY_1、向右發_破MEDIUM_2 共為兩個情境