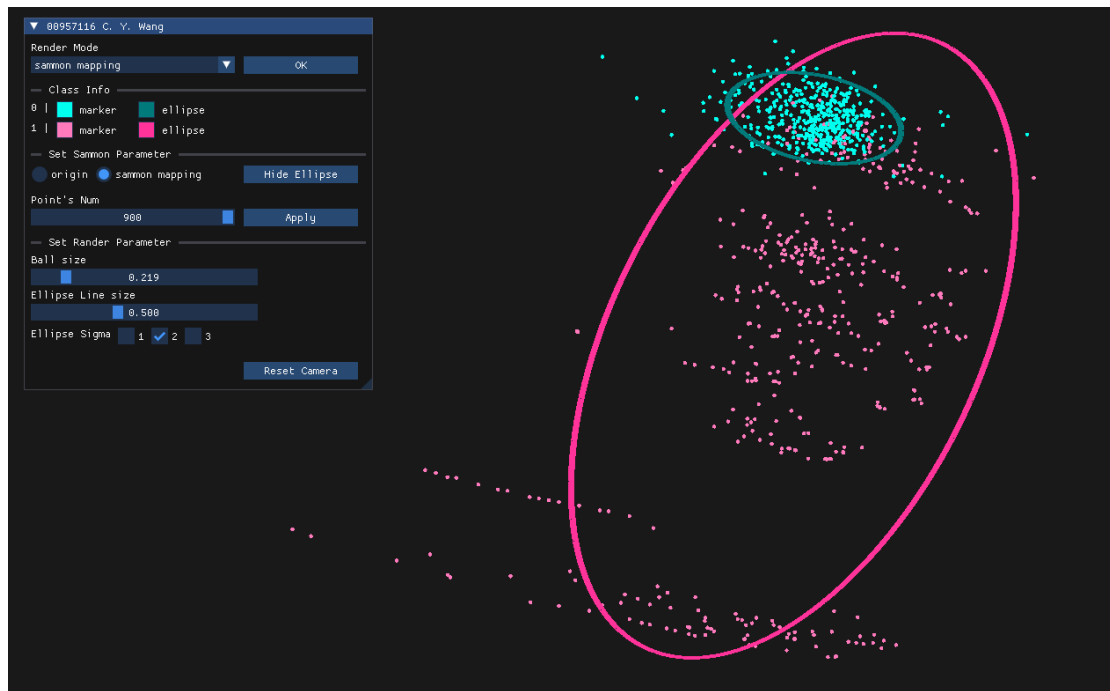


Computer Project4: High Dimensional Data Visualization

00957116 王嘉羽

1 介紹

這次作業是要實作 Sammon Mapping，將高維度的資料以二維的方式顯示。此外，還使用 PCA Method 找出二維資料的主成分分布以及利用特徵值方法找出包覆資料點的橢圓半長短軸，並將其畫出。



2 實作方法 – Sammon Mapping

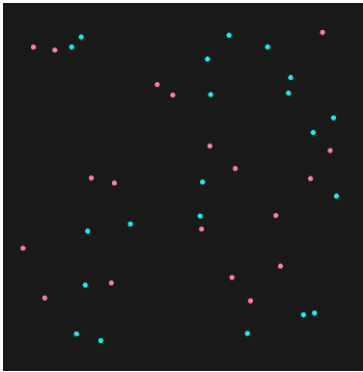
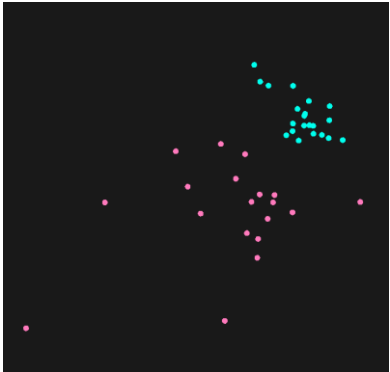
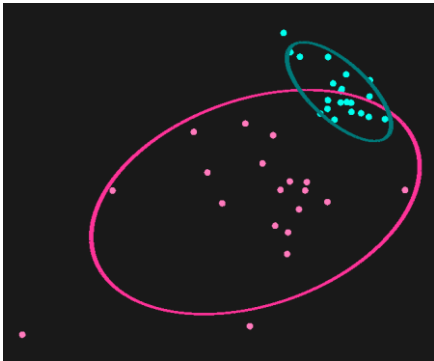
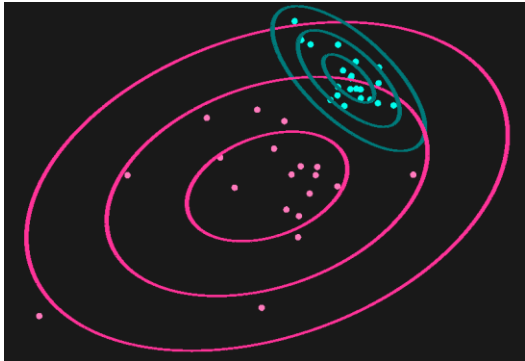
- 隨機在二維空間佈點
- 計算高維度(原本資料)和二維度的距離比例以及兩點的向量
- 如果太近，就往反向移動；太遠，則靠近，持續迭代至收斂為止
- 計算並利用同群點的斜方差矩陣、特徵值及特徵向量繪製橢圓
- 使用幾何著色器，將原本的點繪製成圓型
- 根據螢幕比例繪製圓型，以達到拉伸畫面不會造成變形
- 可以透過 GUI 調整資料點個數、繪製橢圓的 sigma、大小、顏色

3 新增的功能

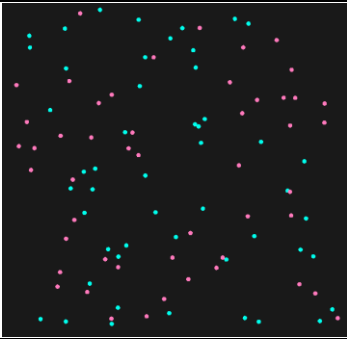
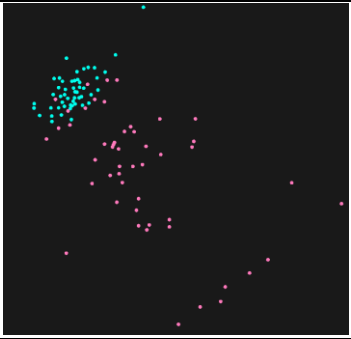
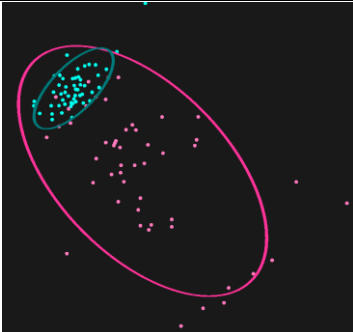
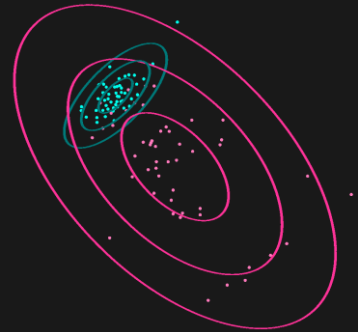
- Set Sammon Parameter: 選擇顯示原始點還是經過 mapping 的點
- Show/ Hide Ellipse: 顯示/隱藏橢圓
- Point's Num: 參與計算的資料筆
- Ball size: 點的大小
- Ellipse Line size: 橢圓線的粗度
- Ellipse Sigma: 要圈多少信賴值數量的點(1: 68.27%、2: 95.45%、3: 99.73%)

4 成果展示

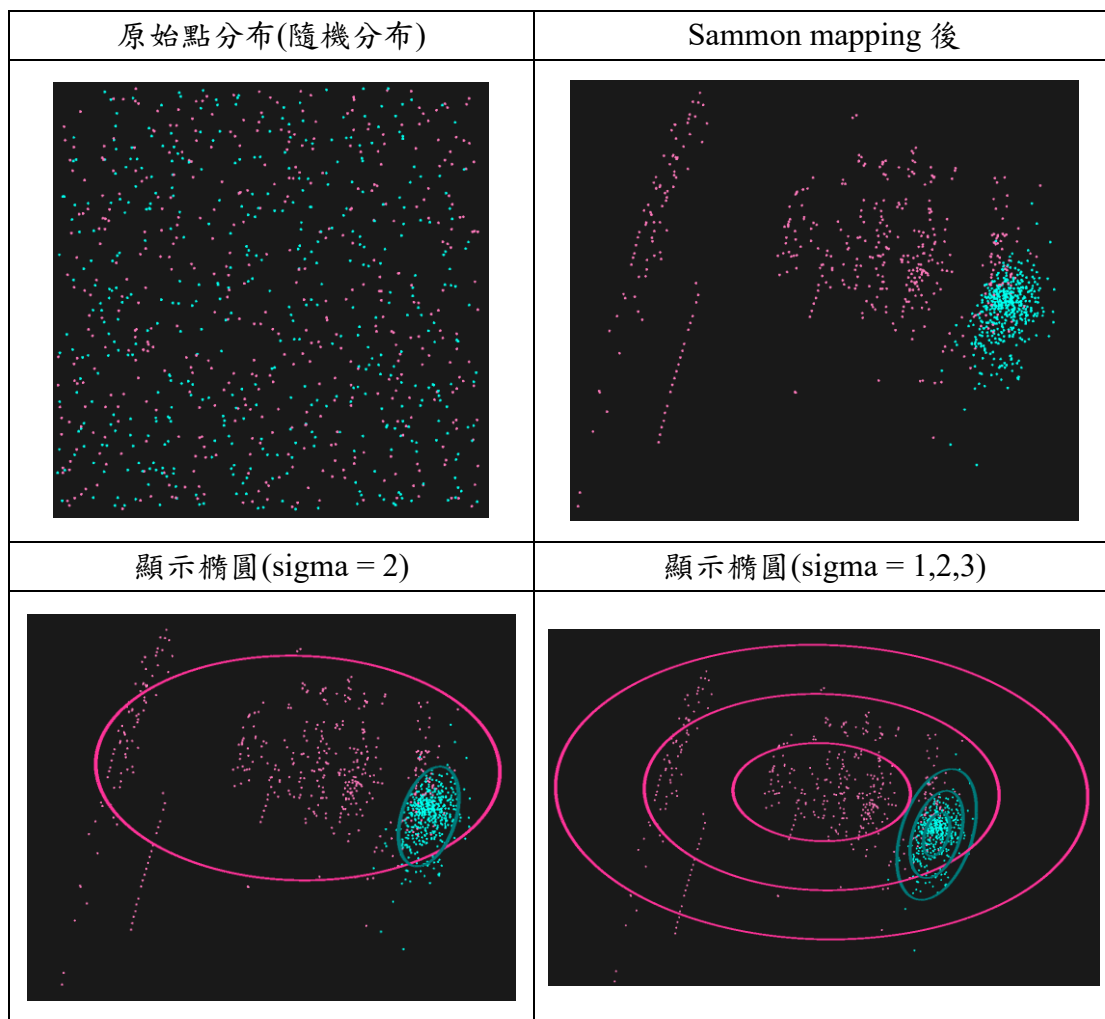
➤ 40 個點

原始點分布(隨機分布)	Sammon mapping 後
	
顯示橢圓(sigma = 2)	顯示橢圓(sigma = 1,2,3)
	

➤ 100 個點

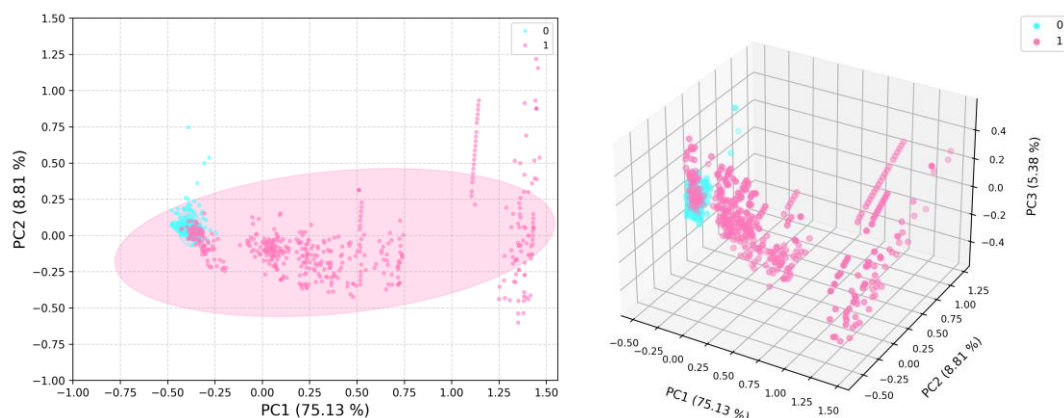
原始點分布(隨機分布)	Sammon mapping 後
	
顯示橢圓(sigma = 2)	顯示橢圓(sigma = 1,2,3)
	

➤ 900 個點



5 加分題 – PCA 降維(不限語言、套件)

我使用的是 py 以及 sklearn 函式庫實作，左圖是降到二維，右圖是降到三維。



6 心得

這次的作業蠻有趣的，之前只在機器學習課有接觸到 PCA、降維、分群等相關的知識，沒想到這次我們竟然要自己寫!! 也了解到了 Python 那些函式庫的方便之處。