

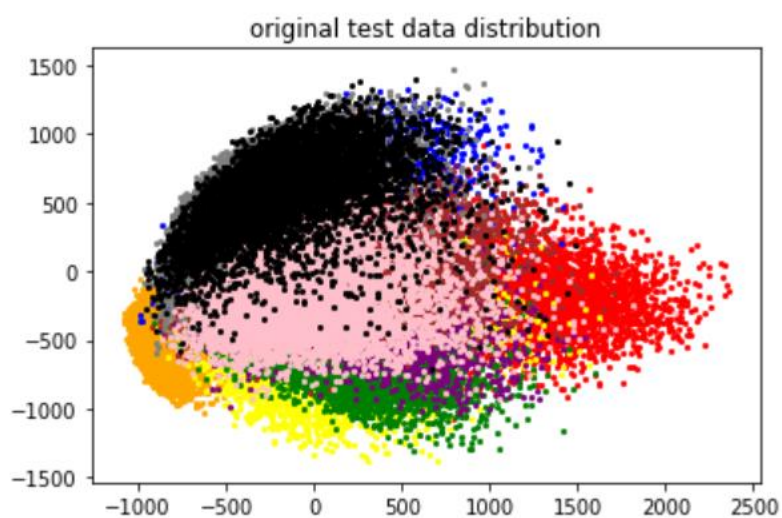
Machine Learning HW#4

0510822 陳紀愷

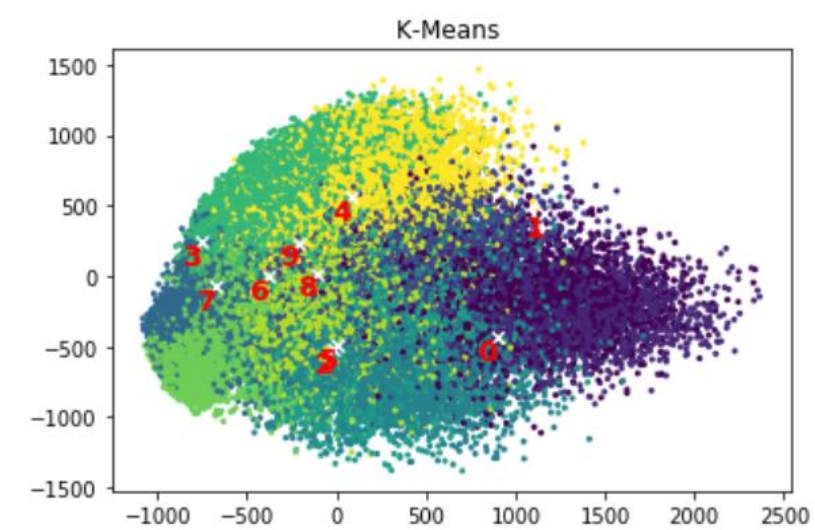
Problem 1: Clustering

(a)(b)

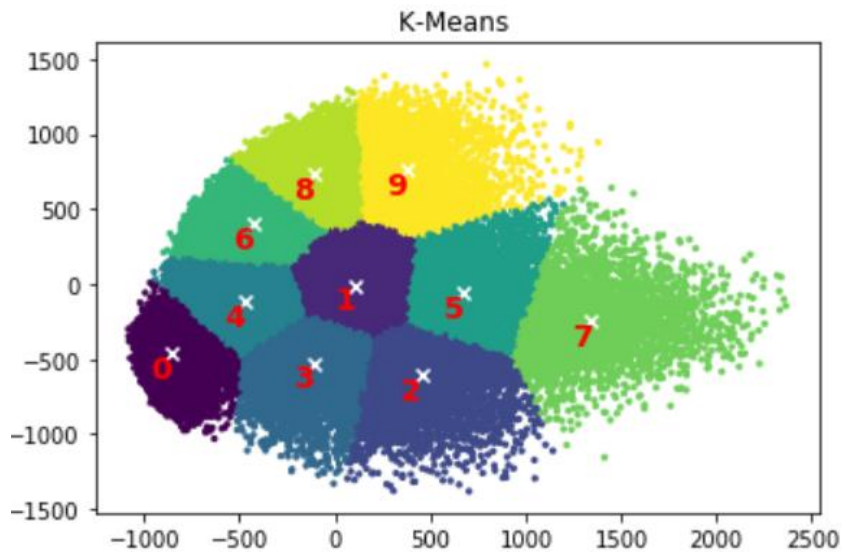
為了可以匯出資料分布圖，下圖皆有對資料做 PCA($n=2$)



↑ 原先的資料分布



↑ 對資料做 K-means 以後再 PCA

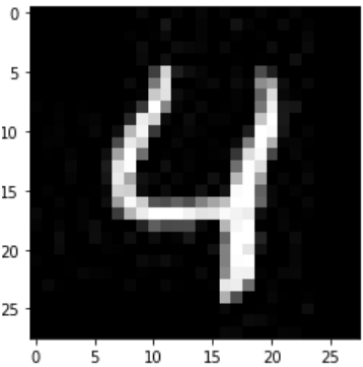
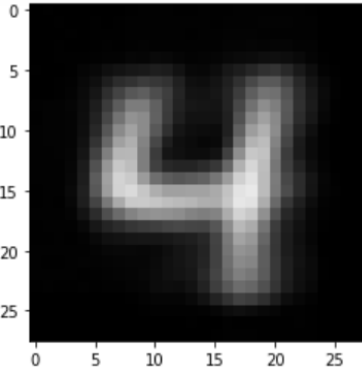
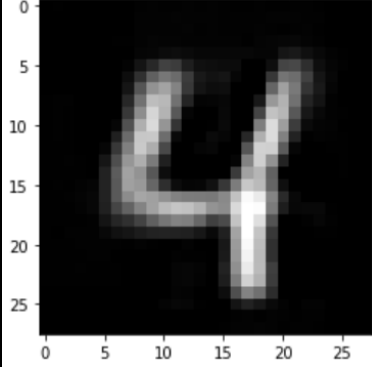
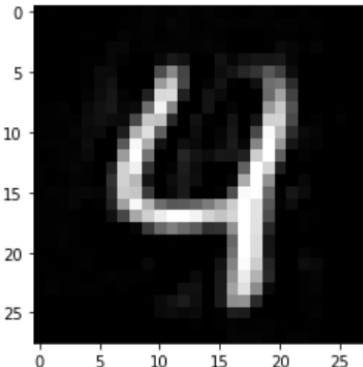
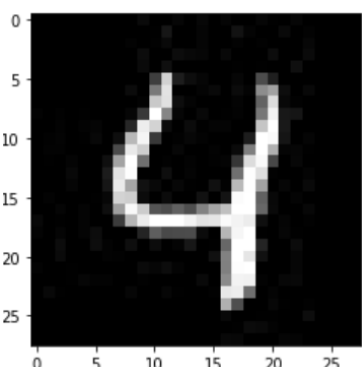


↑ 先對資料做 PCA($n=2$) 以後再 K-means

比較了上面三種圖以後可以發現，直接對資料做 K-means 的資料分布圖和原本的分布圖最相近，可是資料卻會混在一起難以分辨與判斷分類，原因可能是因為在分類時因為特徵較多而多考慮了許多不重要的特徵。先對資料做 PCA 後再 K-means 則因為有先提取出比較重要的特徵(降維)，所以做 K-means 後可以很明顯地分出不同的群體。

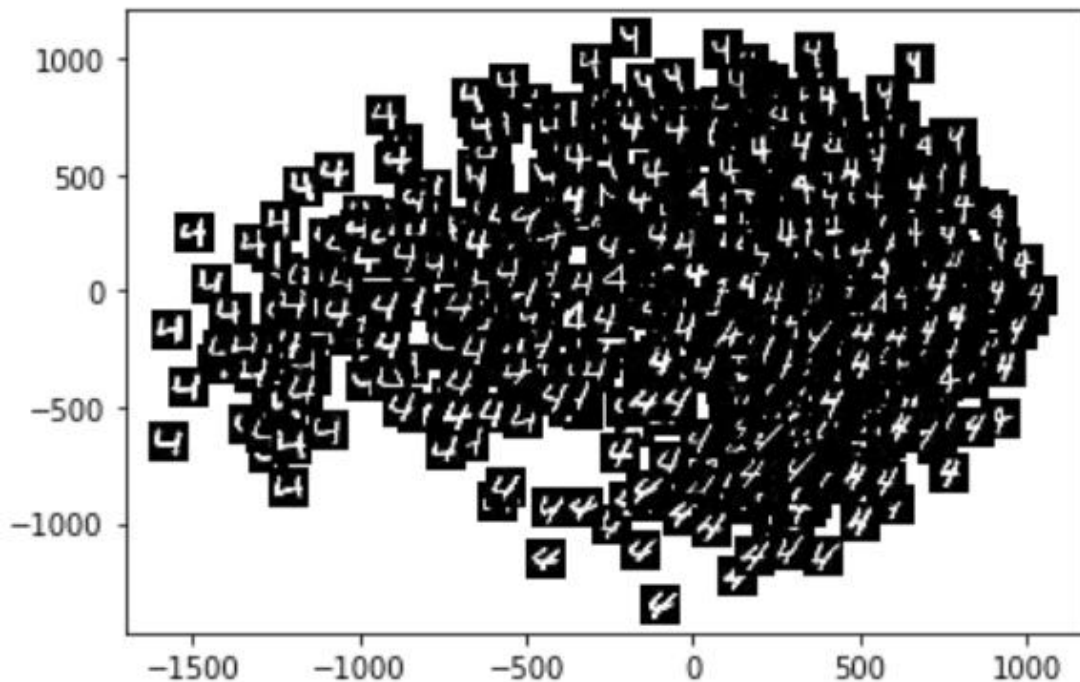
Problem 2: Dimension reduction

(a)

786 components	2 components	16 components
		
64 components	256 components	
		

PCA 是將舊的資料中的特徵去做線性組合，想辦法以有限的 k 個新特徵互不相關(最大化樣本方差)，來把資料從高維空間投影到低維度表示的方法，因此如果留下生成的新特徵越少，代表留下來的特徵越少，因此重建後也會越模糊。

(b)



可以發現比較方正的”4”和歪斜的”4”在分布上有一小段距離。

(c) From the perspective of application, please explain the similarity and difference between LDA and PCA.

LDA 和 PCA 最大不同的點是 LDA 是監督式學習，LDA 在投影的時候也會考量到 label 的存在，希望類別跟類別之間的區別性越開越好，因此 LDA 比較適合用在原本就有很好的分類的地方，像是臉部辨識或是語音辨識等，而 PCA 就相反，適合壓縮沒有明確分類的資料。

(d) From the perspective of mathematical derivation, please use objective function to explain the correlation and difference

between LDA and PCA.

objective function Of PCA

$$\max_w (Xw)^T (Xw) \text{ s.t. } w^T w = 1.$$

objective function Of LDA

$$J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2| = |w^T \mu_1 - w^T \mu_2| = |w^T (\mu_1 - \mu_2)|$$

The difference between PCA and LDA is that LDA is parametric since it assumes unimodal Gaussian likelihoods and LDA will fail when the discriminatory information is not in the mean but rather in the variance of the data ◦