

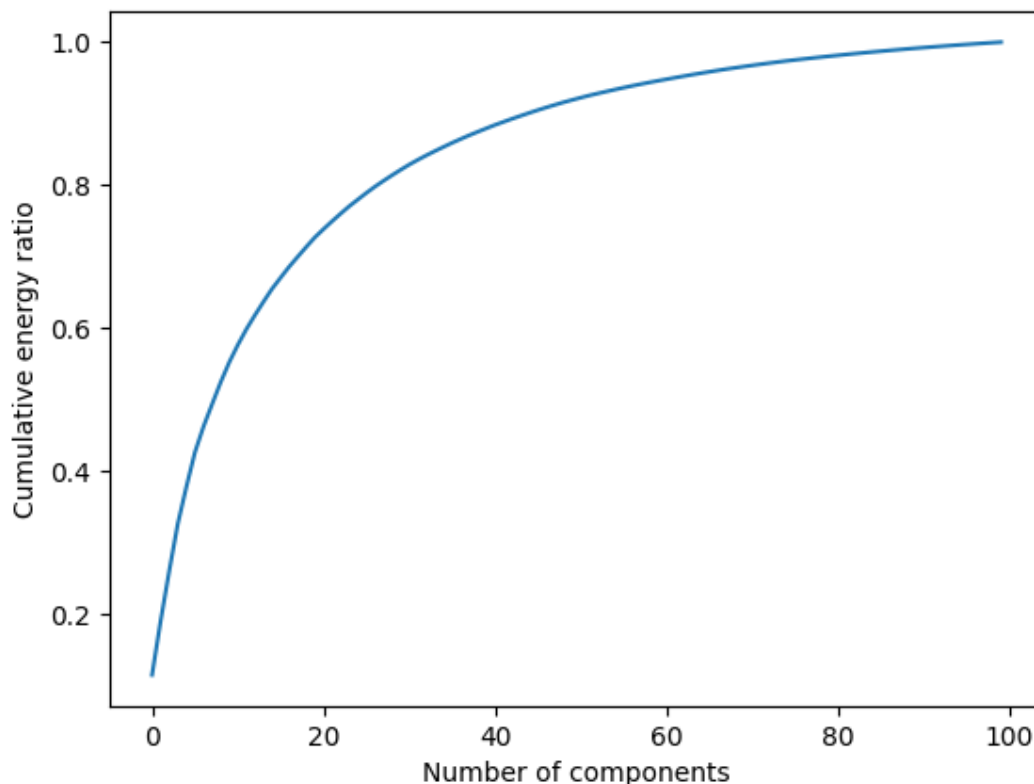
應用學習方式於人與機器之互動

HW1

姓名：吳建澄
學號：NM6111035

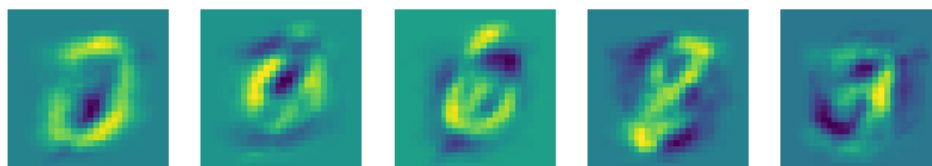
一. PCA – 應用於資料壓縮

1. Eigenvalue 的能量累加曲線圖



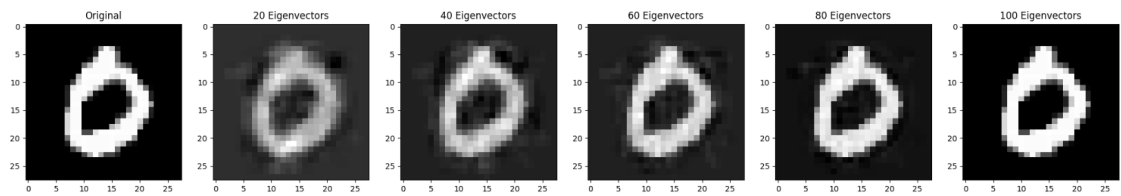
此圖顯示了PCA中每個主成分對數據平方差的貢獻程度。該曲線圖顯示前幾個主成分的貢獻很大，而其餘主成分的貢獻較小，有些甚至可以忽略不計。因此，可以只使用前幾個主成分來表示數據，同時保留較高的信息量。這樣做的優點是能夠降低數據的維度，使得數據更加易於處理和理解，同時減少數據中的噪聲和冗餘信息。然而，缺點是可能會丟失一些信息，因為PCA是通過將數據投影到較低維度空間來進行降維的，可能無法完全還原原始數據。因此，在選擇PCA降維時需要平衡信息的丟失和保留。

2. 不同的 eigenvector 的視覺化結果



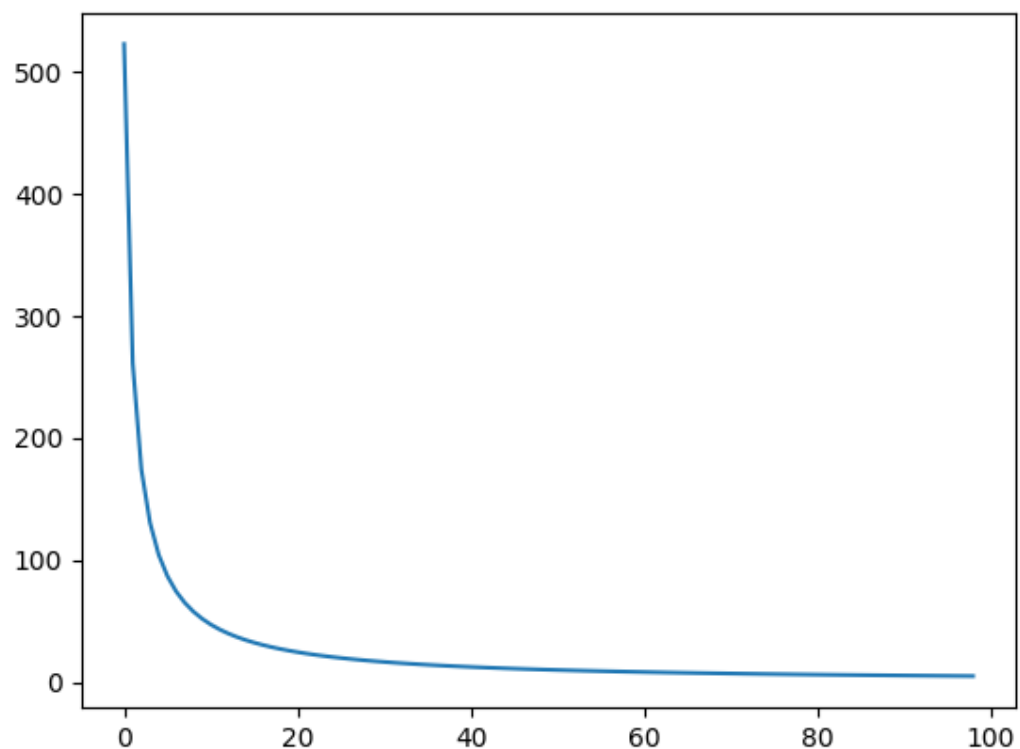
提取了16個 eigenvectors 和它們的特徵值 eigenvalues。然後選擇了前5個主成分來視覺化。它們的影響這些圖像可以用於解釋PCA如何將數據從高維空間降到低維空間，以及如何提取數據的主要特徵。

3. 不同 eigenvector 個數所形成的 PCA space 與其對應的影像重建結果



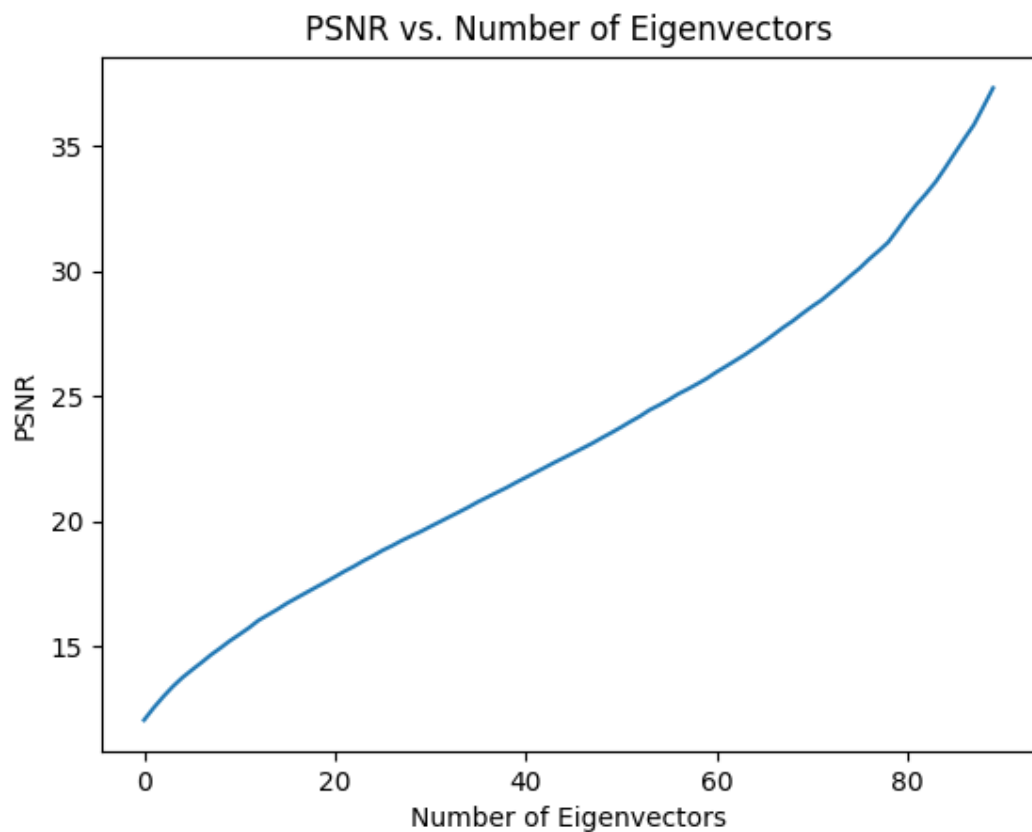
這次所選得 eigenvector 分別為20 40 60 80 100, 可以明顯的感受出來若 eigenvector 的個數越大, 其影像還原度越高, 圖片也越清晰。

4. 使用不同 eigenvector 個數所形成的 PCA space 的影像投影完的檔案大小與原始檔案大小的壓縮比曲線圖



壓縮比在前面會急速下降的原因是 PCA在前幾個 eigenvector 能夠捕捉到大部分的變異性, 新此使用較少的 eigenvector 就能夠有效的壓縮, 但隨著 eigenvector 增加, 能夠捕捉到的變異與下降, 壓縮比也就跟著趨緩。

5. 使用不同 eigenvector 個數所形成的 PCA space 的影像重建結果與原始影像計算出來的 PSNR 值所形成的曲線圖。



根據題4的分析, eigenvector 的個數越大圖片還原度就越多, 而 PSNR 的公式來看, 若兩張圖的差距越小, PSNR 的數值就越大, 因此eigenvector 的個數與 PSNR 為正比關係。

6. 心得

從上面的作業學到了如何使用PCA的函式庫以及了解PCA是如何壓縮資料並降為的運作, 也從觀察不同的 eigenvector 個數對於 PCA 的影響, 包含壓縮比, 圖片會如何改變, PSNR的數值等等, 最後感謝chatGPT的大力幫忙。

二. SVM – 手寫字元辨識

1. 使用 PCA 特徵值做 SVM 輸入的訓練時間、分類時間、及分類正確率的統計。

訓練影像張數	PCA維度	SVM訓練時間	測試影像張數	SVM分類時間	分類正確率
1000	16	30ms	100	4ms	91.0%
5000	16	207ms	500	58ms	92.4%
10000	16	506ms	1000	151ms	94.2%

2. 直接使用原始影像做 SVM 輸入的訓練時間、分類時間、及分類正確率的統計。

訓練影像張數	SVM訓練時間	測試影像張數	SVM分類時間	分類正確率
1000	64ms	100	7.3ms	94.0%
5000	742ms	500	142ms	90.2%
10000	2472ms	1000	624ms	90.6%

3. 上述結果的分析以及心得。

從上面的 SVM 結合 PCA 降維學到了如何利用 PCA 幫助我們資料的壓縮，節省我們的訓練時間，並且能夠提取重要的特徵，從上面結果可以看到，除了在訓練 1000 張時，直接使用原始圖片的結果較好以外，其他的結果都是經過 PCA 降維後在使用 SVM 的結果比較好一些，在訓練以及分類時間上面，經過 PCA 處理後的時間也比使用原圖做訓練的時間減少 50% 以上的時間。