實驗報告

N96041101 黃偉鑫

**環境:**  
作業系統: Windows 10

使用語言:Python 3.5

Python library: numpy 、matplotlib、sklearn、scipy

必要檔案:

train-labels.idx1-ubyte、train-images.idx3-ubyte

**實驗步驟:**

1. 先從網站(<http://yann.lecun.com/exdb/mnist/>)下載train-labels.idx1-ubyte及train-images.idx3-ubyte，因為train-images.idx3-ubyte 需要靠train-labels.idx1-ubyte才能知道現在讀取的檔案是什麼數字所以兩個檔案都需要下載。
2. 下載的檔案與convert.py及 sk\_pca.py放在同一資料夾內
3. 執行convert.py抓取train-images.idx3-ubyte裡的資料轉換成csv格式，以方便運算。

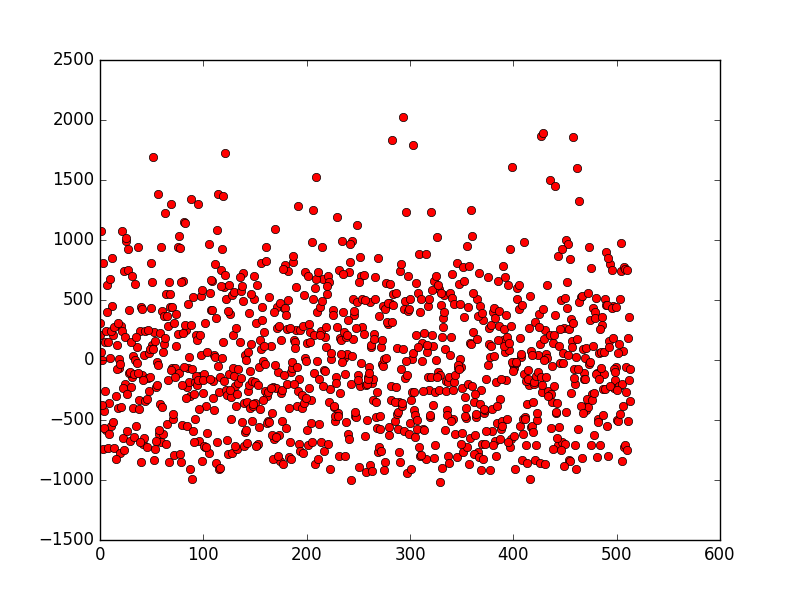
Usage: python convert.py {想要讀取的數量} {想要抓取的數字}

Example : python convert.py 6000 5

1. 會產生一test.csv檔，可用excel或其他軟體開啟。
2. 執行sk\_pca.py進行pac降維運算，並且產生result.csv紀錄結果以及圖像化顯示

**實驗結果:**

(以抓取前6000個檔案數字為5為範例)



**實驗心得:**

因為還無法完成將每點顯示原始的圖片，所以沒有辦法觀察出降成2維後的兩個維度分別代表什麼特徵。嘗試實作的失敗經驗是因為降成2維後再推算回原始圖片的28\*28維度的的準確率太低，無法達到95%以上的正確率，測試結果只能降維到40維以上才能使成功還原的正確率達到95%以上。

**原始碼:**

我將我的專案放在我的[github](Homework_1-A)上，clone下來即可執行。