

學號:0411509 姓名:許家維

此次作業我沿用上次作業的 data 他的參數主要是透過 Wavelet Transformed 得到鈔票圖像的

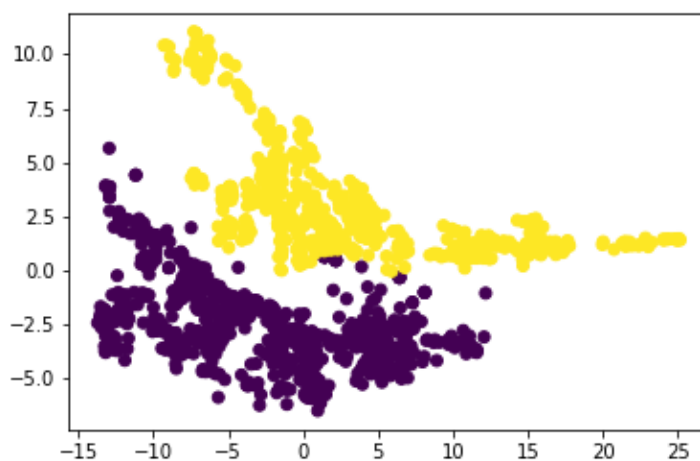
- A. Variance
- B. Skewness
- C. Kurtosis
- D. 圖像亂度(entropy)

目標補足缺失資料

通常補足缺失值有幾個方案解決

1. 案樣本平均處理
2. 刪除該樣本
3. 擬和數據

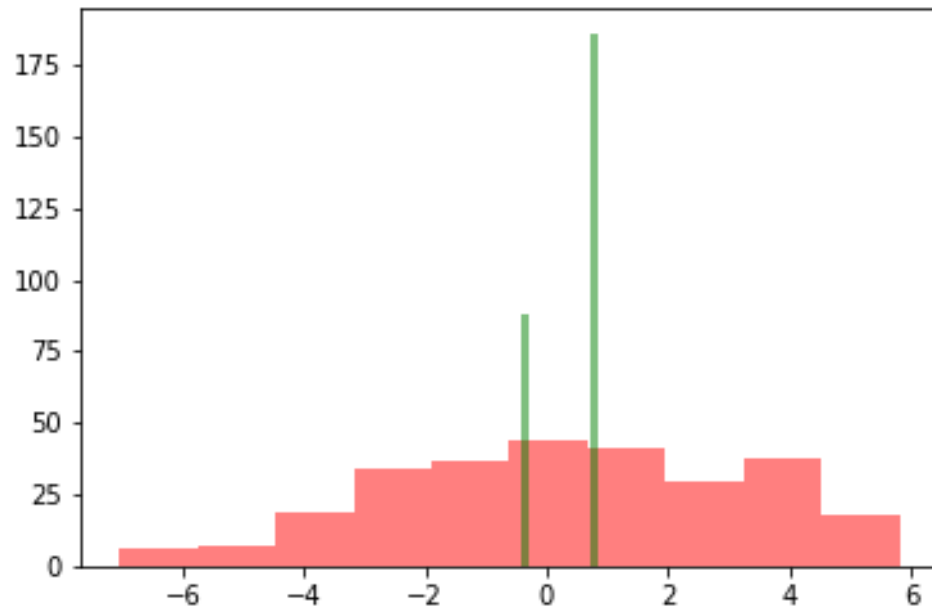
由於我知道鈔票真偽只有 2 類，並且藉由前面的 isomap 降維，發現數據總體大約分為 2 類



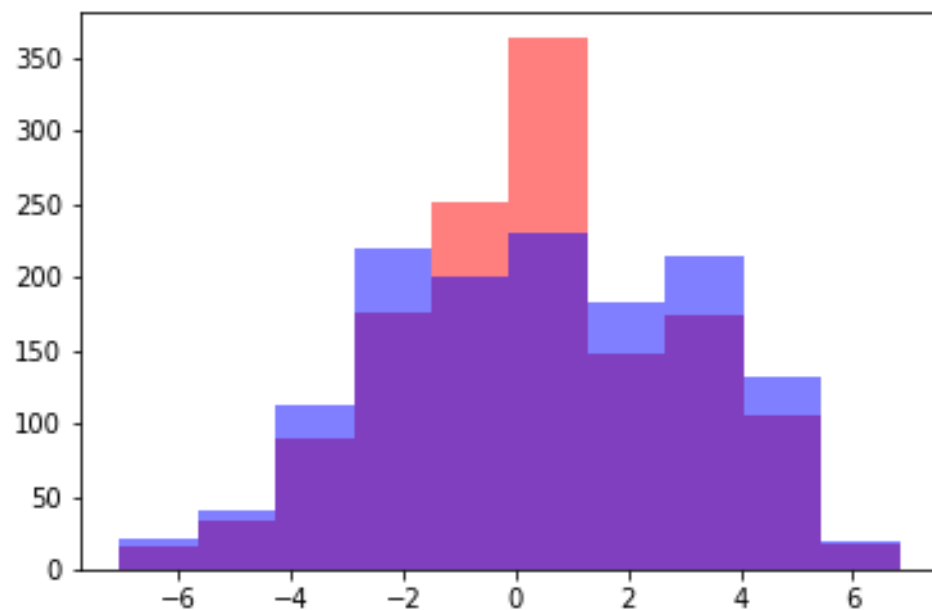
於是我利用 k-mean 演算法，先將數據做分類標籤，而缺失值將會補上該類的平均，理當會比補上總體平均精確。

實作過程

先將 Skewness 任意刪除 2 成的數據其他依樣保持，藉由上述方法補值。



其中綠色為我補的值，而粉色為原始缺失值的值，可以看到簡單分為 2 類還是太簡單了，隨後我將插值後的總體數據和原始數據去比較，其中藍色為原始數據，粉色為補值數據，可以看到僅是透過平均值補值可能十分不足。



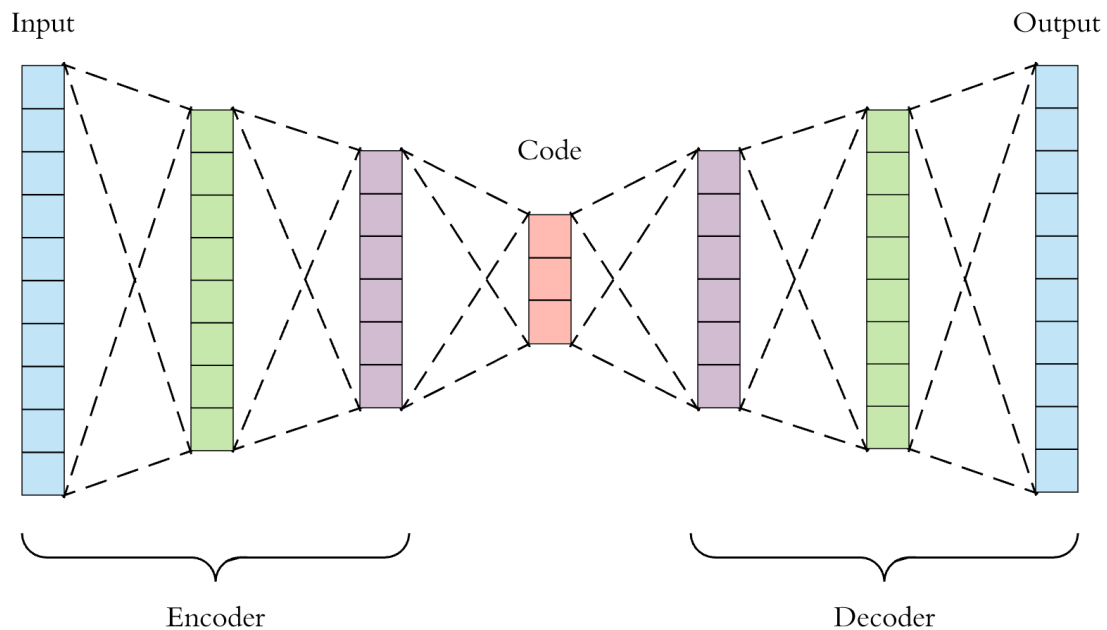
另外關於數據丟失也可能出現在圖像中，所以我選用了標準的手寫圖像作為資料，其中包含了：

1. 28*28 解析度的圖片
2. 實際對應圖片的標籤
3. 資料來源: tensorflow MNIST_data

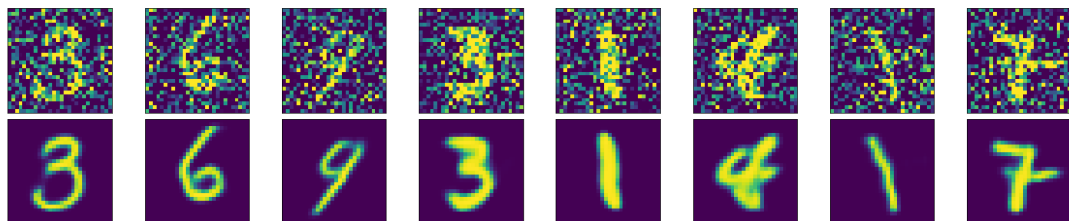
而我透過隨機削去一半的像素，模擬數據缺失的過程

實作過程中利用[1]的原理，搭建 autoencoder 以利還原原始數據

首先使 28*28 的數據被降為成 3D 的資訊，並記錄該數字落在哪個區間中進而分類數據隨後透過 decoder 解碼還原出原始數據，過程如下圖。



下圖為我實作的 demo，上排為缺失的數據圖片，下排為 model 補齊後的數據。



參考資料

- [1] A. M.Saxe, Y.Bansal, J.Dapello, M.Advani, A.Kolchinsky, B. D.Tracey, and D. D.Cox, “on the Information Bottleneck,” *Iclr 2018*, pp. 1–27, 2018.