

hw2-Sparse Coding

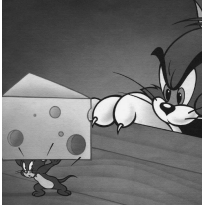
張泳樺-統研所碩一

主題

利用 Sparse Coding 稀疏化矩陣重建圖片

I. DATA

此次圖片分為兩種類型，第一種為完整的圖片如圖所示：



第二種圖片為經過干擾後的圖片，如圖所示：



II. 目的

期望藉由稀疏化矩陣，找到一組能分別表達兩組圖片集的基底向量，能利用這些基底向量將圖片集表示為這些向量的線性組合，達到重建圖片的目的

將輸入的圖片以向量 X 表示，假設有一組基底向量 D_i ，且 α_i 為相對應 D_i 的係數， $D_i = [d_1 d_2 d_3 \dots d_k]$ 即代表 X 的基底， k 表示字典的向量個數，則 X 可以表示為

$$X = \sum_{i=1}^k \alpha_i D_i - \textcircled{1}$$

Loss 可表示為：

$$\min_{\{\alpha_{x_i}\}_{i \in \Omega}} \left\| y - \sum D_i \alpha_i \right\|_F^2$$

III. 模型比較

此模型乍聽之下為 PCA，實則不其然。

A. PCA：

為一種統計中常見的降維方式，利用矩陣中的特徵向量 (Eigenvectors) 和相對應的特徵值 (Eigenvalues)，找到其不同線性變換的方式，且相對於重要的特徵向量會遠小於圖片的維度，因此利用這些相對較少的特徵向量的線性組合，而能達到降維的目的，且這些基底向量為唯一。

B. Sparse Coding：

Sparse Coding 的目的並非利用較少向量來表示原有的圖片，而是將輸入的圖片以一個字典表示所謂的字典可以形容為中華文化的繁體字，輸入的圖片則像我們看的一本書，這本書就是各種不同的文字經過各種排列組合而成新的意義，然而中華文化互古流傳，所傳下來的文字不下數十萬，因此若要精確地包含所有的文字，就有了字典的產生。

同樣地，所構成圖片的元素肯定很多，利用輸入圖片集找到屬於圖片的字典 (即 dictionary)，因此假設圖片有 p 個維度，字典有 k 個維度，則 $k \gg p$ ，利用字典中的向量的線性組合來構成 X (即輸入的圖片)。明顯地，在 $\textcircled{1}$ 式中的解肯定不唯一。

此即為兩方法最主要的不同

IV. SPARSE CODING 模型解釋：

既然 Sparse Coding 的解不唯一又該如何訓練？Sparse Coding 的 Sparse 是在稀疏化字典前的係數即為 a_i ，讓只有少數的 a_i 非零，可以利用 L2 norm 來達到稀疏化的目的。

因此 $\textcircled{1}$ 式 Loss 可表示為：

$$\min_{\{\alpha_{x_i}\}_{i \in \Omega}} \left\| y - \sum D_i \alpha_i \right\|_F^2 + \lambda \|\alpha_{x_i}\|_1$$

V. 程式架構

目標-loss 最小值

利用 loss 越來越小來更新權重 a_i 和字典 D_i 利用分解線性轉換來更新 α_{x_i}

$$\hat{\alpha}_{x_i} \leftarrow D_i^T (y - \sum_{j \neq i} D_j \hat{\alpha}_{x_i})$$

由於 D_i 並非 convex function，因此可以利用 D_i 和 α_i 分開學習，其中 α_i 限制式可以表示為

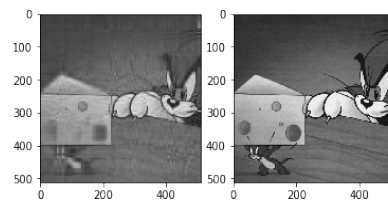
$$\hat{\alpha}_{x_i} \leftarrow \delta_\lambda(\alpha_{x_i})$$

所以在每次迭代中分別要執行：

(1) 固定 D_i 學習 α_i (2) 固定 α_i 學習 D_i

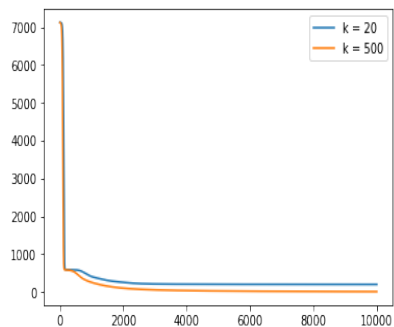
VI. 結果

A. 第一類圖片的結果

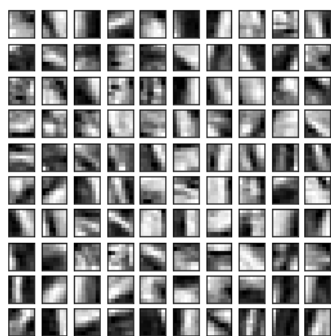


分別將字典數目取 $k = 20$ 和 $k = 500$

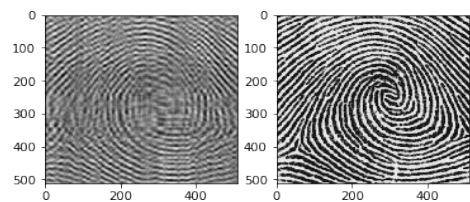
B. 第一類圖片的 $loss$



C. 第一類圖片的字典集

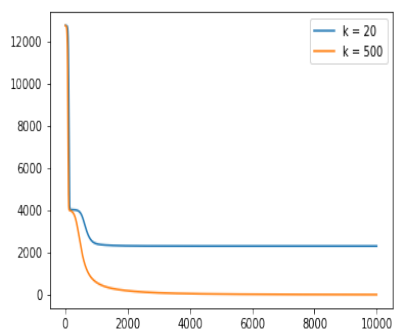


D. 第二類圖片的結果



分別將字典數目取 $k = 20$ 和 $k = 500$

E. 第二類圖片的 $loss$



F. 第二類圖片的字典集



VII. 結論

可以發現 sparse coding 之所以需要較多的字典型式，是因為可以促進收斂的速率顯示的結果也較好