DSP 2025 HW3

TA: Ching-Chia Kao

Mail: d11922015@ntu.edu.tw

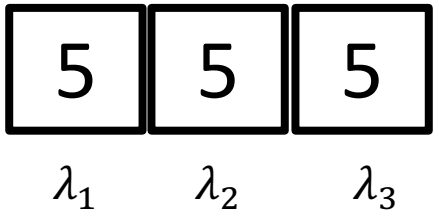
Deadline: 4/17 11:59pm

MNIST Dataset

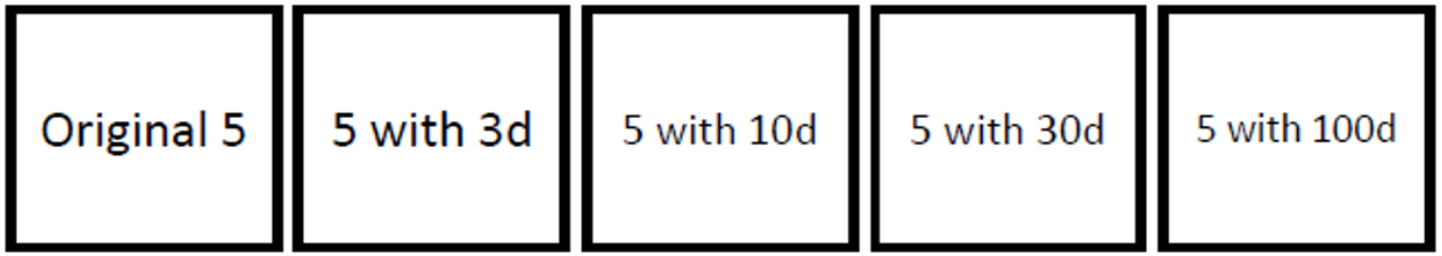
- 70,000 images
- Each image has 28*28 pixels (784)
- Use sklearn to load the MNIST dataset
- Use ‘’gray scale’’ to plot the images

Q1. Extract all the ‘’5’’ images (6313 vectors). Use **centered** PCA (5’s center) to decompose.

(a) Show the eigenvectors with the three largest eigenvalues. (5pt)



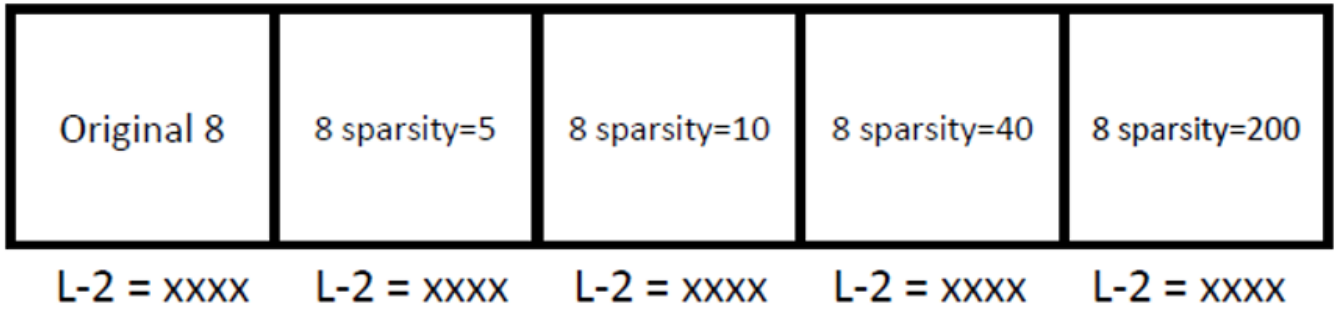
(b) Use the top {3,10,30,100} eigenvectors to reconstruct the first ‘’5’’ image. (5pt)



Q2. Define the first 10,000 images as the training set. Use OMP from sklearn to answer the following questions.

(a) Find the 5 bases of the #10001 image (‘’3’’) with sparsity=5. Show the 5 bases. (5pt)

(b) Find the bases of the #10002 image (‘’8’’) with sparsity={5,10,40,200}. Show the reconstruction images and calculate their reconstruction errors using L-2 norm (Euclidean distance) (5pt)



Q3. Extract all the ‘’8’’ images (6825 vectors). Use LASSO from sklearn to answer the following questions.

(a) Use the first 6824 images as the base set and then then reconstruct the last ‘’8’’ using LASSO. Let $\alpha = \{0.01, 0.001, 0.0001\}$ be the constant that multiplies the L1 term, show number of nonzero coefficients corresponding to each α and the ℓ_2 reconstruction error. (5pt)

(b) Handcraft the Lasso using coordinate descent (20pt)

- Explain your implementation using comments
- Compare the result with (a)
- The efficiency of your implementation will be considered

Homework submission format

- Use NTU cool to submit **{student_id}.ipynb** ---TA will run this file on google colab
- Don’t submit anything else.

Questions

Q: {student_id}.ipynb 無法執行或沒有產生圖片是全扣嗎?

Ans: 對，所以請同學繳交前務必再三確認程式可執行且可產生圖片。

Q: Eigenvectors 可能有正負之差，兩者都可以嗎?

Ans: 都可以