Program Homework 2

OMP on MNIST Dataset

OMP Algorithm(Page 30 in Slides)

- **Initialization**: let the residue signal be $r_0 = x$; set $\Lambda_0 = \phi$ (empty set), the iteration count $l = 1$.

- **Iteration**:

  Find the index s_l that solves the problem,
  $$b_{s_l} = \underset{b \in B_{\bar{\Lambda}_{l-1}}}{\mathrm{argmax}} |b^T r_{l-1}|.$$

  Let $\Lambda_l = \Lambda_{l-1} \cup \{s_l\}$.

  Solve the following least squares problem,
  $$c_{\Lambda_l} = \underset{c_t}{\mathrm{argmin}} \|x - B_{\Lambda_l} c_t\|^2,$$

  (the solution is $c_{\Lambda_l} = \left(B_{\Lambda_l}^t B_{\Lambda_l}\right)^{-1} B_{\Lambda_l}^t x$)
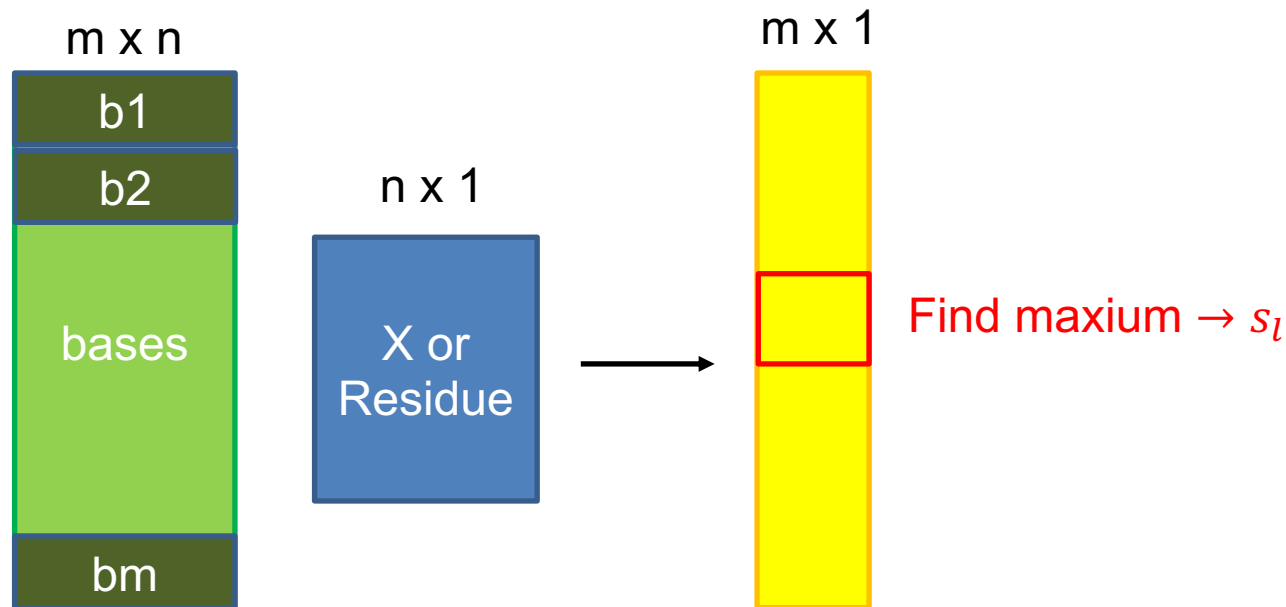
  Let $r_l = x - B_{\Lambda_l} c_{\Lambda_l}$,

- **Stopping criterion:** same as MP

Introduction

- Implement OMP algorithm (page 30 in slides) function
- def OMP(sparsity, x , B)
    - B: bases, dictionary
    - x: input signal
    - sparsity: sparse degree of coefficients
- Return: recovered signal by sparse representation, error(Euclidean distance from x)
- Note: bases are unit-length vectors (page 16 in slides)

- Only data processing, image processing, linear algebra libraries are allowed

How to find s_l

m x n

b1
b2
bases
bm

n x 1

X or Residue

m x 1

Find maxium → s_l

- **Iteration**:
  Find the index s_l that solves the problem,
  $$b_{s_l} = \operatorname*{argmax}_{b \in B_{\bar{\Lambda}_{l-1}}} |b^T r_{l-1}|.$$
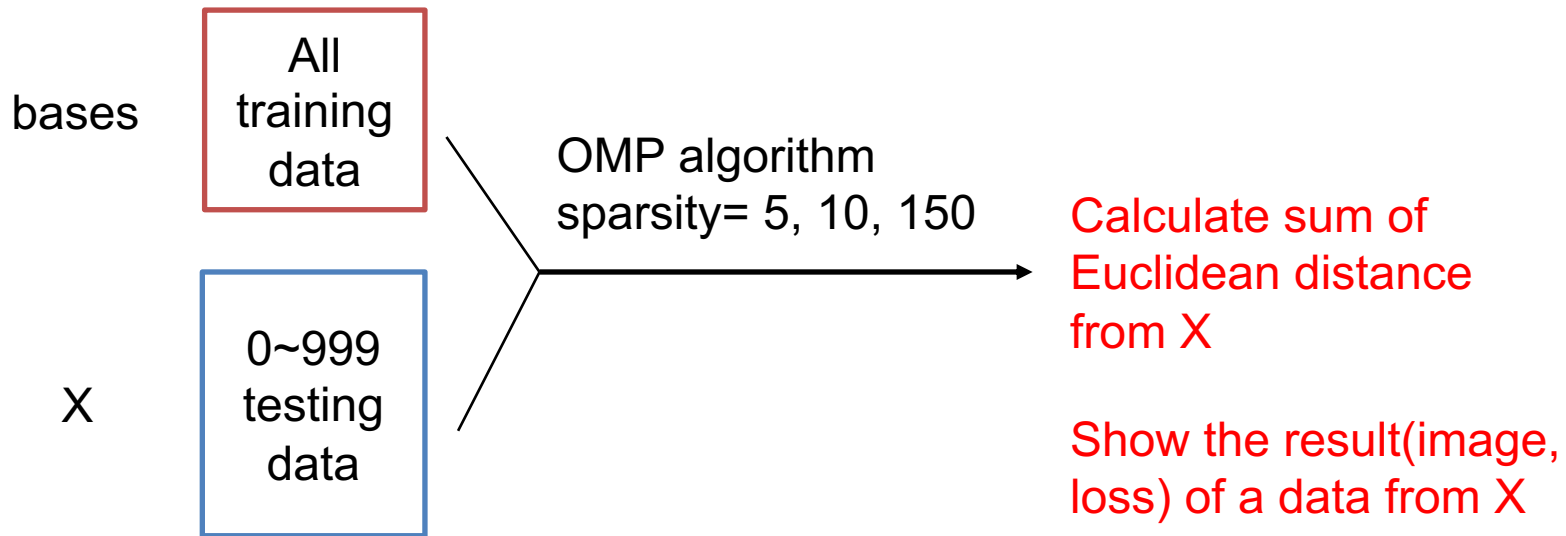 Let $\Lambda_l = \Lambda_{l-1} \cup \{s_l\}$.

How to get B_{Λ_l}

- selected_indices = [ ]
- E.g., selected_indices = [1, 3, 14]
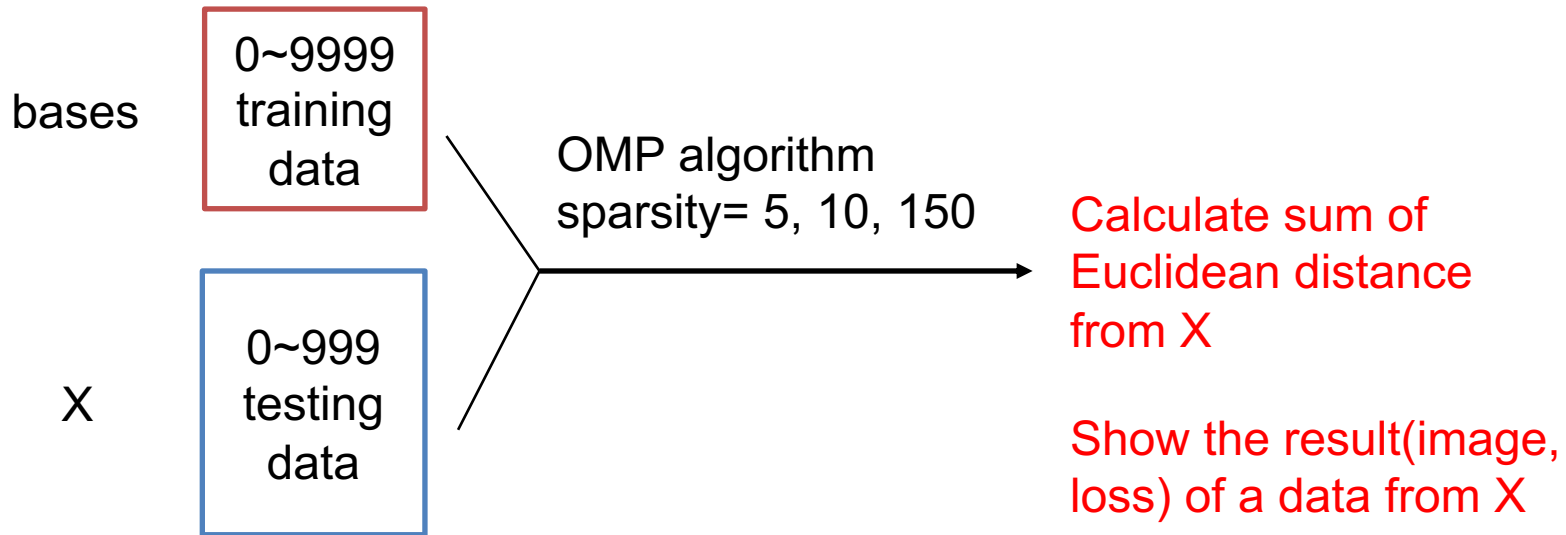- selected_bases = bases[selected_indices, :]

Question (1.a)
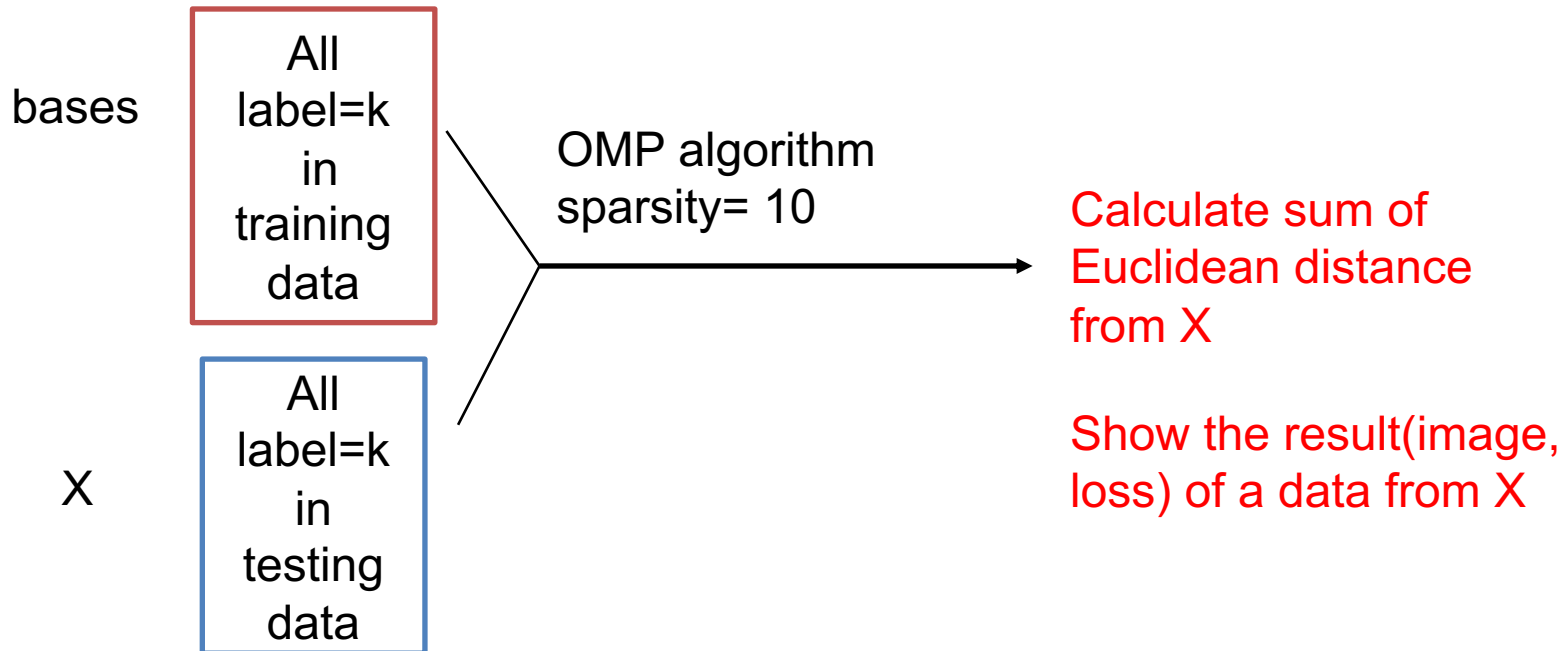
- All training data as bases, set sparsity= 5, 10, 150

bases

All training data

OMP algorithm
sparsity= 5, 10, 150

Calculate sum of Euclidean distance from X

X

0~999 testing data

Show the result(image, loss) of a data from X

Question (1.b)

- 10000 training data as bases(0~9999), set sparsity= 5, 10, 150

bases

| 0~9999 training data |

X

| 0~999 testing data |

OMP algorithm
sparsity= 5, 10, 150

Calculate sum of
Euclidean distance
from X

Show the result(image,
loss) of a data from X

Question (2.)

- All label=k(e.g., k=3, k can be 0~9) in training data as bases, set sparsity= 10

bases

All label=k in training data

X

All label=k in testing data

OMP algorithm sparsity= 10

Calculate sum of Euclidean distance from X

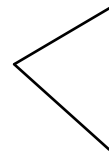Show the result(image, loss) of a data from X

Question (3.)

- Compare to centered PCA

All training data

All label=k in training data

OMP algorithm
sparsity= num of component
when PCA energy= 95%,50%,10%

Centered PCA
energy= 95%,50%,10%

Calculate the Euclidean distance from X

Show the result(image, loss)

X

A data label=k in testing data

Requirements

- Report
- Code
- Readme(library version, how to run the code)
- Due date: **2020/11/23**