**MSC-BDT5002/MSC-IT 5210 Knowledge Discovery and Data Mining, Fall 2017**

Assignment 1

Student No. 20470873

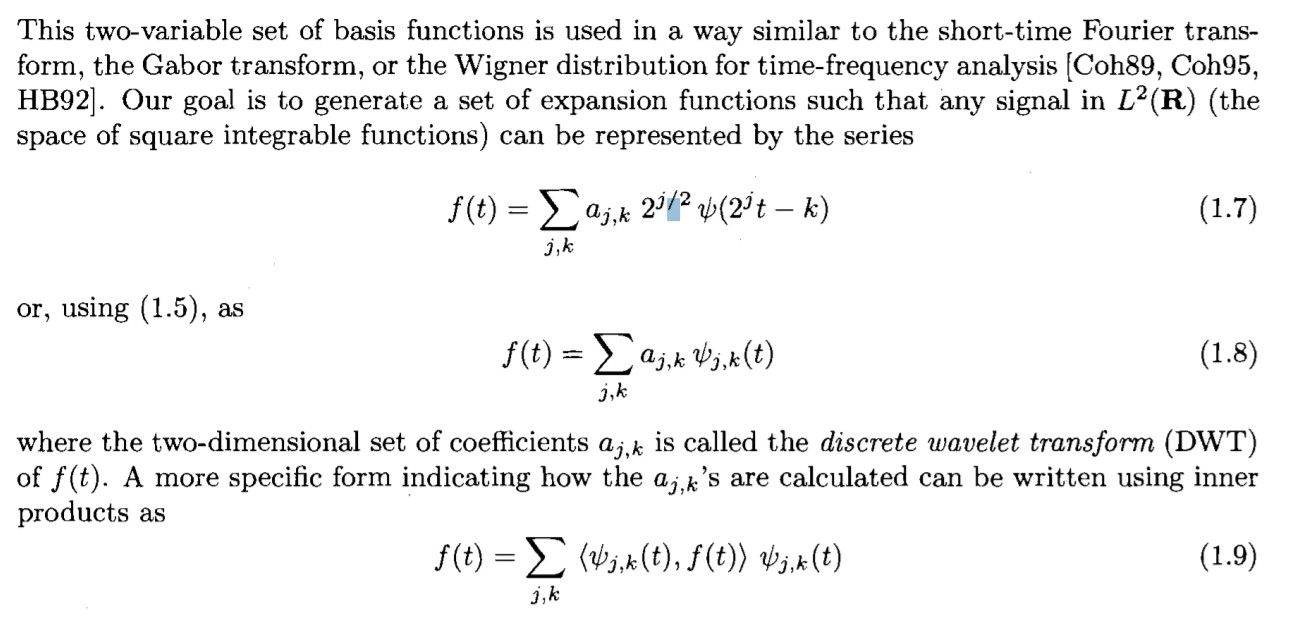
Li Jiawei

# 2.Data Preprocessing

## 2.1 Wavelet Transform

Q1:

The definition of discrete wavelet transform is below:



In my opinion, descrete wavelet transform(DWT) is a technology about linear sinal processing skill. We can build a data vector X with our data, using DWT to tansform it into another X’ with different value. We can keep these two vectors with the same length. In this way, we can create another vector contains the zip data of the remote vector. The most simple way is above, the coefficient of f(t) means the value of 2-D data vector, we call X’ above.

Q2:

1.From the specific S=[1,4,2,3,-2,-1,2,1], we divide it into two part equally, the result is:

A=[1,4,2,3]

B=[-2,-1,2,1]

2.Next,we make some pairs using the number has the same index in these two pairs like(A[i],B[i])(i=0,1,2,3),and we calculate the value of (A[i]+B[i])\*h and put it into index i, also calculate the value of (A[i]-B[i])\*h and put it in to index (i+4). In this question, h is 1/sqrt(2).and we can make another S’ like below:

S’=[-0.707, 2.121, 0.828, 2.828, 2.121, 1.414, 0, 1.414]

3.Next,we calculate the S’’ from S’ using the way like step 2. But we just using the front part of S’. And keep the back part stable. We can calculate the S’’ like below:

S’’=[1.500, 3.500, -2.500, -0.500, 2.121, 1.414, 0, 1.414]

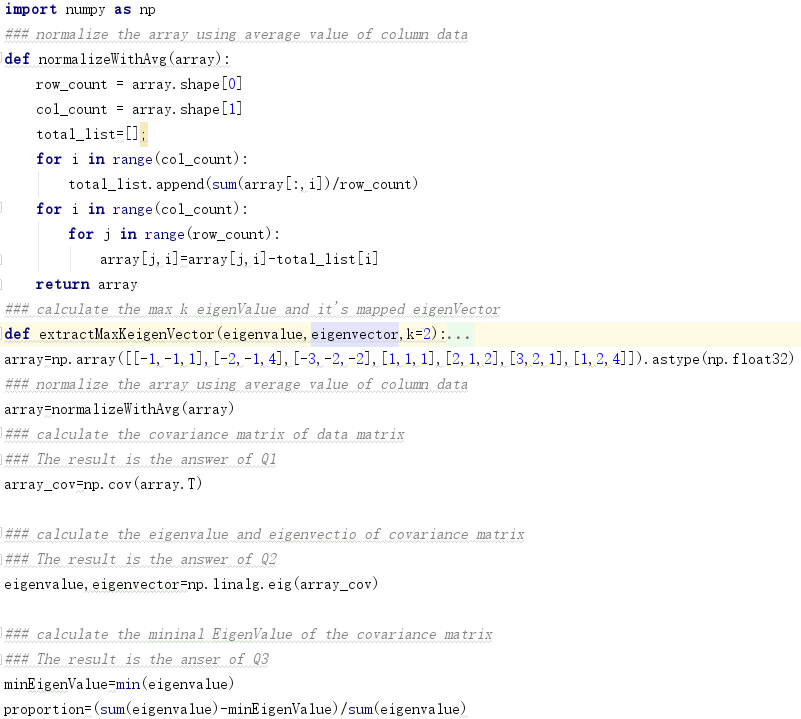
4.Finally,we calculate the S’’’ from S’’ using the way like step 3. But we just using the front quarter part of S’’. And keep another parts stabe. We can calculate the S’’’ like below:

S’’’=[3.536, -1.414, -2.500, -0.500, 2.121, 1.414, 0, 1.414]

S’’’ is the final result of Q2.

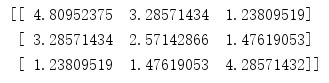
## 2.2 Principal Components Analysis

The code of PCA with python is below:



The output is below:

Q1:

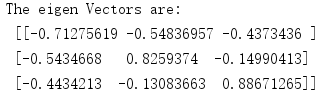


Q2:

EigenValue:



EigenVector:



Q3:

