Assignment 1
CSCI-6390

Yiqing(Alice) Guo

September 21, 2017

For all part that all results have precision of 8

read me
python version: 2.7.11
sample command line input: assign1.py airfoil_self_noise.dat 8 0.0001 out.txt
python file, data file, precision, epsilon, output file
all print text result are save in out.txt
if there are no 5 argument then the output will print in the command windows.
sample command line input with 4 arguments: assign1.py airfoil_self_noise.dat 8 0.0001
if less than 4 argument then the code will using the defualt value.
4 figure are saved as most_correlated.png, anti_correlated.png, least_correlated.png, eigenvectors.png

1 Part 1

Compute the mean vector μ for the 6-dimensional data matrix, and then compute the total variance $var(D)$

The mean vector μ:

$$[2886.38057219 \quad 6.78230206 \quad 0.13654824 \quad 50.86074518 \quad 0.01113988 \quad 124.83594278]$$

Total variance $var(D)$: 9932429.7172

Compute the sample covariance matrix Σ as inner products between the attributes of the centered data matrix.Next compute the sample covariance matrix as sum of the outer products between the centered points

Inner products:

$$\begin{bmatrix} 9932104.79728063 & -5085.67250617 & -1.07878216 & 6557.77175899 & -9.53323084 & -8491.74115188 \\ -5085.67250617 & 35.00093761 & -0.27930199 & 5.41178032 & 0.05859369 & -6.36918252 \\ -1.07878216 & -0.27930199 & 0.00874405 & 0.00551227 & -0.00027147 & -0.1522949 \\ 6557.77175899 & 5.41178032 & 0.00551227 & 242.35026212 & -0.00081328 & 13.43101346 \\ -9.53323084 & 0.05859369 & -0.00027147 & -0.00081328 & 0.00017281 & -0.02834618 \\ -8491.74115188 & -6.36918252 & -0.1522949 & 13.43101346 & -0.02834618 & 47.55979887 \end{bmatrix}$$

Outer products:

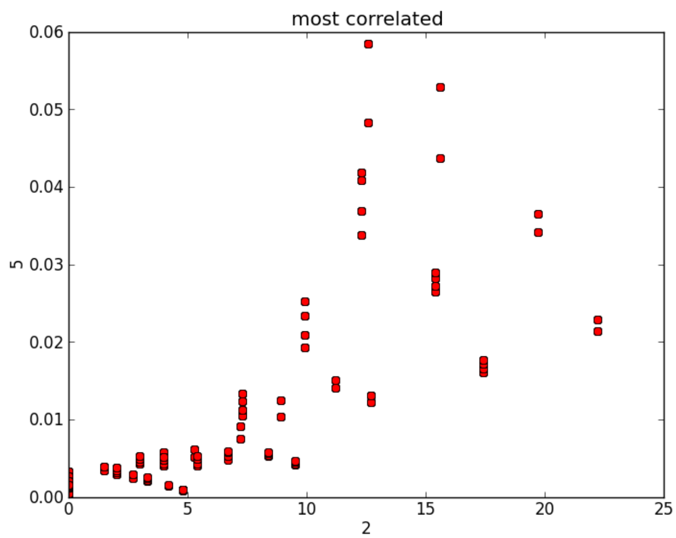
$$\begin{bmatrix} 9932104.79728063 & -5085.67250617 & -1.07878216 & 6557.77175899 & -9.53323084 & -8491.74115188 \\ -5085.67250617 & 35.00093761 & -0.27930199 & 5.41178032 & 0.05859369 & -6.36918252 \\ -1.07878216 & -0.27930199 & 0.00874405 & 0.00551227 & -0.00027147 & -0.1522949 \\ 6557.77175899 & 5.41178032 & 0.00551227 & 242.35026212 & -0.00081328 & 13.43101346 \\ -9.53323084 & 0.05859369 & -0.00027147 & -0.00081328 & 0.00017281 & -0.02834618 \\ -8491.74115188 & -6.36918252 & -0.1522949 & 13.43101346 & -0.02834618 & 47.55979887 \end{bmatrix}$$

Compute the correlation matrix for this dataset using the formula for the cosine between centered attribute vectors.Which attributes are the most correlated, the most anti-correlated, and least correlated? Create the scatter plots for these interesting pairs using matplotlib and visually confirm the trends, i.e., describe how each of the three cases results in a particular type of plot.

Colume	Colume	$cos(\theta)$	Degree
1	1	1.000000	0.00
1	2	-0.272765	105.83
1	3	-0.003661	90.21
1	4	0.133664	82.32
1	5	-0.230107	103.30
1	6	-0.390711	113.00
2	2	1.000000	0.00
2	3	-0.504868	120.32
2	4	0.058760	86.63
2	5	0.753394	41.11
2	6	-0.156108	98.98
3	3	1.000000	0.00
3	4	0.003787	89.78
3	5	-0.220842	102.76
3	6	-0.236162	103.66
4	4	1.000000	0.00
4	5	-0.003974	90.23
4	6	0.125103	82.81
5	5	1.000000	0.00
5	6	-0.312670	108.22
6	6	1.000000	0.00

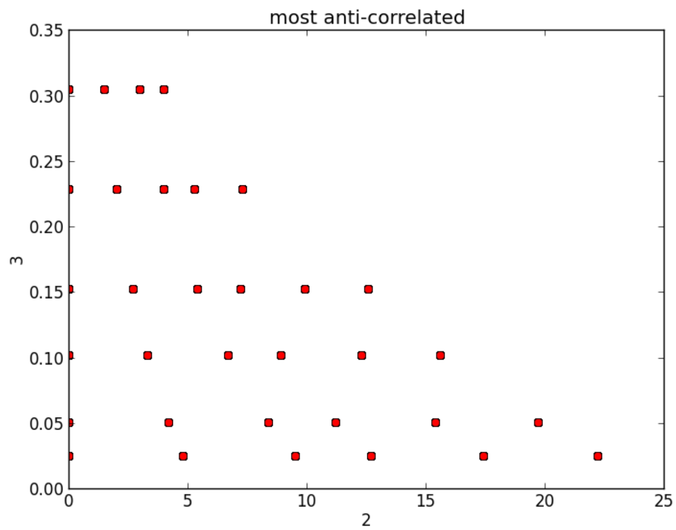
The most correlated is col: (2, 5) , $cos(\theta)$ is 0.753394, and the degree is 41.11

In this we see that one variable increases, the other tends to also increase. we could see a positive relation ship between the two attribute.



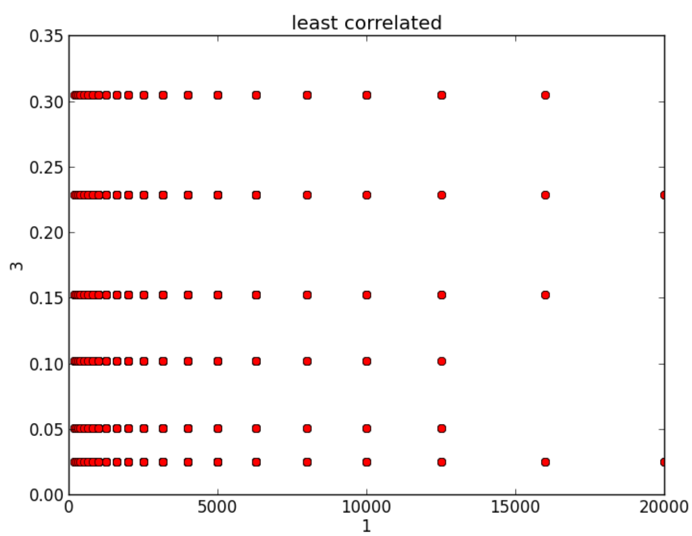
The most anti-correlated is col: (2, 3) , $cos(\theta)$ is -0.504868, and the degree is 120.32

In this we see that one variable increases, the other tends to decrease.



The least correlated is col: (1, 3) , $cos(\theta)$ is -0.003661, and the degree is 90.21

There is really weak correlation in this figure, compare with the figures above we cannot see a relation in this plot.



2 Part 2

Eigenvectors u_1 and $u_2 (\epsilon = 0.0001)$:

$$u_1 = \begin{bmatrix} 1.00070832 \\ -0.00051241 \\ -0.00000011 \\ 0.00066074 \\ -0.00000096 \\ -0.00085559 \end{bmatrix} \quad u_2 = \begin{bmatrix} -0.00049678 \\ 0.0331156 \\ -0.00006833 \\ 0.8849192 \\ 0.00001509 \\ 0.08251522 \end{bmatrix}$$