

**Subject:** [CS 631A] Slight change in Homework 6

**From:** "Sandeep K. Shukla" <sandeeps@cse.iitk.ac.in>

**Date:** 13/11/19, 4:08 pm

I am changing the objective of the Homework 6 for the following reason:

When you use the data for normal grid condition to build your diffusion space, you get a set of eigen values and eigen vectors.

Now when you map a datapoint  $X_i$  of the original space – it maps to  $[\lambda_1 v_1(i), \lambda_2 v_2(i), \dots, \lambda_K v_K(i)]$

A new datapoint  $X_p$  where  $p > N$  ( $N$  is the total number of data points from normal grid condition), then  $v_1(p), v_2(p), \dots, v_K(p)$  will not make sense the length of the eigen vectors is  $N < p$ .

So what I suggest you do is as follows:

Use all data points (normal and stressed) – and construct the diffusion matrix, and then build diffusion map. (Note that first column where 1 or 0 indicates stressed or normal condition should be omitted from the data points).

Then use that to map all data points. Now do clustering on the diffusion space.

Hope is that there will be two major clusters (one cluster for mapped points for normal grid condition and other cluster for mapped points for stressed grid condition).

Then using the highest 3 eigen values, map all data points in 3 dimension. Then plot the data points in 3D (using one color for points whose first column was 1, and another color whose first column was 0). Hopefully you will see a clustering pattern there.

Since we have never tried diffusion map on this data, we would not know whether these predictions will hold. The exercise for the homework is to find if these predictions hold.

-----  
Dr. Sandeep K. Shukla  
Head, Department of Computer Science and Engineering  
Indian Institute of Technology, Kanpur  
Kanpur, India  
<http://www.cse.iitk.ac.in/users/sandeeps/>  
<https://www.securitymooc.in>  
<https://security.cse.iitk.ac.in>

