

Mathematical Foundations for Computer Vision and Machine Learning

Byung-Woo Hong

Computer Science Department, Chung-Ang University

Assignment 3

Jupyter Notebook

- Create a new notebook for Python 3
- Include your name and the student ID in the notebook
- Write python 3 codes for the given assignment
- Try to separate the codes into meaningful blocks (행방에 따서지 말고)
- Write a comment for each block of codes
- Plot the important intermediate results
- Write a short description for each graphical result
- Use LaTeX for mathematical comments in the notebook
- Save the notebook file as [assignment03.ipynb](#)
- Download the notebook as a PDF file [assignment03.pdf](#)

Assignment 3

github

- Start a project with the name `assignment03`
- Include the link to the github project in the notebook
- Upload the notebook `assignment03.ipynb` to the project after the deadline (Note that your github project is visible to public)

(한방에 저장해
놓아도 됩니다)

Assignment 3

Submission to *only* eclass

- Submit the PDF file [assignment03.pdf](#) to [eclass](#)
- Deadline is 11:59 pm on next Thursday. No extension
- Score ranges from 0 to 5

Assignment 3

Score Table

- The results should be correct
- The codes should be written in a modulated way *divide meaningful*
- The comment should be made for each block of the codes
- The important intermediate results should be presented
- The link to the github project should be included

Assignment 3

Programming Assignment: k -means algorithm

- Implement a k -means algorithm for two-dimensional points
- Generate k number of random point clusters
- Demonstrate the k -means algorithm based on the points
- k is an input variable (you pick k for the demonstration)

Assignment 3

Essential Functions: k -means algorithm

- generatePointCluster: (number of clusters, number of points)
- computeDistance: (a pair of points)
- initialiseLabel: (number of clusters) *random label*
- computeCentroid: (number of clusters)
- assignLabel: (number of clusters, list of distances)
- computeEnergy: (list of data, list of labels)

generate data

*클러스터
랜덤 라벨
주기*

for each data x_i

$$\frac{1}{N} \sum_{i=1}^N \|x_i - c_{k(i)}\|^2$$

*Le norm을
average를
구함.*

이걸 이용해서

Assignment 3

Essential Visualisation: k -means algorithm

- Input data
- Initial label
- Initial centroid
- Final label
- Final centroid
- Energy per each iteration

L_1 norm의 양자

$$\sum_i \|x_i - c\| \Rightarrow \sum_i \sqrt{(x_i - c)^2}$$

이분

$$\Rightarrow \sum_i \frac{x_i - c}{|x_i - c|} \Rightarrow \sum_i \frac{2(x - x_i)}{2\sqrt{(x - x_i)^2}}$$

2분 sign function이 되기 때문에
median이 된다. object function을
만들 수 있게

what is
sign
function

$$g(z) = \frac{1}{|z|} \begin{cases} 1 & z > 0 \\ -1 & z < 0 \end{cases}$$

$\Rightarrow L_1$ norm을 사용하면
outlier 를 무시할 수 있음!