

**Lab 11: Unsupervised learning: K-means clustering & Density-Based Spatial Clustering of Application with Noise (DBSCAN)**  
**Due on November 21 11:59pm**

In this lab, you practice K-means clustering and DBSCAN as well as kernel principal component analysis.

Task 1 – K-means clustering

- Load the blobs data set.
- Perform K-means++ clustering with three clusters.
- Make a figure of distortion vs number of clusters by using the elbow method. Find an optimal number of clusters.

Task 2 – DBSCAN

- Perform clustering by using DBSCAN with the blobs data set.
- Perform clustering by using DBSCAN with the moon data set.
- Perform K-means++ clustering with the moon data set
- Discuss about the results between two methods.

Task 3– Visualization with principal component analysis (PCA)

- Once we determine the clusters of the moon data set, we can assign labels to the clusters, making it to the label data set. PCA can provide better visualization of clusters in high dimensions. Perform kernel PCA of the result of DBSCAN for visualization.

Deliverables:

- Submit your completed Python script .ipynb file and your discussion with Markdown to CANVAS
- We will run your files (Run All), so make sure all of your sections are completed.

CANVAS Submission

- Submit your files to CANVAS
  - **Name files → your Lastname\_Firstname**