

# CS:5101 Machine Learning

Term 3 (Dec 2020 - Feb 2021)

## Programming Assignment - 8

### DBSCAN Clustering

**Due Date: 7<sup>th</sup> Jan 2021 11:59 PM**

Follow the instructions given below carefully:

1. For implementing the DBSCAN method, you are allowed to use ONLY those python libraries (libraries means numpy, sklearn etc.) which have been taught till now in the ML course.
2. You must submit your code in a **single** python .ipynb notebook with naming format as follows:  
Firstname\_Lastname\_assignment8.ipynb
3. For each question, create a separate text block containing the question followed by a code block containing the solution.
4. Follow each and every instruction given in each question carefully.
5. Your code must be properly commented explaining each step clearly.
6. If any of the above instructions are not followed, penalty will be there for the same.
7. Your code and answers will be checked for plagiarism and if found plagiarised, zero marks will be provided for assignment 8. So make sure you actually code and solve the questions rather than noting down the answers.
8. NOTE: Question 1 is of 5 marks which constitutes the total marks for this assignment. Question 2 is additional for bonus of 2 marks.

## Questions

### 1. Outlier Detection

- (a) Task 1:[2 marks] Apply DBSCAN on the dataset given in data.csv to group the data into clusters and also predict the outliers (noise points). Plot in 2D the clusters obtained along with the noise points for a clear visualisation.
- (b) Task 2:[1 mark] Add a column "Outliers" in data.csv which will contain value -1 for noise points and value 0 otherwise. You have to rename data.csv to Firstname\_Lastname\_assignment8.csv and submit. Marks will be given to best 5 noise point predictions.
- (c) Task 3:[2 marks] Write a report (PDF format) with the 2D plots of before and after applying DBSCAN on the dataset and write your observations. Mention the number of unique clusters and number of noise points you obtained. Also mention the values of eps and min\_samples that you have used and why. Name the report Firstname\_Lastname\_assignment8.pdf and submit.

### 2. Image Segmentation (BONUS)

- (a) Task:[2 marks] Perform image segmentation on the given image "db\_image.jpg" using DBSCAN. Plot the original image and the segmented image and write your observations in the report from Question 1 along with the eps, min\_samples, number of clusters and number of noise points. Also mention which portion of the image was predicted as noise.