# Assignment 8

# Dimensionality Reduction

*The purpose of this assignment is to use Python to learn how to perform K-means clustering in Python.*

This assignment provides you with an opportunity to demonstrate the achievement of the following course learning outcomes:

* Understand and apply the Python programming language
* Understand and be able to understand and employ dimensionality reduction techniques.

Key Information

* **Type:** *Individual*
* **Weight:** 6.25%
* **Delivery:** Course website upload
* **Due Date:** End of lab session

## Expectations

You are expected to complete this assignment individually.

Respect for academic integrity is crucial to your success. Make sure you understand what constitutes acts of academic dishonesty in the page: [What is Academic Dishonesty?](http://mcmaster.ca/academicintegrity/students/whatis.html)

## Instructions

*Using R/Python, you are to complete the following questions.* ***Please submit your answers (CODE USED AND OUTPUT) as IPYNB* *file to the course website submission folder.***

*Each of the following questions need to be answered clearly in your solution.*

1. Import the dataset and run a complete analysis of it.
   1. Your target is column name “gnd”
2. Split the data into training and testing
3. Apply any feature selection algorithm and choose the best features.
4. Use any classifier and report the accuracy
5. Apply PCA dimensionality reduction technique to the data.
6. Plot a 2-dimensional representation of the data based on the first and second principal components, explain the results versus the known classes (display each class with a different color).
7. Repeat step 6 for the 20th and 21st components and comment on the result
8. Plot a 2-dimensional representation of the data based using t-SNE, compare the output with PCA’s output in 6 and 7, and explain the results versus the known classes.
9. Use the any classifier to classify 6 sets of dimensionality reduced data (using 2, 4, 10, 30, 40, 256 PCA components). Plot the classification error for the 6 sets and the number of chosen components.

## Rubric

To achieve full marks on this assignment, you must have answered all questions above correctly with code submitted that has no errors.