**SEP 788/789 Deep Learning and Advanced applications**

**Assignment 1**

**Professor - Hamidreza Mahyar**

[**mahyarh@mcmaster.ca**](mailto:mahyarh@mcmaster.ca)

**Due Date: 6th Feb 2022**

1. Theoretical Questions (8 marks)

* Explain PCA in details
* What is cross validation?
* Explain the difference between classification and regression
* What is Precision-Recall, F1 score, Confusion Matrix(CM) and Accuracy?

1. Programming question (12 marks)

Using the iris dataset (https://archive.ics.uci.edu/ml/datasets/iris), design and develop a machine learning model for the algorithms given below and predict the outcome. Compare the results of F1 score, CM, Accuracy and timings for various algorithms. Algorithms are:

* K-Nearest Neighbours (KNN).
* Gaussian Naive Bayes (NB).
* Support Vector Machines (SVM).

Your code should include

* Visualize the dataset and its class distribution
* Perform cross-validation
* Compare various model wrt Confusion Matrix, Accuracy and F1 score,
* Calculate timing for training and evaluation for each algorithm and compare the timings

Note: You can write algorithms for various models, however, scope of this assignment is to use built in library functions i.e. scikit-learn,