# **CVI620 – Assignment 2**

Summer 2025

| Total Mark: | 7.5 marks (7.5% of the total course grade) |
| --- | --- |
| Submission file(s): | * Python files for train and inference * Assignment2.docx (this document with some sample tests) |
| Deadline | * July 28th, 2025 |

If you are unable to complete the assignment on-time for any legit reason, please provide documentation explaining your absence (e.g., an appointment confirmation or a work letter).

Please submit the submission file(s) through Learn@Seneca. Make sure to use GitHub and provide the link to your GitHub account for all your contributions in the box below:

|  |  |
| --- | --- |
| Project GitHub repository: | https://github.com/kogun86/CIV-Assignment-2 |

**Please attach some of your test images along with any required explanations in this document.**

1. In Folder Q1, there is a dataset in which we aim to estimate the house price using two features: the number of bedrooms and the basement area.

Use Multiple Linear Regression for this task. Display the coefficients of the model and calculate the MAE (Mean Absolute Error) and MSE (Mean Squared Error). Search about RMSE (Root Mean Squared Error) and explain the trade-offs between these metrics. Finally report RMSE score of your model.

Perform this task using both LinearRegression and SGDRegressor.

Additionally, study the MAPE (Mean Absolute Percentage Error) metric using [this link](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_percentage_error.html), and apply it to evaluate your model.

* Mean Absolute Error
  + Measures average absolute difference between predicted and actual values.
  + Easy to interpret
* Mean Squared Error
  + Penalizes larger errors more heavily due to squaring.
  + Useful when large errors are worse than small ones.
* Root Mean Squared Error
  + Square root of MSE. Has the same units as the target.
  + More interpretable than MSE
* Mean Absolute Percentage Error
  + Error as a percentage, making it scale-independent.
  + Problematic when actual values are near zero

1. For the Cat and Dog dataset provided in the Q2 folder, perform classification using all the methods you know and try to achieve the best possible result. Compare the algorithms carefully and tune the parameters so that the best result can be obtained.

Save the trained model and test it on several images from the internet. Was the model able to correctly predict the images?

The model was able to recognize some images correctly, but the average accuracy score of 0.6 is too low to rely on the model's results, because with a random choice of 2 options, the average accuracy score would be 0.5.

1. The MNIST dataset is one of the most well-known datasets in the field of image processing. It contains 60,000 images related to handwritten digits from 0 to 9 and is provided as a CSV file in the Q3 folder. In this file, each image is represented as a flattened vector. Classify this dataset using different methods and try to achieve at least 90% accuracy.