# **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: |  |

**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.

**Definations**

**MAE:** the average of difference b/w actual and predicted values

**MSE:** the average of the squared differences between actual and predicted values

**RMSE:** the square root of the average of the squared differences between actual and predicted value

**Interpretation**

which one to use is largely dependent on the dataset and the problem. I was able to find the followin when searching about these three errors:-

*outliers are something in a dataset that are largely different from the rest of the data*

**MAE:** is less sensitive to outliers

**MSE:** is more sensitive to outliers and gives huge error value due to large outliers, a few articles also list that this changes the unit because it squares the error but i did not understand what they were trying to say by it

**RMSE:** is also sensitive to outliers and gives more weight to larger errors it is in the same unit

(again i not sure what they mean by this statement about unit)

https://www.youtube.com/watch?v=fk7bzKFDmk8&t=576s

https://dev.to/mondal\_sabbha/understanding-mae-mse-and-rmse-key-metrics-in-machine-learning-4la2

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 predictions are correct as the accuracy score for both the models was pretty low for LR model it was 60% and it was 50% for KNN  
Below is the results from both models for a bunch of cat and go images from the internet

A screen shot of a computer

AI-generated content may be incorrect.

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.