# **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/Silpar55/CVI-ML> |

**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.

MAE, MSE and RMSE are great formulas to calculate the error margin in our model,

however, each of them provides different ways to understand the behaviour of our model

and we need to know which one give us a better understanding based on our goal

MAE is simple and tell us the amount of range that our prediction is wrong, for example,

if our prediction is 10 and our MAE is 5, it means that the real value probably will be between

5 - 15. This is easy to understand but treats every error equal, good for robust projects but bad for other sectors

MSE gives an exaggerating error value based on how big the margin is, for example, if the error margin is 5, the MSE is

25, this might be a lot but if we compared to an error margin of 10, we'll see that MSE is 100, this provides that bigger

erros should be penalize more than small ones which is good for optimization but sensitive.

Finally, RMSE Offers a little bit of the exaggeration of MSE and the robust of MAE, it penalizes bigger errors than small ones

but with less priority than MSE.

MAPE is another way to display MAE, and it gives a more intuitive way to measure MAE, instead of saying that the prediction

is off 1000, we can say that the prediction is off 20%, this is better since the error is always relative to the preds

A screenshot of a computer error

Description automatically generated

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 models trained correctly predict my images

A collage of a dog and a cat

Description automatically generated

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.

Using KNN

A black background with white numbers and symbols

Description automatically generated

Using Logistic Regression

A black background with white text

Description automatically generated