COMP 472 Artificial Intelligence Summer 2025

Mini project 1

Due date: as indicated on Moodle

You can work in a group of a minimum of 2 or a maximum of 3. One submission per group please. 50% off if working alone or in a group that has more than 3 members.

Objective:

To get familiar with Python and some of its AI libraries.

Instructions:

In this project, you will develop a machine learning app that classifies handwritten digits (0–9) from the MNIST dataset. You will need to use at least 2 libraries:

1. **NumPy** for data exploration and manipulation.
2. **scikit-learn** to build a simple baseline model using logistic regression.

Python has a dataset of about 1800 grayscale images of handwritten digits. Please find more info about it and use it in this project. NumPy will help you in storing and **normalizing the data**. You may find **matplotlib** useful to show how the images look like.

When you build a classification model, you split your data into training data and testing data. The split is usually 80% for training and 20% for testing.

Use the built-in LogisticRegression model. Do a bit of research to find out what is LogisticRegression and make sure to have a brief idea about how it works. Only the basics; do not get into the details.

You need to evaluate the performance of your classifier. How accurate is it in recognizing handwritten digits? Find out more about the classification_report and the confusion_matrix. Make sure to print them on the screen to get an idea about the performance of your app.

Documentation and Presentation: Document your program's functionality, and code structure. You will give a demo to your TA. The TA's will post time slots on Moodle for the demos. All group members must be present during the demo. All team members must be familiar with all parts of the code. The TA will

ask individual team members specific questions about the code. Each student must show complete understanding of the submitted code and what each line of code is doing.

You must submit the code on Moodle once you are done as a zip archive. The name of the archive must be of the following format:
<Student1 first name Student 1 last name> <Student 1 ID>-<Student2 first name Student 2 last name> <Student 2 ID> - <Student3 first name Student 3 last name> <Student 3 ID>.zip

You must code this project in Python.

Enjoy, lean and have fun 😊