EEN1072 Data Analysis and Machine Learning II

Release Date = 04-03-2025
Due Date: 10-04-2025 (5pm)
Total Marks: 25

This assignment is designed for students to demonstrate their ability to perform data analysis and machine learning with an aim to gain hands-on experience. You are required to demonstrate your ability to apply machine learning to resolve some real world problems. The assignment will be evaluated based on your ability to demonstrate a complete understanding of the data analysis and machine learning processes rather than the accuracy of the model. You are required to design the neural network architecture, optimize performance of models by fine tuning various parameters and calculate accuracy of your trained model. You have to document your data analysis and machine learning journey starting from data sourcing to pre-processing, analysis, training and validation of your solution.

Assignment Tasks:

Q1: Convolution Neural Networks for Objects Detection

[50 Marks]

This question is related to implementation of Convolutional Neural Network to automatically detect objects in any given image. This is an open world problem however students are free to use a limited set of objects for the detection e.g. you can select a set of objects and only those objects will be detected. You are free to take any dataset available online for the object detection, however, it is

expected that you have at least 1000 data points (images) used for the training the model. You are also expected to collect a separate dataset of at least 100 data points (images) for testing the model's accuracy.

Please note that you are not allowed to take any pre-trained object detection neural network models and you are required to design your own convolutional neural network from scratch.

The breakdown of the tasks is as follows;

1.1. Briefly describe the dataset, data points, data source, labels and training/testing/validation split.

(Report) **[5 Marks]**

1.2. Apply different data pre-processing techniques e.g. data cleansing, data reshaping, data normalisation/scaling, augmentation and preparation.

(Report + Code) **[10 Marks]**

1.3. Design a convolutional neural network and briefly explain its architecture including rationale for all architectural decisions e.g. how many hidden layers and why?

(Report) **[5 Marks]**

1.4. Train a neural network model for object detection, test and validate it using the appropriate metrics for performance evaluation.

(Report + Code)**[20 Marks]**

1.5. Evaluate performance of your trained models and introduce at least two different variations to your model architecture for performance improvement.

(Report + Code) **[10 Marks]**

Q2: Recurrent Neural Networks for Trajectory Prediction

[50 Marks]

This question is related to implementation of Recurrent Neural Network to automatically predict trajectory using the previous sequence of input in real-time. You are free to use any kind of sequential data for the trajectory prediction e.g. car, ships, people (indoor or outdoor). However, the selected trajectory data must have some sort of sequential input of geo-location/coordinates. You are free to take any dataset available online for the trajectory, however, it is expected that you should have at least 1000 data points (trajectories e.g. a mix of short or long trajectories) used for the training the model. You are also expected to collect a separate dataset of at least 100 data points (trajectories) for testing the model's accuracy.

Please note that you are not allowed to take any pre-trained trajectory prediction neural network models and you are required to design your own recurrent neural network (LSTM or advanced) from scratch.

The breakdown of the tasks is as follows;

1.1. Briefly describe the dataset, data points, data source, labels and training/testing/validation split.

(Report) **[5 Marks]**

1.2. Apply different data pre-processing techniques e.g. data cleansing, data reshaping, data normalisation/scaling, augmentation and preparation.

(Report + Code) **[10 Marks]**

1.3. Design a recurrent neural network and briefly explain its architecture including rationale for all architectural decisions e.g. how many hidden layers and why?

(Report) **[5 Marks]**

1.4. Train a neural network model for trajectory prediction, test and validate the trained model using the appropriate metrics for performance evaluation.

(Report + Code)**[20 Marks]**

1.5. Evaluate performance of your trained models and introduce at least two different variations to your model architecture for performance improvement.

(Report + Code) **[10 Marks]**

Submission Guidelines

The final submission should be submitted on the loop before the due deadline. A final submission should contain;

1. A report (PDF or Word document) documenting all assumptions, design decisions, and findings. You can include any diagrams, visualisations, plots, and tables in the report. You should strive to make your work completely reproducible using only the report document: include details on everything you tested and all results. Document and justify all design decisions.

2. Two or more separate notebooks (at least one for each question) containing all of your relevant code. In case you have tested multiple versions of the datasets (e.g. after cleansing) or trained different models for the same task, do not overwrite the previous code. You can either submit multiple versions of the notebook or separate code sections for each iteration/repetition of any task.

3. Datasets used for training and testing must be hosted on github accounts and your notebook code must be able to directly connect to the datasets hosted online. You are free to use any other online platform for data hosting, but it is up to the students to make sure that their code should work irrespective of where the data is hosted.

Important:

- Do **NOT** include code as images (e.g. screenshots of code) in your report. Include code snippets as text. Your report must be submitted as word doc or PDF and not as a zip. A zip submission of the file will not be accepted.
- Your notebook must have relevant comments in each section of the code clearly demonstrating your complete understanding of the code. A notebook without any comments will receive penalty points.

Plagiarism: Please read and strictly adhere to the DCU Academic Integrity and Plagiarism Policy. Note that reports are automatically checked against each other and against external web sources for plagiarism. Any suspected plagiarism will be treated seriously and may result in penalties and a zero grade (see Sec 6.2 of the DCU Academic Integrity and Plagiarism Policy). You are not allowed to copy any code (in full or parts) from jupyter notebooks provided by the dataset owner or third party for their own analysis. In case of suspected code plagiarism, you will be asked for an interview to demonstrate your programming skills and ability to code independently.

Grading: The assignment is worth 25% of the overall mark for the module. Marks will be awarded based on the quality of the resulting report. In particular, I will be checking to see if you are handling data correctly, carrying out exploratory analysis to gain insights, correctly performing model selection, and critically, documenting everything in a clear and concise way. The submitted code will also be checked to ensure that the work is your own.

Late Submission: Please note that any last minute requests to extend assignment submission deadlines will not be entertained. Loop submission for assignments will be automatically disabled at the end of the deadline **10-04-2025 (5pm)**. There will be a separate submission option available for the late submission which will incur a **penalty of 30%** on the awarded grades. Loop will remain open for late submission until **13-04-2025 (5pm)**. No submissions will be accepted after the late submission deadline.