COP528 Applied Machine Learning Coursework

In this coursework you will apply what you have learnt in this module to solve real-world problems, both with classical machine learning and popular deep learning methods. The coursework has two tasks, the first task is about designing a pipeline to use ML methods to solve a predefined task (60%), and the second task is about using a Convolutional Neural Network (CNN) for image analysis (40%).

Task 1: Machine Learning Pipeline (60%)

Design a pipeline, an evaluation strategy, and a set of experiments to determine the best parameters and machine learning algorithm, based on the results of empirical evaluations derived from a dataset (for achieving this, you could compare different algorithms if needed). Data used for this task: UCI machine learning repository contains many ML datasets for various applications. You are free to select a dataset (collected and uploaded to UCI repository after 2018 where you can check the year column of the dataset) from the website (https://archive.ics.uci.edu/ml/datasets.php).

Task 2: Deep Learning for image classification (40%)

You will be provided with an image dataset, where each image contains meaningful objects, e.g., parachute, oil box and truck etc. You could either use an existing CNN network architecture or design your own CNN network for classifying test images into the predefined classes.

Task 1 and 2 submission guidelines

Submission: The code, report and reflection must be submitted on 12:00AM Friday [18th March 2022] of this module.

- Project Report: PDF, up to 6 pages (IEEE double column format, template will be provided) on Learn.
- Software: Jupyter Notebook.

Assessment Criteria: The project will be marked based on the code quality and the report (as described below). The report will have two parts: for each of the two tasks in the coursework, the marking criteria is given as below (Task 1: 60% and Task 2: 40%).

Code:

- **Code Quality (10%):** You need to submit clean, structured and well commented code so that the instructor could just change the test file name to run your code and get the evaluation results.

Report:

- **Introduction (10%)**: Describe the problem you are working on, application background, the machine learning task(s), and an overview of your results.

- **Data and preliminary analysis (10%)**: Briefly describe the dataset. You could use some visualisation or statistical methods to make assumptions that will influence your design.
- **Methods (15%)**: Present your machine learning approach for solving the task. The proposed approaches should be evidenced by a working piece of code of your software. You should demonstrate that you have applied ideas and skills. It may be helpful to include figures, diagrams, or tables to describe your method.
- **Experiments (25%)**: Present and discuss the experiments that you performed for the task. You may show what things you tried, what hyperparameters or architectural choices you tested, model training and evaluation strategies, what is your best model, impact of various components of your system. Justify the methods/parameters when applicable. Quantitatively evaluate and/or compare your results, e.g. performance metrics, statistical tests, learning curve and plots. You should include graphs, tables, or other figures to illustrate your experimental results.
- **Reflection (20%)** Summarise your key results, what problems did you encounter? What are the good findings and what could make your performance better if you try it again?
- **Writing/Formatting/ Referencing (10%)** Is the report well-structured, clearly written and nicely formatted? Is the technical content presented at the right level, concise, and focused? Is the code well formatted with helpful commenting when necessary?

Further Information:

- You are recommended to use Python scikit-learn, TensorFlow and Keras for implementation. You can use data pre-processing code blocks that are available online (to avoid plagiarism, ensure that you provide the source and acknowledge the author of the code you have used. You must add this information in a comment accompanying the code fragment that you borrowed. If you have adapted the code, state "Code adapted from: [provide the source]". If you have copied the code and have made no changes to it, state "Code copied from: [https://uark.libguides.com/CSCE/CitingCode]").
- It's ok if your results are not "good". What matters is that you demonstrate your knowledge and effort made to gain a good understanding and practical skills related to machine learning in detail through the project.
- The report structure and contents are indicative. Components which are relevant to your project should be demonstrated. Some questions may be standard for your project and only a brief mention is enough. You do not need to address all of them in full detail.
- You may consult any textbooks, online resources, or publicly available implementations for ideas and code that you may adapt into your strategy or algorithm. You need to clearly cite your sources in your code and your writeup. You should not use another students' code for the class for your project.
- IT contact: science.it@lboro.ac.uk.