Summative Assignment

COMP2261 Artificial Intelligence – Machine Learning 2021/2022

Deadline for submission: 17 January 2022 14:00 (GMT)

1   Introduction

You are to build and compare predictive models using the data provided below. You are to define a machine learning problem, i.e., the input/output, and what they can be used for.

2   Data

Download: https://tinyurl.com/COMP2261MLDATA

3   Submission and Marking Scheme

3.1   Jupyter Lab/Notebook file (0%)

The file should have the name of [your_Banner_ID].ipynb, and it should include all the codes you used for the project. Comments should be given when necessary. It is not going to be marked, but no-submission or not-working code will result in a zero mark for the whole summative assignment.

3.2   PDF report (100%)

The file should have the name of [your_Banner_ID].pdf, max 4 pages (including everything). Template: https://tinyurl.com/comp2261mltemplate

3.2.1   Title, Abstract, Keywords, Introduction (12%)

Provide a brief introduction to the project. Precisely define the problem you are solving, i.e., formally specify the inputs and outputs. Frame the problem as a machine learning task. Explain why this task is interesting and important.

3.2.2   Exploratory Data Analysis and Data Preparation (24%)

Summarise the main characteristics of the dataset, using tables and statistical graphics, and/or other data visualisation methods. Describe how you split the dataset and present stats such as count, mean, etc. Describe how you constructed and/or transformed the data/feature.

3.2.3   Learning Algorithm Selection (8%)

Choose three machine learning algorithms, which are not necessarily taught in this submodule. They cannot be based on deep learning or reinforcement learning. Explain why the chosen algorithms are appropriate for the project.

3.2.4   Model Training and Evaluation (32%)

Describe the training process, including the parameters involved and how they fit, concerns about underfitting and/or overfitting, and concerns about the convergence of the optimisation. Describe the hyperparameter selection and tuning process, including the hyperparameters involved and how they were selected and tuned, the candidate values that were considered, and the performance metric that were used for optimisation. Describe the evaluation metrics and explain why they are appropriate.

3.2.5   Model Comparison (15%)

Describe how you compared the trained models, using tables and statistical graphics and/or other data visualisation methods.

3.2.6   Conclusion and Discussion (9%)

Summarise the project: what the project was about, what you did, what the results were, the major limitations of the approach you used, what could have been done to improve the procedure and the result, the lessons you learned through this project.

3.2.7   References (0%)

List all the external references.

Note: for any decision you made, provide clear and appropriate justifications. Marks are generally awarded for a good experimental procedure that supports the results, not for achieving the best performance. **Deduction** of the final mark may be made for not submitting files as requested, e.g., file names/extensions, report formatting (referencing, captions, etc.)