CM3015 Machine Learning and Neural Networks

Project Idea Title 1: Deep Learning on a public dataset

What problem is this project solving?

Choosing, based on a quantitative evaluation, a well performing machine learning model for used with a publicly available dataset.

What is the background and context to the question above in 150 words or less?

Pick a dataset from Kaggle.com – choose one that interests you or you think is important – for example: tweets, faces, lung scans, skin diseases, student grades…

Develop a deep learning classificatation/regression model for your chosen dataset by following the methodology of Deep Learning with Python. Aim to find the best model – work from simple to deep and employ the advanced techniques of Chapter 7.

List some recommended sources for students to begin their research

- F.Chollet, Deep Learning with Python, 1st ed.
- Kaggle.com

What would the final product look like?

(e.g. presentation, usability, functionality, results)?

A research project – the final product is a report

What would a prototype look like?

What would it show?
What does it need to prove?
*What **IS** important to make clear?*
*What is **NOT** important at this stage?*

The prototype is a baseline model that achieves a common sense prediction.

It is not important to achieve the accuracy of any puplished paper on this dataset (or any of the Kaggle public notebooks)

What kinds of techniques/processes are relevant to this project?

Jupyter notebooks

Tensorflow, matplotlib and associated Python libraries

What would the output of these techniques/processes look like?

Model code

Validation plots

Prediction on test set

How will this project be evaluated and assessed by the student (i.e. during iteration of the project)?

What criteria are important?

Does the model significantly improve on a commonsense baseline Have

I investigated all the alternatives

For this brief, what would a minimum pass (e.g. 3rd) student project look like?

- Any original model that runs and produces a prediction
- A basic evaluation of the model on the public dataset
- Report is well-strutured

For this brief, what would a good (e.g. 2:2 – 2:1) student project look like?

In addition to minimum pass criteria:

- A sequence of orginal models of increasing depth and sophistication
- An evaluation of the different models using the public dataset, which makes it possible to draw conclusions about the effectiveness of different models and choose a preferred model
- Report: Correct application of the DL methodology; good standard of written, technical English

For this brief, what would an outstanding (e.g. 1st) student project look like?

In addition to the good criteria:

- Replication of a hight quality published model(s) on the chosen dataset
- An evaluation of the models to the standard of academic research
- The report is a self-contained explanation and account of theory and experiment. There is a literature review and the work is contextualised. Critical comparison of best model and reimplemented model(s) and results in the literature.