TU Dublin, Tallaght Campus
Department of Computing

Applied AI and Deep Learning – Full Time and Part Time 2019
CA2 (25% Weighting)

CA Overview:

Students will be given a unique data set, based on random values using the SKLEARN blob dataset creator. Each student will be assigned parameters to create their data set for use in this CA.

Students will be given a unique combination of:
- Standard deviation of the dataset   (Range: 100 – 150)
- Features   (Range: 50 – 100)
- Random Number   (used for blobs and NumPy seed)

Data set features same for all students:
- 5 classes
- 100000 instances

CA Points:

Students will be given randomly assigned set of values. Students will be expected to develop the deep artificial neural network with the best performance (including addressing underfitting and overfitting), from the beginning to a final saved model.

All steps should be separated in a Jupyter notebook (sections), and three versions of the notebook are required on final upload of the elapsed CA: PDF, HTML and the .ipynb Jupyter notebook.

Note: The PDF will be run through turn it in, for plagiarism evaluation, where the institution's marks and standards will be applied.

Students are expected to discuss each piece of work, expanding on the rationale for selection of techniques and the processes applied to get to the final model, including the use of any visual aids such as tables, plots etc.... The word count is a minimum of 2000 words.

Code on its own, will not be considered as a passing grade.

The is an additional PDF for instructions on how to create the unique data set using the make_blob() function.

Please add your name and student number as the first block in markdown on the notebook.

CA Headings:

The CA will be graded under the following headings. Where each heading should have a rationale and description expanding the approach and its findings, code that executes the work, and in some cases a visual aid to further compound your findings. Use standard Jupyter notebook markdown, to crate headings and sections. The following headings should guide your project:

- Introduction
- Creating the dataset
- Data exploration
- Data pre-processing
- Model selection and training testing technique selection
- Model exploration to determine network topology
- Initial hyperparameter investigation (learning parameters and optimization)
- Grid search
- Final Model Presentation and performance evaluation
- Saving model and weights
- Opening model and weights and running model on new unseen data
- Summary and Conclusion

CA Notes:

At all stages, precautions must be taken (such as checkpointing) to ensure work is competed by the CA deadline.

It is up to the student to select the notebook platform (such as locally or on Azure notebooks) and it is the student's responsibility to ensure that the platform has the required version of Python, Anaconda and required libraries.

Items such as the grid search results and model epoch training results should be left visible in the notebook (that is verbose is not applied) so that they are in the final uploads.