COMP219 - 2018 - First CA Assignment
Individual coursework
Simple Machine Learning Model

Assessment Information

Assignment Number	1 (of 2)
Weighting	10%
Assignment Circulated	Wednesday 10 October 2018
Deadline	Wednesday November 21 2018, 15:00
Submission Mode	Electronic
Learning outcome assessed	2. Ability to choose, compare, and apply suitable basic learning algorithms to simple applications; 3. Ability to explain how deep neural networks are constructed and trained, and apply deep neural networks to work with large scale datasets
Purpose of assessment	To implement machine learning algorithms on a dataset
Marking criteria	The marking scheme can be found in Section 3
Submission necessary in order to satisfy Module requirements?	No
Late Submission Penalty	Standard UoL Policy.

I enforce a "no error policy" in this module: If your code does not compile, your mark will be capped at 40%. Thus, you may get a higher mark for an incomplete solution than for an advanced sketch.

If you want to show me your attempt to add some features that does not compile TOGETHER with your working code, please feel free to submit **two** ZIP files clearly indicating which one of them contains working code and which contains an incomplete one. In this case, you will not be penalised and you can get a higher mark.

1   Objectives

This assignment requires you to *implement* and *evaluate* one (or multiple) simple machine learning models on a dataset.

2   Requirement and Description

Language and Platform   Python (version 3.5 or above) and Tensorflow (newest version). You may use any libraries available on Python platform, such as numpy, scikit-learn, panda, etc.

Dataset   You can use any dataset which is convenient for you. It is recommended that you select one that is different from those frequently used in various places, such as Iris. Unless exceptional circumstance, it is recommended that the dataset is not too small (e.g., no less than 200 items) and not too big (e.g., no more than 100,000 items). There are a few suggested repositories where you can find plenty of datasets:

- scikit-learn toy dataset: http://scikit-learn.org/stable/datasets/index.html

- UCI machine learning repository: `https://archive.ics.uci.edu/ml/datasets.html?sort=nameUp&view=list`

- `https://www.springboard.com/blog/free-public-data-sets-data-science-project/`

Learning Task   You can choose either classification (preferred) or regression.

Learning Model/Algorithm   You may choose at least one learning algorithm from the following list (but not limited to):

- decision tree learning

- SVM

- naive Bayes

- (deep) neural network

- k nearest neighbor

Assignment Tasks   The implementation task (as suggested in the Objectives) is to learn a model from the dataset you select.

The evaluation task is to apply model evaluation on the learned model. For the materials on model evaluation, you may take a look at the metrics explained in the lecture "model evaluation", e.g., accuracy, error, confusion matrices, cross validation results, etc.

You need to write a proper document explaining the above two tasks.

Submission files You submission needs to contain the following two files:

- a package containing your source codes (with the instruction on how to run them) and

- a document explaining your implementation and model evaluation results.

3 Marking Criteria

The assignment is split in a number of steps. Every step gives you some marks. At the beginning of the submitted document, please include a check list indicating whether the below marking points have been implemented successfully. Unless exceptional cases, the length of the submitted document needs to be within 4 pages (A4 paper, 11pt or 12pt font size).

Step 1: Loading Data 10%

Successfully load the dataset and use python commands to display the dataset information, e.g., the number of data entries, the number of classes, number of data entries for each classes, etc.

Step 2: Training 30%

Successfully train a model. This step is further divided into smaller sub-steps.

Step 2.1: code for training, 10%

You write a code that is dedicated for training task. To get this mark, you need to have a clear comments in your code explaining that which part of the code is for this task.

Step 2.2: successful training, 20%

Training can be done successfully. Here you need to have a simple test command to validate the trained model. For example, one can randomly select a few samples from the test dataset to the predications of the trained model against the true labels.

Step 3: Model Evaluation 40%

Apply model evaluation method to evaluate the simple machine learning model you trained. This part includes the following two aspects:

Step 3.1: explain your experimental design, 20%

Here you need to explain which method you are using, and how you design your evaluation experiments.

Step 3.2: document your evaluation results, 20%

You can get your mark of this part if you write down your experimental results.

Extra 20%

You can see that marks for the steps described add up to 80%. In order to get 20% extra you need to train more than one models, and compare them in the model evaluation. You may be able to see e.g., one model is better than the other in terms of some metrics.

4   Deadlines and How to Submit

- Deadline for submitting the first assignment is Thursday, 21 November at 3pm.

- Submission is via the departmental submission system accessible (from within the department) from
  `http://intranet.csc.liv.ac.uk/teaching/modules/module.php?code=COMP219`.
  Please export your project and submit the ZIP file.

  If you want to show me your attempt to add some features that does not compile TOGETHER with your working code, please feel free to submit **two** ZIP files clearly indicating which one of them contains working code and which contains an incomplete one. In this case, you will not be penalised and you can get a higher mark.