Coursework I FAQs

Nirmalie Wiratunga, Sadiq Sani and Kyle Martin

October 8, 2018

1 What am I expected to do for CW1?

For full details of the coursework, please refer to the coursework specification document on Moodle. However we summarise a few key points next.

In coursework 1, we are expecting you to demonstrate an understanding of two algorithms; the Artificial Neural Network (ANN) and k Nearest Neighbour (k-NN). We would like you to demonstrate a full understanding of the theory behind these algorithms, and show that you know how to adjust their hyperparameters to find an optimal solution to a given problem. This will culminate in a hybrid system which combines both algorithms.

When submitting, you should submit both a written report and your full working code in your zipped submission. The code should be structured as three ipynb files - ann.ipynb, knn.ipynb and hybrid.ipynb. The report should be broken down as follows.

- **Part 1 - Introduction:**

  1. Select two datasets from the provided list.
  2. Discuss your choices and how you will split them into train and test sets, considering what factors have led to this decision.
  3. Note that all of the sections below should be completed for **both** datasets.

- **Part 2 - The ANN Classifier:**

  1. Paragraph detailing students understanding of hyperparamaters.
  2. A graph for each hyperparameter comparing the different values.
  3. Paragraph explaining the results.

- **Part 3 - The k-NN Classifier:**

  1. Paragraph detailing understanding of weighted and unweighted kNN.
  2. Graph comparing weighted and unweighted kNN at different values of k.
  3. Paragraph explaining the results.

- **Part 4 - The Hybrid Classifier:**

  1. Paragraph detailing understanding of the hybrid system functionality.
  2. Table comparing the results of the ANN, kNN and hybrid.
  3. Paragraph explaining the results.

There is no need for a conclusion - the report should finish with the discussion of your hybrid classifier results.

2 How do I create the graphs for the CW?

We encourage the use of the matplotlib package when creating the graphs for your coursework. Though we will not penalise the use of Excel (or similar packages), you will find it much easier to export your results directly as a graph, rather than saving them to a file before exporting them to a spreadsheet.

3 What template should I use for the written components of the CW?

A template has been provided for the written component of the coursework. It is on moodle and in a word format. Additionally, we have provided formatting guidelines to keep you right. You should use the existing template (see cm4017-cw1-template.doc), rather than attempting to recreate it from scratch.

4 Am I expected to comment my code?

Since we have provided the code which will act as a basis for your experimentation, we will already have knowledge of its functionality and background mechanics. Therefore, we do not require strict documentation or commented code. However, if you feel that any of the methods you have adopted or code you have used is confusing or unclear, it would be helpful if you could comment that part. This will help us to understand your thought process, and might enable us to grant marks for that (even in situations where the code itself may be incorrect).

Please Note: The process of your experimentation should be completely clear from your report for both pieces of coursework - we should not have to refer to the original code for clarity. **Documentation of code should never be used as an excuse for an unclear report.**

5 If the datasets which you use provide a test and train file, is there then no requirement to split them into test and train data?

That is correct. If train-test split is already provided then you can certainly use them as is. In some situations (depending on the power of your machine) you may have to work with a smaller samples of the train or test sets (as we had to do in Lab 3 with the training set). What ever strategy you choose simply state it clearly and use the same strategy when applying different configurations of your algorithm to that dataset.

For the more keen and ambitious student; you might even want to work with several disjoint train - test splits and average the accuracy results over several runs.