

CM20252 & CM50263 Artificial Intelligence

Coursework: Programming 3

Özgür Şimşek

Date set: 23 April 2018

Date due: 7 May 2018, 11:00 am

Total marks: 100

CM20252 students: This coursework will determine 20% of your mark for the unit.

CM50263 students: This coursework will determine 16% of your mark for the unit.

Where to submit: CM20252 Moodle page

What to submit: Completed Jupyter notebook (.ipynb file)

You will be given a Jupyter notebook to work with. You must follow the instructions on this notebook and submit this particular notebook. Otherwise you will receive 0 marks.

Late submissions: We will follow the university policy on late submissions.

Coursework submitted after the deadline will receive a maximum mark of 40 (out of 100).

Coursework submitted after five working days will receive a mark of zero.

Feedback: Within two weeks of the submission deadline, your Jupyter file will be returned to you (via Moodle), showing the marks you received from each part of the coursework. You can get additional feedback from the unit leader or one of the tutors via appointment.

You are required to work individually.

This coursework will be marked anonymously. Please do not include any identifying information on the files you submit.

Do not plagiarise. Plagiarism is a serious academic offence. For details on what it is and how to avoid it, please visit the following webpage:

<http://www.bath.ac.uk/library/help/infoguides/plagiarism.html>

Part 1 (45 marks)

You will build a spam filter using the naïve Bayes classifier. You will be given two data sets, one for training your classifier and one for testing your classifier. In the Jupyter notebook, you will find detailed, precise instructions that will guide you in developing your classifier. This part of the coursework is expected to take about four hours (or less) but please note that your personal experience may differ from this estimate.

Part 2 (55 marks)

You will develop a classifier for a different problem, given a set of training instances (called the *training set*). Your classifier will be evaluated on a different set of instances (called the *test set*).

Please note that the test set will not be available to you while you are developing your classifier. In addition, you will not know what the best possible level of performance is.

To develop your classifier, you are welcome to use any algorithm you find suitable but you must understand the algorithm fully. You may choose to develop a Naïve Bayes classifier, similar to the one you developed in Part 1.

Your mark will depend on (1) how well your classifier performs on the test set, and (2) a short discussion of your classifier and of the process of developing a classifier (for example, you can discuss alternative classifiers you experimented with).

With a good implementation of Naïve Bayes, you will receive 15 marks. If you executed Part 1 correctly, this will give you a total of 60 marks, using Naïve Bayes alone.

Each of the following will allow you to receive additional marks: high level of accuracy in the test set (the higher the accuracy, the higher your mark will be), experimentation with alternative classifiers, and evaluating (and discussing) how various classifiers perform as the size of the training set increases.

Additional details are provided in the Jupyter notebook.

You must follow the instructions in the Jupyter notebook precisely.