18s1: COMP9417 Machine Learning and Data Mining

Assignment 2 – Introduction

May 7, 2018

Aims

Learning objectives of this assignment:

- a self-selected task to extend aspects of the course material

- involves practical aspects of the machine learning problem, i.e.

    - implementing or modifying algorithms and/or
    - experimental evaluation of algorithms on data set(s)

- exercise written communication skills in motivating, recording and summarising work done on a specified task

Submission

The hand-in for this assignment has two parts:

- files containing program code to do something interesting with data set(s) and/or results of running programs on data set(s)

  – compressed archive of files

- a report on what you did.

  – single document in PDF format

NOTE: there is a default **2MB limit** on the size of `give` submissions. If this is a problem then let me know and this can be changed or an alternative submission method will be arranged.

Marking

Total: 30 marks available (rounded up to the nearest integer)

Part 1: [12 marks]

8 marks: solving the basic problem as described in the topic
4 marks: extra features, or 1 person solving most or all of a > 1 person problem

Part 2: [10 marks]

6 marks: describing the problem and your solution
4 marks: good presentation and communication of results

Balancing: [8 marks]

5 marks: difficulty
3 marks: achievement

Marking

Why have balancing ? To take account of different levels of difficulty of the tasks undertaken, and different levels of accomplishment of those tasks.

The difficulty assigned to each project in the spec is added to one of the three levels of achievement judged by the overall impression obtained from the combination of Parts 1 & 2:

difficulty	achievement	
5	high	3
4	good	2
3	acceptable	1
2		
1		

Part 1

Marks will be gained by:

- evidence of good design or planning by breaking down the problem into sub-components

- rigorous collection of results

- use of comments and notes to record decisions taken and reasons for them in the process of the work

Part 1

Marks will be lost by:

- programs failing to compile or run

- missing results files

- no clear information on contents of files submitted (e.g. in README)

- evidence of plagiarism

Part 2

Marks will be gained by:

- evidence of thorough testing of an idea

- good presentation and summarisation of key results using tables, graphs, etc.

- simple, clear and relevant explanations

- well-formatted, well-organised, spell-checked and grammar-checked documents

Part 2

Marks will be lost by:

- inappropriate length (aim for around 2 pages per group member — extra figures, tables, etc. can go in an appendix of *reasonable* length)

- digression, rambling or waffling to fill space unnecessarily

- errors or inconsistencies in presentation, such as
  - incorrect description of algorithms or their properties
  - poor algorithm selection for a task
  - errors in evaluation like not using an independent test set or cross-validation if this is required
  - statements or conclusions not based either on your experimental results or referenced sources
  - incorrect or inappropriate use of statistical tests

- evidence of plagiarism

Deadline

Sunday June 3 23:59:59 (contact LiC if you need an extension)