

Course Project

ICS2207 - Machine Learning: Classification, Search and Optimisation

Submission Checklist - Very Important (seriously)	
<i>Failure to satisfy these submission requirements may result in non-acceptance of your submission or reduced marks.</i>	
	The deadline is strictly on Friday 20 th January 2023 at 23:59.
	You included the complete plagiarism declaration form.
	You included the completed statement of completion (template below).
	Report is in PDF - no Word documents or any other format; no exceptions.
	Source code is included in the submission. No links to Dropbox, GitHub, or anything else; no exceptions.
	Archives are in ZIP format - no RAR, 7z, or any other format; no exceptions.
	Uploaded size limit is 100Mb - the PDF report, source code, and any relevant datasets must fit. If there is no space, you may provide an external link to the datasets ONLY (e.g., Dropbox, GitHub, etc.). Report and source code must <u>always</u> be in your VLE submission. Don't let external links expire until you have been graded.
	Your name and student ID are both on the front page of the report.
	Projects must be submitted only through VLE – submissions made by email or any other way apart from VLE will not be considered; no exceptions.
	A draft and final submission area is set up in VLE. Only projects submitted in the final submission area will be graded. Projects submitted to the draft area will not be considered at all; no exceptions.
	It is your responsibility to ensure that your upload is complete, valid, and not corrupted. You can reupload the assignment as many times as you wish within the deadline. Double-check! Corrupted uploads cannot be graded.
	Plagiarism is a serious offence and will not be tolerated.
	This is NOT a group project.

Designing a keyboard using genetic algorithms

- This project is based on “Using AI to Create the Perfect Keyboard” described on YouTube here:

<https://www.youtube.com/watch?v=EOaPb9wrgDY>

- Like the above project, you must also implement a genetic algorithm to help design an ‘optimal’ keyboard. However, you are also required to experiment with the following alternatives and ideas:
 - Use a **different text corpus** (the original project used the arXiv abstract dataset). As a suggestion, you may want to experiment with books available in plain text. You’ll find plenty at <https://www.gutenberg.org/>.
 - In addition to the crossover operation described by the original author, experiment with **two-point crossover** (partially matched crossover may be useful).
 - The original author did not describe a **mutation operation**. Design one, use it, and see what happens.
 - Experiment with using **elitism**.
- You must evaluate your work well. Make sure that you study the effect of population size, crossover types/rates, mutation types/rates, comparison to QWERTY and/or AZERTY layouts, use of elitism, and so on.
- You are required to write a high-quality report. Make sure to:
 - Introduce your work and ‘set the scene’.
 - Discuss all your design decisions (e.g., choice of corpus, chromosome design, crossover and mutation operator design, elitism, etc.).
 - You should perform a good evaluation. I expect proper experimental procedure discussing your setup, expected outcomes, results, and a good discussion.
 - A proper conclusion covering important and interesting observations.

Statement of completion – MUST be completed and included in your report

Item	Completed (Yes/No/Partial)
Implement 'base' genetic algorithm	
Two-point crossover	
Implemented a mutation operation	
Elitism	
Good evaluation and discussion	
<i>If partial, explain what has been done</i>	

Marking Breakdown

Description	Marks allocated
'Base' genetic algorithm	45%
Two-point crossover	5%
Mutation	5%
Elitism	5%
Evaluation	25%
Overall report quality	15%