COMP7015 Artificial Intelligence **Semester 1, 2022/23**

Course Project Instructions

Instructor: Dr. Kejing Yin Oct. 14, 2022

1 Requirements

1. You should form a group of **one to four** people to carry out the project.
2. Milestones (important dates):
   (a) **Oct. 21, 2022 (Friday)**: Register your group information (names and student IDs of your group member) and topic selection. You will be notified your group number afterwards. Link: `https://hkbu.questionpro.com/comp7015reg`
   (b) **Nov. 4, 2022 (Friday)**: Submission of a **Progress Report** in Moodle.
   (c) **Dec. 3, 2022 (Friday)**: Submission of the final work, including all **source codes**, a **Final Project Report** and a **Presentation Video** in Moodle.
3. In the **Progress Report**, you should briefly describe your *motivation of the topic selection*, *work progress (with contribution of each group member clearly stated)*, *challenges encountered*, and *plans to tackle the challenges* in no longer than two A4 pages.
4. The **Final Project Report** should be at least five A4 pages, covering *a brief introduction of the topic you selected*, *an overview description of your methods used*, *detailed descriptions of the experimental design*, *the results you obtained*, *a conclusion*, and *the contributions of each group member*. There is no upper page limit, but it is suggested not to exceed 10 A4 pages.
5. In the **Final Project Report**, you are required to faithfully report the contribution of each group member. Your final score will depend on how much contribution you make to the project.
6. The **Presentation Video** should be 10-15 minutes, covering your solutions to and the main results of the course project. Each group member should clearly present his/her contribution to the project and the entire presentation should be organized in a logic way.
7. You are highly encouraged to use as much visualizations as possible (e.g., figures, tables, and graphs) to make your presentation clearer.
8. In the **Presentation Video**, you are required to display the presenter's faces clearly to allow the lecturers to identify your identity.
9. In the final submission, you are required to submit three files to Moodle: 1) the Final Project Report in .pdf format, 2) the presentation video in any video format, and 3) all other files/documents and the source codes compressed in a .zip file.
10. We have zero tolerance to cheating and plagiarism. Your project report and source codes will be scanned by anti-plagiarism platforms, anyone who plagiarizes and whom were plagiarized will receive zero score and be reported to the departmental exam committee for further penalty.
11. **Resubmission of previous work**: Please be noted that you cannot submit your previous work (e.g., published papers and work submitted to other courses) or otherwise it will be considered self-plagiarism.

12. **Third-party library policy**: In option 1, you need to implement the algorithms on your own. Using third-party libraries are not allowed. In options 2 and 3, you can use third-party libraries but you need to explicitly acknowledge them in your report and presentation.

13. If you take a cue from other sources, *e.g.*, books, literature, and online resources, you should clearly cite them in your Final Project report. Note that copying source codes from online resources is explicitly prohibited.

2 Topics

You can choose one topic from the following:

Topic 1: Map Navigation

Suppose you are an AI algorithm developer in a tech enterprise that develops a mobile map and navigation app targeting the Californian market. You are responsible for developing algorithms to help your users find the best route given an initial location and a destination.

The real map of California are provided in two csv files: the "CaliforniaRoadNetwork_Nodes.csv" file and "CaliforniaRoadNetwork_Edges.csv" file. The intersect of roads are called "nodes" and their coordinates are recorded in the former file. Therefore, the roads are connections between nodes and are called "edges", which are recorded in the later file. To make it simpler, we assume that all roads are straight and two-way. We can compute the length of a road by

$$l = \sqrt{(x_{\text{start}} - x_{\text{end}})^2 + (y_{\text{start}} - y_{\text{end}})^2},$$

where (x, y) denotes the coordinates of the start and end nodes.

In this project, you are required to:

1. Find the shortest route of the following start nodes and destinations. State what algorithm you use and the reasons why you use that algorithm.

Start Node	End Node
0	1894
4	3115
18	9186
25	15061
33	21040

2. Randomly sample $1,000$ combinations of start and end nodes, apply BFS, graph DFS, uniform-cost search, greedy search, and A* search to the random samples, measure and compare the time and memory the different algorithms require. If you find profiling memory consumption difficult, you can count the maximum number of elements in the frontiers and visited lists as an approximation of memory usage.

3. Summarize and compare the solutions found by different algorithms and the time and memory they require. Based on your observations, discuss the advantages and disadvantages of different algorithms.

4. Implement tree-like search, depth-limited search, and iterative deepening depth-first search. Test the algorithms you implemented using the randomly sampled start and end nodes.

Topic 2: Coupon Acceptance Prediction

Suppose you are a data scientist working at a company that develops a mobile payment app. You are responsible for making recommendations of different coupons to vehicle drivers. To make more effective recommendations, you would like to build a machine learning model to predict whether the driver would accept the coupon or not.

For this problem, we can use a dataset collected via a survey on Amazon Mechanical Turk. The survey describes different driving scenarios including the destination, current time, weather, passenger, etc. A detailed description can be found in `https://archive.ics.uci.edu/ml/datasets/in-vehicle+coupon+recommendation`.

In this project, you are required to:

1. Select 2–3 suitable machine learning models for this dataset and read in details the documentation in the scikit-learn package about the models you selected. Introduce the models in your reports and presentation.
2. Select proper method (hold-out, K-fold cross validation, etc.) to evaluate the performance.
3. Measure the performance using the metrics introduced in the course. Compare the different machine learning models using the metrics. Discuss the differences between different performance metrics.
4. Try to tune the hyperparameters in the models you selected and observe how they affect the performance.

Topic 3: Open Topic

If you are not interested in the two problems introduced above, you can propose another problem/dataset to work on for your course project. If you decide to do so, there are a few more requirements:

1. Submit a project proposal (at most one A4 page) by Oct. 27, 2022.
2. Since there is only about 6–7 weeks for the project, so the scope should not be very large. Having said so, the workload should be at least similar to the two problems above.
3. In the project proposal, you should briefly state the problem and precisely describe what you plan to do, including what dataset or algorithms you plan to use, how would you evaluate the results, etc.
4. You do not need to limit yourself to the algorithms/models introduced in the course. Any AI-related methods can be used as long as they are suitable for the problems you proposed.
5. If needed, the course instructor and TAs will give brief feedback to your proposal to make sure that it is feasible.