RMIT UNIVERSITY

Machine Learning & Computational Machine Learning
COSC 2673 & COSC 2793
Assignment 3

Assessment Type	Individual assignment. Submit online via Canvas → Assignments → Assignment 3. Marks awarded for meeting requirements as closely as possible. Clarifications/updates may be made via announcements/relevant discussion forums.
Due Date	Week 13, Thursday 1st of June 2023, 11:59pm
Marks	20%

1 Overview

This assignment is designed to help you become more confident in designing machine learning systems. In this assignment you will conduct a virtual presentation (pre-recorded), presenting a brief summary and critical analysis of the project work that is done in Assessment Task 2, as well as improvements/extensions that could be made for his/her own work based on a literature review of the state-of-the-art approaches.

The assignment will consist of:

- Critical analysis of an ML technique applied to solve a real-world ML problem.
- Conduct a review to identify the state-of-the-art approaches relevant for a particular problem.
- Compare and contrast your approach with state of the art approaches to solve similar problems.
- Research how to extend the modelling techniques that are taught in class.
- Demonstrate theoretical understanding of ML techniques.

To complete this assignment, you will require skills and knowledge from lecture and lab material for Weeks 1 to 12 (inclusive).

This assignment has ONE deliverable:

1. **A recorded video presentation**.

2 Learning Outcomes

This assessment relates to the following course learning outcomes (CLOs):

- **CLO 1:** Understand the fundamental concepts and algorithms of machine learning and applications
- **CLO 2:** Understand a range of machine learning methods and the kinds of problem to which they are suited
- **CLO 5:** Understand major application areas of machine learning
- **CLO 6:** Understand the ethical considerations involved in the application of machine learning.

3 Assessment details

3.1 Task

Using machine learning in real-world settings involves more than just running a data set through a particular algorithm. You need to have skill in researching and understanding what others have done to solve similar problems. In this assignment, you will do a critical analysis of a ML project, as well as improvements/extensions that could be made based on a literature review of the state-of-the-art approaches. You will also demonstrate theoretical understanding of ML techniques.

The projects you are to investigate for this assignment is given in Section 4. There are two projects and you **should** select the same project you worked on for assignment 2. Irrespective of the project you should record a presentation **less than 11 minutes long** that includes the following:

- *(approximately 2 minutes)* Present a **critical analysis** of the model(s) that you have developed.
- *(approximately 5 minutes)* **A review of the techniques in literature** that are used to solve same/similar problems and compare and contrast your approach to the approaches found in your review.
- *(approximately 3 minutes)* Discuss **how the project can be extended to cover some challenging scenarios**. The challenging scenarios that can be considered for each project can be found in section 4.

You may use slides in your presentation. However, content on the slides that are not adequately explained (verbally) would not receive any marks.

The instructions for recording a presentation is provided separately on the assignment 3 canvas page. Please follow the instructions carefully and make sure the access privileges are configured correctly so that the markers can access the recording. **Incorrectly submitted/corrupted recordings or recordings that cannot be accessed will receive zero marks.**

If the recorded video is over 11 minutes long. Only the part up to 11 minutes will be accessed. The over length part of the recording will be ignored.

3.2 Critical Analysis

The first part of your presentation should include a critical analysis of the solution that you developed in assignment 2. The analysis may include:

- A brief summary of your solution to the problem.
- Other approaches considered while developing the solution.
- Analyse the model and its outputs.
- Limitations of the model you developed.
- Why do you think the approach is adequate to solve the problem.

3.3 Review

Next, you should conduct a review to identify the following items and discuss your findings in the presentation. You review should identify:

- Other works in literature that solves the same/similar problem as the one you have solved.
- Other techniques and algorithms that can be used to solve the problem given to you.

See Section 4 for specific details relevant to your project.

A good literature review:

- Should follow a logical structure.
- Should not just provide a list of related papers. Papers should be discussed in relation to why are they relevant for this problem, what is good about them and what are the limitations.
- Should discuss literature from peer-reviewed sources. Wikipedia or web discussion forums are not considered as peer-reviewed sources.

Remember that good literature review provides factual statements that summaries the work in literature in a way that is useful for the listener to understand the context and follow your rational. Statements such as:

"Ruwan did <xyz>. Andrew did <xyz>. Wei did <xyz>"

is not a literature review. This is an annotated bibliography. Instead, you should aim for statements such as:

"To solve the problem <xyz> Ruwan did <xyz>, while this method can handle <xyz> it has the limitations <xyz>. To overcome this limitation, Andrew proposed <xyz> ..."

3.4 Extensions

Finally you should discuss how your system can be extended to handle the specific situation given under section 4. This should include a justified problem formulation and well though out approach. Your discussion should also include how you would know if the proposed solution is adequate to the task. No experiments (implementations) are needed at this stage.

4 Projects

The projects you are to investigate for this assignment is given in this section. There are two projects and you **should** select the same project you worked on for assignment 2. Each project has different requirements, so ensure you are aware of these differences.

Project 1: Classify Images of Colon Cancer

Summary: Assume you are a machine learning engineer working for a biomedical startup company. You have just developed a machine learning system that can classify histopathelogy images of colon cells. You have used a modified version of the "*CRCHistoPhenotypes*" data-set to develop two ML models to perform the following two tasks:

- Classify images according to whether given cell image represents a cancerous cells or not (`isCancerous`).
- Classify images according to `cell-type`, such as: `fibroblast`, `inflammatory`, `epithelial` or `others`.

The next step is to do a presentation to your management and other ML engineers in your company, critically analysing your model and discussing the related works and extensions. Your talk should include the following:

Critical analysis: Under this section you should analyse the model that you developed and discuss the design choices you made during the model development stage. See Section 3.2 for things that you might discuss under this category.

Review: You talk should include a discussion on:

- Other works in literature that solves the same/similar problem as the one you have solved.
- State-of-the-art techniques used for solving image classification problem.

As a starting point, the following two papers are provided to you. You should discuss these in your presentation and compare the approach you have taken to what is discussed in the literature. For higher grades (HD/DI) you should include more paper in your review (see rubric for how many papers to include). The important aspects to be included in the review are discussed in section 3.3.

- K. Sirinukunwattana, S. E. A. Raza, Y. Tsang, D. R. J. Snead, I. A. Cree and N. M. Rajpoot, *"Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images,"* in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1196-1206, May 2016. Paper Link.

- He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). Paper Link

Extension: A typical Histopathelogy images has a high resolution. For example the original images in the "*CRCHistoPhenotypes*" data set has a resolution of 500x500. However an individual cell in that image is relatively small (fits within a rectangle of 27x27). If a model is developed to process the high resolution images (or a down sampled version of it), then the fine scale information of the cells will be lost and the classifier would not be good. To solve this problem, individual cells were detected in the original histopathelogy images (see figure 1) and, was extracted in to individual images of 27x27 (in the data-set given to you in assignment 2) in the data set provided to you in assignment 2. In order to make a complete system (end-to-end), you also need to automate the detection and extraction of individual cells from the original histology image with 500x500 resolution. What are the machine learning techniques that are applicable for this task and discuss how you would formulate the problem. No experiments are needed at this stage.

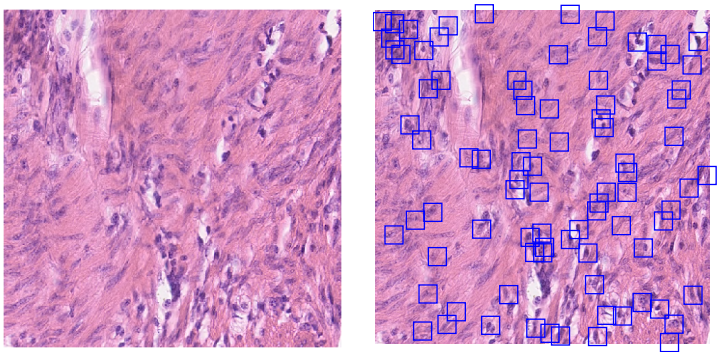

Figure 1: Example histopathelogy image and the detected cells The extension is to do this cell detection automatically.

Project 2: Learning to do Packet Scheduling in Routers

Summary: Assume you are a machine learning engineers working for a technology company that is manufacturing routers. Your just developed a *reinforcement learning* based scheduling algorithm for a router that is being designed currently by your company. Your scheduling algorithms can handle traffic with different QoS constraints and works under two different scenarios:

Scenario 1: For each timeslot the scheduler selects a queue and transmits the first packet in the queue (ignore time taken to switch queues).

Scenario 2: For each timeslot the scheduler selects to send a packet from the current queue or switch the queue (i.e. one timeslot is taken to switch queues).

The next step is to do a presentation to your management and other ML engineers in your company, critically analysing your model and discussing the related works and extensions. Your talk should include the following:

Critical analysis: Under this section you should analyse the model that you developed and discuss the design choices you made during the model development stage (e.g. elements of the reinforcement learning problem). See Section 3.2 for things that you might discuss under this category.

Review: You should conduct a review to identify:

- Other works in literature that solves the same/similar problem as the one you have solved.
- Other techniques and algorithms that can be used to solve the problem given to you.

As a starting point, the following two papers are provided to you. You should discuss these in your presentation and compare the approach you have taken to what is discussed in the literature. For higher grades (HD/DI) you should include more paper in your review (see rubric for how many papers to include). The important aspects to be included in the review are discussed in section 3.3.

- Ferrá, H.L., Lau, K., Leckie, C. and Tang, A., 2003, August. Applying Reinforcement Learning to Packet Scheduling in Routers. In IAAI (pp. 79-84). Paper Link.
- Wang, Q., Nguyen, T. and Bose, B., 2020, February. Towards Adaptive Packet Scheduler with Deep-Q Reinforcement Learning. In 2020 International Conference on Computing, Networking and Communications (ICNC) (pp. 118-123). Paper Link

Extension: Now your company has has started another business - manufacturing traffic lights for intersections. Most traffic lights operate on a predetermined phasing (that is, a timing when the lights change), which does not take into account the number of vehicles that are waiting at, or approaching, the intersection. This leads to a sub-optimal throughput of vehicles. For example, some vehicles may be waiting at a red light while the intersection is otherwise clear. A more intelligent policy, can increase the throughput.

You are asked to investigate if the scheduling system you developed can be adapted to improve the throughput of traffic at an intersection with traffic lights. You need to formulate the reinforcement learning problem and discuss the approach you might take. Also discuss the similarity and differences compared to the packet scheduling solution you have developed.

The intersection for the above analysis can be assumed to have only two roads crossing (north-to-south and east-to-west) and at least one pedestrian crossing (across east-to-west road). An example is given in Figure 2.

5 Submission

You have to submit all the relevant material as listed below via Canvas.

1. Share-point link to your recorded presentation.

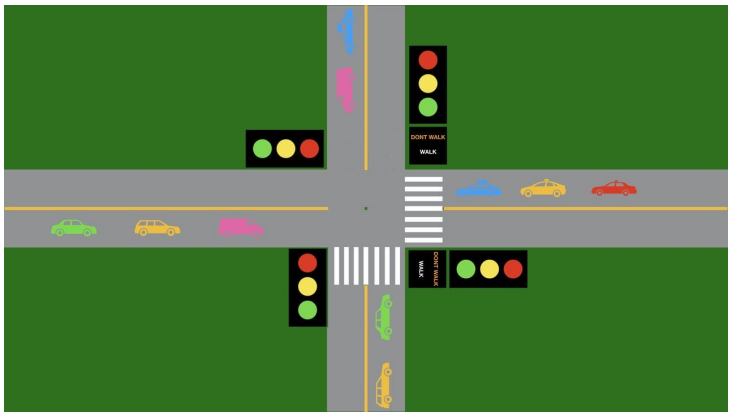

Figure 2: Example Traffic intersection.

The instruction for recording the presentation will be provided on canvas. *Modification to the content pointed by the submitted link, after the due date will incur late penalty.* **Incorrectly submitted/corrupted recording links or links that cannot be accessed will receive zero marks.**

After the due date, you will have 5 days to submit your assignment as a late submission. Late submissions will incur a penalty of 10% per day. After these five days, Canvas will be closed and you will lose ALL the assignment marks.

Assessment declaration:
When you submit work electronically, you agree to the assessment declaration - `https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/assessment/assessment-declaration`

6 Teams

Not relevant. This is an individual assignment.

7 Academic integrity and plagiarism (standard warning)

Academic integrity is about honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas. You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e. directly copied), summarised, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods
- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites. If you do not acknowledge the sources of your material, you may be accused of plagiarism because you have passed off the work and ideas of another person without appropriate referencing, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct. Plagiarism covers a variety of inappropriate behaviours, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to the following: `https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity`.

8 Marking guidelines

A detailed rubric is attached on canvas. In summary:

- Critical analysis 30%;
- Review 40%;
- Extension 30%;