

Assignment 3: Take Home Exam

31005 Machine Learning Spring 2019

TASK

Answer ONE of the following questions (at the end of this document). Answers should be around 800-1,000 words in length. The structure of your report will vary slightly for different questions. The required information is stated in the respective questions.

The questions are challenging, open research questions. It is natural if you are not 100% sure about your answer or you could not answer the question completely. Your answer will be assessed by your reason in the argument and your insight.

You are allowed (and practically required) to do additional reading on the subject to answer the questions. The references should be cited properly. **The task is individual work. Direct discussion of the exam question in any form is considered dishonest.**

Criteria

100	<p>The answer is relevant to the question.</p> <p>The answer is technically sound, which can be assessed either by theoretical proof, or by reasonable hypothetical arguments, or by empirical evidence such as experimental support.</p> <p>The proposed solution can be practically feasible. This point is either convinced by logical argument in the report or supported by empirical evidence.</p> <p>The report contains sufficient background research of existing solutions to related or similar problems.</p> <p>More possible methods are considered. The proposed method is well motivated among the alternatives.</p> <p>the report has a clear structure and is well written.</p>
50	<p>The answer is mostly relevant to the question.</p> <p>The answer is mostly technically sound, but contains obvious issues, or lacks support.</p> <p>Feasibility discussion is missing, but it is possible for the readers to accept that the proposed solution is feasible.</p> <p>There is no effort in studying the background research of existing solutions to related or similar problems. No alternatives are considered.</p> <p>The report has a clear structure and is written with care and easy to follow.</p>

0	<p>The answer is mostly irrelevant to the question.</p> <p>OR</p> <p>The answer disagrees with existing understanding of the problem, and no logical reason has been provided.</p> <p>OR</p> <p>Feasibility discussion is missing, and it is hard to be convinced that the proposed solution is feasible. There is no background research of existing solutions to related or similar problems. No alternatives are considered.</p> <p>OR</p> <p>The report is unreadable.</p>
---	--

Marks

This assignment contributes 30% to your final mark.

SUBMISSION

Due date 11:59pm 9 Oct 2019.

You need to submit an exported PDF of your Jupyter Notebook, including a PLAIN TEXT link to the Github file to UTSONline. (Same as A1/2)

Late Penalty 10 marks per day (round-up) past deadline.

Extension Extensions may be granted if arranged with the Subject Coordinator before the deadline and if decent progress/effort has been made at the time of application. We use github commission history of your draft report as evidence for progress/effort.

If your performance in an assessment item or items has been affected by extenuating or special circumstances beyond your control you may apply for Special Consideration. Information on how to apply can be found at <http://www.uts.edu.au/current-students/managing-your-course/classes-and-assessment/special-circumstances/special>.

Due to the size of our class, extensions will lead to delay in marking your assignment **and the final grade**.

GROUPWORK

This assignment is an individual task.

QUESTIONS

QUESTION 1

Following your graduation, you are hired by a polling organisation as a data analyst. As social media has exploded and transformed the way people interact with each other, it would be a great idea to use messages collected from social media to predict how the user can be converted to change his/her support. List three challenges to solving this problem. With reference to existing approaches, describe the design of your system. Discuss the ethical and social consequence of this study.

QUESTION 2

Ensemble methods have been very successful in building classifiers. The hot topics include how to create diverse classifiers and how to fuse the decisions from individual classifiers, in particular how to establish the weights that individual classifiers contribute to the ensemble's answer. Describe two existing approaches to solving this problem, discuss their advantages and disadvantages. Make a plan to address one issue or two (related to learning the weights or creating diverse classifiers), briefly describe your new method. Explain the reason why the developed method could outperform the conventional ones.

QUESTION 3

Marketing or advertising companies would be very interested in being able to predict whether a Twitter message will spread as a meme or not, and even better, construct it so that it will spread. Why is this a hard problem to solve? Describe two approaches using data analytics to predict whether a tweet will go viral or not. How would you validate these approaches? Discuss the ethical and social consequence of this study.

QUESTION 4

One of the themes in the machine learning models we've looked at this semester is large numbers of parameters that are changed by tiny amounts. Why do so many apparently different models use such similar techniques? Are there other ways to approach the problem of learning? Are there also commonalities in the way the amounts to be changed are determined?

QUESTION 5

Consider if you are in front of a gambling machine. The machine has n arms, pulling each will yield a random amount of reward. The average reward yielded by each arm in long-run is a fixed certain value, but the money you receive in individual rounds is random. E.g. you can **expect** pulling arm-2 will produce a return of r_2 , but the actual returns are random values. The expected return of each arm is unknown – you know there is a fixed value, but not knowing what the value is. The task is to i) design a strategy to earn reward as fast as possible (“fast” is defined in terms of the number you pulling the arms); ii) identify the main challenge in designing such strategies; iii) discuss the up bound of the performance of the optimal policy.