COMP9417: Project Guidelines

March 26, 2021

Overview of Guidelines

- The deadline to submit the report is 5:00pm April 28. The deadline to submit your predictions is 2:59pm April 26 (Sydney time).
- Submission will be via the Moodle page
- You must complete this work in a group of 3-5, and this group must be declared in Moodle under Group Project Member Selection
- The project will contribute 30% of your final grade for the course.
- Recall the guidance regarding plagiarism in the course introduction: this applies to all aspects of this project as well, and if evidence of plagiarism is detected it may result in penalties ranging from loss of marks to suspension.

Dataset & Evaluation

You must choose one of the following projects and register your choice at the latest by Tuesday, March 30 using this form. The project choices are:

1. TracHack Challenge under Chellenge 21.2
2. Machine Learning in the Unknown (internal challenge - see Moodle for spec).

Both data sets will be released on April 5th.

Note: Please note that, if you choose to participate in the TracHack project, you need to sign an NDA before being able to access the data and receive the working environment setup instruction, so it is advised that you register your choice as soon as possible to avoid any potential delay in accessing the data.

Objectives

In this project, your group will use what they have learned in COMP9417 to construct a classifier for the specific task as well as write a detailed report outlining your exploration of the data and approach to modelling. The report is expected to be 10-12 pages (with a single column, 1.5 line spacing), and easy to read. The body of the report should contain the main parts of the presentation, and any supplementary material should be deferred to the appendix. For example, only include a plot if it is important to get your message across. The report is to be read by the client, and the client cares about the big picture, pretty plots and intuition. The guidelines for the report are as follows:

1. Title Page: tile of the project, name of the group and all group members (names and zIDs).
2. Introduction: a brief summary of the task, the main issues for the task and a short description of how you approached these issues.
3. Exploratory Data Analysis: this is a crucial aspect of this project and should be done carefully given the lack
    - f domain information. Some (potential) questions for consideration: are all features relevant? How can we represent the data graphically in a way that is informative? What is the distribution of the classes? What are the relationships between the features?
4. Methodology: A detailed explanation and justification of methods developed, method selection, feature selection, hyper-parameter tuning, evaluation metrics, design choices, etc. State which method has been selected for the final test and its hyper-parameters.
5. Results: Include the results achieved by the different models implemented in your work, with a focus on the f1 score. Be sure to explain how each of the models was trained, and how you chose your final model.
6. Discussion: Compare different models, their features and their performance. What insights have you gained?
7. Conclusion: Give a brief summary of the project and your findings, and what could be improved on if you had more time.
8. Reference: list of all literature that you have used in your project if any. You are encouraged to go beyond the scope of the course content for this project.

Project implementation

Each group must implement a minimum of two classification methods and select the best classifier, which will be used to generate predictions for the test sets of the respective task. You are free to select the features and tune the methods for best performance as you see fit, but your approach must be outlined in detail in the report. You may also make use of any machine learning algorithm, even if it has not been covered in the course, as long as you provide an explanation of the algorithm in the report, and justify why it is appropriate for the task. You can use any open-source libraries for the project, as long as they are cited in your work. You can use all the provided features or a subset of features; however you are expected to give a justification for your choice. You may run some exploratory analysis or some feature selection techniques to select your features. There is no restriction on how you choose your features as long as you are able to justify it. In your justification of selecting methods, parameters and features you may refer to published results of similar experiments.

Code submission

Code files should be submitted as a separate .zip file along with the report, which must be .pdf format.

Peer review

Individual contribution to the project will be assessed through a peer-review process which will be announced later, after the reports are submitted. This will be used to scale marks based on contribution. Anyone who does not complete the peer review by the Thursday of Week 12 (6 May) will be deemed to have not contributed to the assignment. Peer review is a confidential process and group members are not allowed to disclose their review to their peers.

Project help

Consult Python package online documentation for using methods, metrics and scores. There are many other resources on the Internet and in literature related to classification. When using these resources, please keep in mind the guidance regarding plagiarism in the course introduction. General questions regarding group project should be posted in the Group project forum in the course Moodle page.