Xi'an Jiaotong-Liverpool University
西交利物浦大學

INT104 – Artificial Intelligence

Coursework – Lab Report

Introduction

In this coursework, a spreadsheet has been provided to perform a set of data analysis. The spreadsheet contains the following information: the index of student, gender of student, the programme that a student is enrolled, the grade that the student is in, total marks that a student is awarded and the mark of 5 exam questions (indexed as Q1, Q2, Q3, Q4 and Q5).

The first column is the ID of the student. The gender of the student is represented as "1" and "2". The grade of the student is either "2" or "3". The programme of the student is represented as "1", "2", "3" and "4". The full mark for 5 exam questions are 8 marks (Q1), 8 marks (Q2), 14 marks (Q3), 10 marks (Q4) and 6 marks (Q5) respectively.

The coursework requires students to find a better solution of features that may predict the student's group according to mark distribution. The student has 6 hours (Week 5 and Week 6) in lab sessions to perform basic experiments. A report for the result obtained should be submitted by the last Friday of week 13.

Please note that the experiment performed, and the result obtained in all lab sessions can be documented in the report as long as the student thinks such experiments and results support the content of lab report.

Tasks

1. Observe the distribution of raw data and analyse what makes the classification of programmes difficult.
2. Try various transform of data. Visualisation may be helpful. The expected output is to make the samples of the same programme having a similar distribution.
3. Compare the candidate features and comment on how the resulting clusters are associated with the programme information, given clustering the data into four clusters.
4. Try to build several classifiers that predict the student's programme according to mark

distribution. Evaluate the performance of classifiers.

5. Write a report that analyses that which set of features is better than others for the task of predicting student's programme.

Requirement

You MUST use Python to perform the dedicated tasks. Over the lab session, a Teaching Assistant (TA) will be assigned to you to support your work. However, please bear in mind that it is NOT the TA's responsibility to teach you Python programming nor design the experiment for you.

After the lab session, you should write a lab report that documents the experiment you performed, the results you obtained and the discussion that justifies your recommended way of feature extraction.

A Python script that guides the experiment with a different dataset has been provided separately. The student is kindly reminded that 1) implementing the provided Python script with the given dataset ONLY will result a mark of zero; 2) answering all questions raised in the Python script does not guarantee a high mark.

The length of the lab report after the lab session should be no more than **6 pages** in double columns (refer to IEEE format) excluding reference lists. The report could be simply titled as "lab report" but the student could also title the report in their own favourite way. The report does not need a cover page, the student should write their names under the title with student ID provided. The student should also name their assigned TA under their own names. Though literatures could be cited to support the ideas in the report, it is NOT necessary to review relevant literatures in the report hence it is absolutely no problems to cite no papers in the report. The lab report should be handed in with a single PDF file without source code attached. The use of Latex is strongly recommended.

The use of AIGC is allowed for proofreading and brainstorming ONLY. Copying an AI generated solution to the task will not guarantee that you could pass the coursework. Generating a report via AIGC is strictly forbidden. It is the critical thinking, experiment design and the analysis of result that matter. You MUST fully understand your code and the experiment you have designed in this coursework.

The coursework will be marked purely on report. Please refer the following pages of marking criteria for the submitted report.

Marking Criteria

Lab report:

Editorial and Language Issues (15 marks)

1. Editorial issues will be marked in a proportional way. Each 1% of content with formatting issues will deduct 1 mark.
2. Each 1% of content that cannot be understood by readers will deduct 1 mark.
3. Each 1% of the content without academic style will lose 1 mark.
4. A total of 15 marks maximum could be deducted for editorial and language issues.

Task 1 & 2 (25 marks)

- The distribution of raw features is observed. (2 marks)
- There are at least 3 transforms being performed for the raw features. (4 marks each, in total 12 marks)
- The features are compared via at least one matric. (8 marks)
- A conclusion is drawn from the comparison. (3 marks)

Task 3 (25 marks)

- At least three sets of features are clustered. (3 marks each, in total 9 marks)
- For each set of features, there are at least three clustering configurations are tried. (3 marks each, in total 9 marks)
- There are justifications provided on the experiment design, i.e., the choice of clustering methods and configurations. (5 marks)
- The best performed clustering results are compared with the distribution of programmes. (2 marks)

Task 4 (25 marks)

- At least three sets of features are used for classification of programmes. (2 marks each, in total 6 marks)
- For each set of features, there are at least three sets of hyper parameters are tried. (2 marks each, 6 marks in total)
- An ensembled classifier is built with attempt of 2 sets of different hyper-parameters. (3 mark each, in total 6 marks)
- There are justifications provided on the experiment design, i.e., the choice of

classification methods and configurations. (5 marks)
- Comment on why the best choice of features and classifiers. (2 marks)

Discussions and Conclusions (10 marks)

- Present a new finding being discovered by the experiment results presented. (4 marks)
- Justify how the discovered finding is supported by the experiment results. (3 marks)
- Based on the existing results, what further work could be performed to strength the presented new finding. (3 marks)

Penalties:

- -10% of awarded marks: for improper citation
- -20% of awarded marks: for severe improper citation (multiple improper citation or duplication of a whole paragraph)
- University academic integrity penalties apply.
    - Any report has a similarity higher than **15%** will be reviewed for plagiarism identification. According past experience, a report with a similarity higher than **25%** is highly likely to be identified with an academic integrity issue.
    - Duplicated experiment results and diagrams are highly likely to be investigated for potential academic integrity issues.

Submission

1. Only submissions in PDF format are accepted.
2. Submit your lab report via the dedicated Learning Mall coursework link before the Friday of week 13.
3. Please name your submission file as ID_FirstName_LastName_C1.pdf (e.g., 1234567_FirstName_Surname_C1.pdf).
4. Late submission policy of XJTLU applies.