

INT104–Artificial Intelligence

Coursework 2

Introduction

Providing comfortable medical services is an important direction for the development of anesthesia technology today. However, the drugs and anesthesia methods need to be considered according to the individual conditions of patients and different treatment purposes. To this end, we designed a questionnaire covering 15 questions that can reflect the physical status of patients. According to the scores of each question, we can judge the appropriate anesthesia method for them.

A spreadsheet of mark is provided where the mark for each question is listed for each patient. There are more than 5000 patients in total who comes from 2 results. The results in the spreadsheet are labelled as ‘0’ and ‘1’. There are also a few patients that belong to other results, which are labelled as ‘2’. Such data could be treated as outliers and could be simply deleted. This coursework requires the student to design a classifier to classify patients to the results they belonged to, according to the mark earned for each question.

There are three lab sessions planned for this coursework, where you must demonstrate your progress to your assigned TA in each lab session. Each lab session corresponds to a task in this coursework, which will be described in the following sections. After all tasks are finished, you are expected to submit a lab report documenting the whole process of your attempt on design a classifier.

Though Python is recommended to be used as the implementation tool for this coursework, other tools / programming language such as Microsoft Excel, Weka, MATLAB, Java, and C++ are also acceptable. The student could use MULTIPLE types of tools throughout three tasks in this coursework. **The lab report should be submitted in PDF format**, which could be produced by any typesetting software such as Microsoft Word, Latex (recommended) and Markdown.

This coursework will be marked by the lab report submitted with a given marking criteria. Over each lab session, your TA will check your progress with a demonstration required. **Missing any lab session will result to a proportion of marks being deducted.** Please refer to the marking criteria attached in this instruction for details. The demonstration process during each lab may include showing your code and report draft, which demonstrates that the time management ability of the students and avoids possible cheating on programming.

Task 1: Dimensionality reduction

Many Machine Learning problems involve thousands or even millions of features for each training instance. Not only do all these features make training extremely slow, but they can also make it much harder to find a good solution. This problem is often referred to as the curse of dimensionality. Fortunately, in real-world problems, it is often possible to reduce the number of features, turning an intractable problem into a tractable one.

Principal Component Analysis (PCA) is by far the most popular dimensionality reduction algorithm. First it identifies the hyperplane that lies closest to the data, and then it projects the data onto it. It seems reasonable to select the axis that preserves the maximum amount of variance, as it will most likely lose less information than the other projections. Students are required to find the ways to make the reasonable dimensionality reduction, where PCA is recommended in this task. There are no dedicated number of feature dimension requirement. With the presented data, the cumulative sum of explained variance ratio of each principal component could be presented as data visualization.

In this task, the student should try to extract data features and make the dimensionality reduction of the given dataset and decide which data feature will be used as the input of classifier that to be developed in the next task. In the lab report, a full and detailed justification on data preprocessing and dimensionality reduction should be presented.

Task 2: Build Classifiers

With the selected data features, this task requires the student to design at least **three** classifiers to classify which results a patient is enrolled, according to the marks awarded for each question in an questionnaire. For a better performance of the proposed systems, data features may be used as the input of the candidate system, which replace the raw data given in the dataset.

The candidate classifier should be built in a supervised way. The classifier could be the method taught in the lectures such as Support Vector Machine (SVM) and Decision Tree. However, the student is also encouraged to try methods beyond the scope of the lectures delivered such as deep neural networks and Bayesian graphical models.

In the lab report, the student should specify what method is used to build the candidate classifiers, what data feature is used as the input of the candidate system and what result is obtained. The key part of program such as data pre-processing, the training process of the model and the inference process of the system should be stated with details. However, the use of screenshot should be avoided, where a text-based description (such as text-based source code with line number) is preferred when necessary. (NOTE: source code presented in the lab report should be in Courier New or similar fonts.)

The student is then required to recommend ONE classifier among the three candidate classifiers. The process of classifier evaluation should be fully demonstrated. The decision of recommendation should also be fully justified. The process of classifier evaluation and the justification of classifier choice should be documented in the lab report.

Task 3: Unsupervised Patients Classification

In this task, the student is required to classify patients to different groups **in an unsupervised manner**, according to marks awarded from the questionnaire. The classification of groups should enable more targeted treatment by classifying patients in the same group. There are no dedicated ways of classification hence the student should make their choice of principle applied to classification (i.e., **there are no dedicated number of groups that the patients should be classified into**).

In the lab report, the student is expected to present the full details of the unsupervised classification process and fully justify the final decision made for patient classification. Specifically, the student should interpret the principle that is followed for the classification. As the case of earlier tasks, screenshots of source code should be avoided.

Lab Report

Please note, a title should also be included in the lab report. (Please do NOT use “lab report” as the title of the report.). **The page limit for the lab report is 8 pages** excluding abstract, reference list, cover page and appendix. The majority of marks in this coursework will be awarded **according to the content of lab report**. Please be aware that **a lengthy report does not guarantee a high mark for this coursework**. It is not necessary to attach the source code in the lab report.

A separate lab report template will be provided as a reference. However, **the outline provided need not to be strictly followed**.

The use of ChatGPT is prohibited except for proofreading purposes.

Marking criteria

Marking Criteria	Item	Marks	Description
Editorial and Language Issues (15)	Formatting	5	Format should follow IEEE-like style with double columns: <ul style="list-style-type: none">● Font size: 10 pts● Line spacing: 1● Alignment: Justified● Font: Times New Roman● Margins: Conventional
	Organization	5	The paper should be well organized showing a clear structure of the report as instructed in this material.
	Language Issue	5	The quality of language in the report.

Task 1 (20)	Dimensionality Reduction	10	Extracting reasonable dimension of feature from the original data.
	Data Visualization	5	Visualize the explained variance ratio of each extracted features.
	Live Demonstration	5	Show your progress to TA over lab session
Task 2 (25)	Description of Classification Methods	5	Describe the selected methods of classifications with justification.
	Training Classifiers	5	Training at least three classifiers (no more than five) and obtain predictions.
	Classifier Selection	5	Using the obtained classifiers to classify data and evaluate performance of the candidates.
	Live Demonstration	10	Show your progress to TA over lab session
Task 3 (20)	Unsupervised Classification	10	Classify patients into several groups via one unsupervised classification method.
	Critical Thinking in Interpretation	5	The students should interpret the classification results with a reasonable interpretation of language.
	Live Demonstration	5	Show your progress to TA over lab session
Lab Sessions (20)	Time Management	20	The assigned TA will judge whether the coursework has been proper progressed with lab session. If you miss any lab session, you may lose a few marks as a result of incomprehensive progress evaluation.
Repetition check			This coursework will be checked by Turnitin for the repetition rate of your submission. If a repetition rate higher than 20% is founded (more than 2% from a single source), the report will be checked for the identification of plagiarism. Possible penalties include a deduction of marks or a mark of zero.

Submission

1. Only submissions in PDF format are accepted.
2. Submit your lab report via the dedicated Learning Mall coursework link.
3. Please name your submission file as IDnumber_FirstName_LastName_C2.pdf
4. Late submission is NOT allowed. The deadline of submission is set to the Thursday of Week 14.