Artificial Intelligence and Computer Graphics
(AIG710S)

First Group Assignment

Due on 25/05/2020

Problem 1.

Mark: 80

The dataset in `bank − marketing − campaign.zip` represents a collection of examples from a marketing campaign organised by a bank to get its clients to place a term deposit. The dataset has 21 columns, described in Table 1.

Your task is to build a classifier based on this dataset. The classifier should use the *regularised logistic regression* algorithm, with the regularised *cross-entropy* as its cost function. You will use the sum of the magnitude of all the coefficients (also known as `Lasso` regularisation or $\mathtt{L}_1$ regularisation) as your regularisation technique.

Assessment Criteria

The following criteria will be followed to assess your submission:

- Data cleaning and preparation in `Julia`;
- Implementation (from scratch) of the *regularised logistic regression* algorithm in `Julia`;
- Design and implementation of the classifier;
- Performance metrics[1], including:

  accuracy: the proportion of correct predictions (clients correctly predicted to have placed a term deposit or not) over all predictions;

  precision: the proportion of clients the classifier predicted have placed a term deposit actually did so;

  recall: the proportion of clients that actually placed a term deposit which was predicted by the classifier.

[1] It is advised to use a confusion matrix.

Table 1: Dataset Metadata

Column Name	Description	Type	Column Name	Description	Type
age	age	numeric	job	type of job	categorical
marital	marital status	categorical	education	education	categorical
default	credit default?	categorical	housing	housing loan?	categorical
loan	peronal loan?	categorical	contact	contact type	categorical
month	last contact month	categorical	day_of_week	last contact day of week	categorical
duration	last contact duration	categorical	campaign	total number of contacts	numeric
pdays	number of days from previous campaign	numeric	previous	number of contacts before this campaign	numeric
poutcome	outcome of previous campaign	categorical			
emp.var.rate	employment variation rate	numeric	cons.price.idx	consumer price index	numeric
euribor3m	euribor[a] 3 month rate	numeric	nr.employed	number of employees	numeric
y	has the client placed a term deposit?	binary			

[a]Euribor stands for Euro Interbank Offered Rate. Note that the campaign took place in Europe.

Submission Instructions

- This project is to be completed by groups of maximum two (2) students each.
- For each group, a repository should be created either on `Github`[2] or `Gitlab`[3]. The URL of the repository should be communicated by Thursday, May 14^{th} 2020, with all group members set up as contributors.
- The submission date is Monday, May 25^{th} 2020, midnight.
- A submission will be assessed based on the clone of its repository at the deadline.
- Any group who fails to submit on time will be awarded the mark 0.
- There should be no assumption about the execution environment of your code. It could be run using a specific framework or simply on the command line.
- In the case of plagiarism (groups copying from each other or submissions copied from the Internet), all submissions involved will be awarded the mark 0, and each student will receive a warning.

[2] `https://github.com`
[3] `https://about.gitlab.com/`