MLPA Final Assignment

Submission Deadline: Monday 29th April 2019 at 11:59 p.m.

NOTE: This document is 4 pages long. Make sure you read all 3 pages as you may lose marks if you do not follow the instructions correctly.

Task Description: In this assignment you will develop a classifier that uses data to predict the outcome of a Bank marketing campaign.

Team work: Find a partner to work on this project. (Two person teams are encouraged, though you may work alone if you prefer. No >= 3 person teams please). Be aware, however, that if you work as a team everyone on the team gets the same mark, irrespective of who does what. Also, if you have a falling out with your teammate I will not be involved in resolving the problem. Also, the marking is the same irrespective of whether you work by yourself or with another person.

What you are given:

Thre are two files available on webcourses:

1. *trainingset.txt*: this file contains the training instances. This file lists the descriptive features and the target feature level for each instance.
2. *testset.txt:* use this file to test accuracy of your model
3. *queries.txt*: this file contains the query instances. This file lists the descriptive features for each instance. However, the target feature level has been overwritten with '?'

Submission Deadline and Late Submissions:

• **Deadline:** Monday 29th April 2019 at 11:59 p.m.
• Marks will be deducted for late submission - 10% per day late.

How do you submit your solution?

• You can submit your assignment work through the Assignment Submission Form I have set up on the Webcourse module.

What do you need to submit:

If you are using a classifier seen in class, you need to submit 2 **separate** files (don't bundle them in a zip file submit the two files separately): (1) a solutions file, and (2) the Weka model file (descriptions below).

In case you are using a classifier not seen in class, you must submit the same 2 files as above plus a third file containing (3) a short description of the classifier.

1. **The solutions file:**

- Naming convention: This file should be named using the following convention
- 'studentnumber1+studentnumber2.txt', where studentnumber1 is the student number of the first member of the team and studentnumber2 is the student number of the other member of the team. For example, the file C1234567+D9876543.txt is the correct name for the solution file for a team comprising of the students C1234567 and D9876543. If you are working by yourself name your solution file 'studentnumber.txt', e.g. C1234567.txt
- Contents: The file should list your classifier's target variable predictions for each of the query instances in the queries.txt file. Each line in the file should list one query number followed by a comma followed by your classifier's prediction for that query, i.e.:

<query>,<prediction>

- The box below illustrates what someone looking at a portion of your solutions files should see

```
1,TypeA
2,TypeB
3,TypeB
4,TypeA
5,TypeA
```

2. **The model file for your classifier.**

a. Naming convention: This file should be named using the following convention**'studentnumber1+studentnumber2.model**', where studentnumber1 is the student number of the first member of the team and studentnumber2 is the student number of the other member of the team. For

example, the file C1234567+D9876543.model is the correct name for the model file for a team comprising of the students C1234567 and D9876543. If you are working by yourself name model file 'studentnumber.model', e.g. C1234567.model

b. Contents: Weka provides the functionality to save a model once it has been trained. It outputs this model into a .model file. Once you have selected the model type and hyperparameters that you think will give you the best results then you should train a model and use Weka to output this model. This is the file you should submit as part of you solution. This model should expect the data to be in same format as the one I have shared them with you.

3. **The short description of your classifier (this is optional and only necessary if you are using a model we did not discuss in class)**

a. Naming convention: This file should be named using the following convention'**studentnumber1+studentnumber2.pdf**', where studentnumber1 is the student number of the first member of the team and studentnumber2 is the student number of the other member of the team. For example, the file C1234567+D9876543.model is the correct name for the model file for a team comprising of the students C1234567 and D9876543. If you are working by yourself name model file 'studentnumber.model', e.g. C1234567.model

b. Contents: This should be a short paragraph describing the algorithm you have used and why you chose it. You can base your description on Weka's documentation regarding the classifiers. This is in order to avoid people randomly trying classifier and not understanding why they are working better (or worse) than others in this particular assignment.

Marking Scheme

Marks are awarded based on the accuracy of the classifier.

Marks may be deducted for the following reasons:

(a) Late submission (including submissions that are incomplete by the time the deadline has passed): 10 marks per day late.

(b) Incorrect submission: 10 marks will be deducted if your submission is does not follow the stated formats. The reason that I do this is that if you do not correctly format you submission this slows down the correction process for everyone. I will be strict in deducting these marks. If your submission does not follow the guidelines

these marks will be deducted. Examples of the types of errors that will result in these marks not be awarded include:

• Solution file named incorrectly
• Leaving blank lines between solutions in the solutions file
• Having trailing blank spaces after key values before commas in the solutions file
• Forgetting to put commas between the fields in the solutions file
• The solutions file not being a .txt file, for example submitting your solution as an .rtf or other file format (or bundling your files as a zip file submission)
• The description file not being a .pdf or .doc file, for example submitting your solution as an .rtf or other file format (or bundling your files as a zip file submission)