Technological University Dublin

ARROW@TU Dublin

Assessment & Feedback Cases

Learning & Teaching Practice Exchange

2014

Practical Work with Report

Brian Mac Namee
Technological University Dublin, brian.macnamee@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/ltcassess

Part of the Higher Education Commons

Title: Practical Work with Report

Lecturer: Brian MacNamee

Programme and year on which assessment was offered

MSc Computing

Description

A short piece of writing on a particular subject.

What have you found are the advantages of using this form of assessment?

- Tests students skills to develop machine learning solutions to "real-world" problems
- Lots of flexibility in the work that students could do
- Connects classroom work to the bigger machine learning practice and research communities

What have you found are the dis-advantages of using this form of assessment?

- Flexibility was a challenge for some students
- Marking was challenging as students could pick very different projects to work on, some more difficult than others.
- Students could use tools which meant solutions could be developed without huge understanding of machine learning

Alternatives

Essay type assessments can be separated into collection /review of relevant resources, development of plan/schedule/synopsis of proposed work, concept map/flow chart of work in progress, executive summaries, bibliographies etc

Assessment in Practice

- Not suitable for very large classes due to the report nature of the submission and the marking burden.
- Suitable for group work.

Assessment Time

- Preparation time: ~4 hours
- Student time to complete: ~20 hours
- Marking time: ~8 hours
- Ease of Feedback: Written feedback to each group

Additional Resources

Assignment Description

2011 – 2012 Machine Learning Assignment
Become A Machine Learning Superstar!

In this assignment you will be required to demonstrate your knowledge of machine learning techniques, your ability to apply technical skills to build actual machine learning models, and your capacity to evaluate state-of-the-art machine learning approaches. The assignment is composed of three tasks each of which is based on real applications of machine learning.

Marking Scheme

Marks will be allocated as follows:

- Question 1: Similarity Based Learning **(25%)**
    - Case structure: 15%
    - Similarity measures: 10%

- Question 2: Information based, Probability & Error Reduction Based Learning **(55%)**
    - Implementation: 25%
    - Kaggle standing: 5%
    - Model evaluation: 15%
    - Discussion: 10%

- Question 3: Advances in Machine Learning **(20%)**
    - Description of approach: 8%
  - Description of results: 8%
  - Discussion of applicability to Kaggle problem: 4%

Submission

Important details are as follows:

- The assignment can be undertaken individually or in groups of **no more than 3**.
- Plagiarism will not be accepted and will result in zero marks.
- Submission will be in the form of a document and should be made through the module webcourses site.

Important Dates

Important dates are as follows:

- Assignment handout date: 1st March, 2012
- Submission date: 20th April, 2012

Submission

Important details are as follows:

- The assignment can be undertaken individually or in groups of **no more than 3.**
- Plagiarism will not be accepted and will result in zero marks.
- Submission will be in the form of a document and should be made through the module webcourses site.

Important Dates

Important dates are as follows:

- Assignment handout date: 1st March, 2012
- Submission date: 20th April, 2012

Question 1: Similarity Based Learning

Table 1 shows an extract from a dataset (a larger extract from the dataset is given in Appendix A) that has been provided to you in order to build an SMS spam detection system using techniques from **case based reasoning.** The dataset has 6 descriptive features and one target feature, **Type**, indicating whether a message is a *spam* or a *ham* (genuine) message. The descriptive features are:

- **Text:** the message text itself
- **Time:** the time the message was received at
- **Location:** the location from which the message was sent
- **Number:** the number from which the message was sent
- **Router hops:** the number of network routers the message visited in getting from the sender to the receiver's phone

Table 1: An extract from an SMS spam detection dataset

Text	Time	Location	Number	Router Hops	Type
Oi when you gonna ring	14:11	Dublin	087 4520256	1	Ham
Yes but can we meet in town cos will go to gep and then home. You could text at bus stop. And don't worry we'll have finished by march … ish!	10:39	Sligo	089 4564563	2	Ham
V nice! Off 2 sheffield tom 2 air my opinions on categories 2 b used 2 measure ethnicity in next census. Busy transcribing. :-)	13:57	Dublin	088 1974937	4	Ham

07734396839 IBH Customer Loyalty Offer: The NEW NOKIA6600 Mobile from ONLY £10 at TXTAUCTION!Txt word:START to No:81151 & get Yours Now!4T&	6:30	Dublin	088 8682000	4	Spam
I don't know u and u don't know me. Send CHAT to 86688 now and let's find each other! Only 150p/Msg rcvd. HG/Suite342/2Lands/Row/W1J6 HL LDN. 18 years or over.	4:42	Cork	089 9092835	1	Spam
goldviking (29/M) is inviting you to be his friend. Reply YES-762 or NO-762 See him: www.SMS.ac/u/goldviking STOP? Send STOP FRND to 62468	5:35	Unknown	088 9038000	5	Spam

Based on the scenario given above you must perform the following tasks:

a. Design a **case structure** that could be used to represent the text messages in the case base to be used for this problem, providing justifications for all decisions that you make, and examples of what this structure might look like. **(You do not need to implement this case structure.)**

b. Design a **similarity measure** that could be used to compare cases in this case base, providing justifications for all decisions that you make, and simple examples of how this similarity measure would work. **(You do not need to implement this similarity measure.)**

Steps

Steps for this part of the assignment are as follows:

- Design a case structure
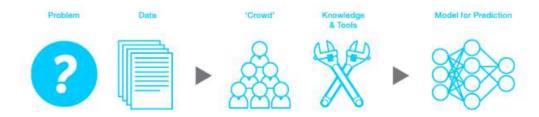- Design a similarity measure

Submission

Submission for this part of the assignment should be a document of no more than **three pages** describing your approaches to (a) and (b) above.

Question 2: Become a Kaggle Machine Learning Superstar

As described on the Kaggle website[1]: *"Kaggle is an innovative solution for statistical/analytics outsourcing. We are the leading platform for predictive modeling competitions. Companies, governments and researchers present datasets and problems - the world's best data scientists then compete to produce the best solutions. At the end of a competition, the competition host pays prize money in exchange for the intellectual property behind the winning model."*

This part of the assignment requires you to compete in a Kaggle competition. All Kaggle competitions require you to build a predictive machine learning model which will be evaluated against other Kaggle competitors.

Steps

Steps for this part of the assignment are:

- Form a team (you should use the same team for all sections of this assignment) and register on Kaggle
- Pick a Kaggle contest to enter
  - Pick whichever contest interests you most - aim high!
  - Be very careful of the dates associated with the contest you choose. Although the contest does not have to be complete before the end of this assignment, make sure that the Kaggle contest dates will give you enough time to work on your model before the contest closes.
- Develop models to address the problem in the contest
  - You can use any machine learning tools you feel are appropriate, but SAS Enterprise Miner or the Weka toolkit are recommended.
- Perform a local evaluation of your models separate to the Kaggle evaluation
  - Use whatever approach to evaluation you think is most suitable
- Submit models to be evaluated by Kaggle

Submission

The primary submission for this part of the assignment will be a document of no more than **eight pages** (including all diagrams and references) describing the following:

---

[1] http://www.kaggle.com/about

- The Kaggle contest you entered, and why you chose it
- The model that you build including important parameters and justification for all choices (include screenshots from whatever tools used as appropriate)
- A discussion of the local evaluation strategy you adopted
- A discussion of the evaluation results
- A screenshot of your Kaggle performance report
- Suggestions for what you might do to continue the work if you had time

After the submission date for the assignment there will be an in-class discussion session on all of the Kaggle contests entered. At this discussion your team will be asked to briefly describe the contest you entered and your approach.