

Faculty of Engineering, Environment and Computing
7072CEM - MACHINE LEARNING

Assignment Brief 2019/20

Module Title MACHINE LEARNING	Individual/Group Project (2 people), but individual paper submission.	Cohort Sep/Jan	Module Code 7072CEM
Coursework Title Machine learning algorithms for solving real-world classification and clustering problems			Hand out date: 16th September 2019
Module Leader Dr. Vasile Palade			Due date: 5th October 2019, 6.00pm
Word Limit: 4-6 A4 pages	Coursework type: Assignment		% of Module Mark 100%
<ul style="list-style-type: none">• Submission arrangement: online via CU Moodle.• File types and method of recording: Submit your report as a PDF or Word document using the 'Assignment' link in the 7072CEM Moodle page.• Mark and Feedback date: 2 weeks after submission• Mark and Feedback method: provided in Moodle			

Module Learning Outcomes Assessed:

1. Examine the fundamental concepts of machine learning, their implementation and application.
2. Prepare appropriate preparation of data sets and evaluate the performance of different learning algorithms on these data sets.
3. Appraise the different learning methodologies, their associated algorithms and their appropriateness to solve real-world problems.
4. Select and apply learning algorithms to various practical scenarios and evaluate their performance.
5. Critique trends in the current machine learning developments.

Task and Mark distribution:

1. See next pages

Notes:

1. Please notify your registry course support team and module leader for disability support.
2. Any student requiring an extension or deferral should follow the university process.
3. The University cannot take responsibility for any coursework lost or corrupted on disks, laptops or personal computer. Students should therefore regularly back-up any work and are advised to save it on the University system.
4. If there are technical or performance issues that prevent students submitting coursework through the online coursework submission system on the day of a coursework deadline, an appropriate extension to the coursework submission deadline will be agreed. This extension will normally be 24 hours or the next working day if the deadline falls on a Friday or over the weekend period. This will be communicated via email and as a CU Moodle announcement.

7072CEM COURSEWORK:

MACHINE LEARNING ALGORITHMS FOR SOLVING REAL-WORLD CLASSIFICATION AND CLUSTERING PROBLEMS

Title: **Machine learning algorithms for solving real-world classification and clustering problems**

Individual Research Paper: 100% of the module mark

CONTEXT

During this module, you learned about different machine learning techniques, associated concepts and applications. We explored a number of classification algorithms, such as Generalized Logistic Regression, Linear Discriminant Analysis, Optimized K-nearest Neighbour, Bayesian and Statistical Methods, Support Vector Machines and Decision Trees. Also, we covered clustering algorithms, such as K-means, and feature selection and extraction methods, such as PCA. In this assignment, you will have to select an application related to a classification, clustering, or anomaly detection problem, and explore how best to apply machine learning algorithms to solve it. Basically you are free to choose any of the following datasets (or another one

agreed with your tutor in advance) and apply 3 classification or clustering techniques on it.

1. [Bags of Words](#) (Classification)
2. [Daily and Sports Activities Dataset](#) (Classification and Clustering)
3. [Dresses Attribute Sales Dataset](#) (Regression)
4. Or [other](#) (choose as desired but match techniques with the dataset)

You can combine and choose from the above mentioned algorithms or you can use or come up with a new classification or clustering algorithm!

THE PURPOSE OF THIS COURSEWORK IS TO

- Examine the fundamental concepts of machine learning, their implementation and application.
- Perform appropriate preparation of a dataset and evaluate the performance of different learning algorithms on this dataset.
- Gain practical experience in selecting machine learning algorithms for solving a real-life classification or clustering problem.
- Demonstrate effectiveness in project teamwork and leadership.

YOU WILL BE REQUIRED TO:

- Work in groups of 2, 1, or 3 (3 not preferred, and only when the project is very complex, and it should be approved by the lecturer);
- Actively participate in all activities;
- Submit progress on your work regularly to get formative feedback and improve the final submission;
- Give some group presentation if required, which will contribute to your formative assessment.

Your final submission will include a scientific paper of up to 6 A4 pages (written individually), based on the experience and results gained during the project work. You will have to acknowledge the contributions of all group members in your paper.

You are encouraged to target a certain conference or journal and submit the proposed paper to it. Submission guidelines can be found on the conference or journal web page you choose to submit to.

List of reputed conferences and journals:

- 1- [IJCNN Conference](#)
- 2- [NIPS Conference](#)
- 3- [International Conference of Machine Learning](#)
- 4- [Machine Learning Journal](#)
- 5- [Neural Networks Journal](#)
- 6- Others (please let us know)

THE PAPER SHOULD BROADLY INCLUDE THE FOLLOWING SECTIONS:

- Abstract
- Introduction (where you introduce the problem along a short literature review of related work; if the literature review is longer, it is recommended to be a section on its own)
- Problem and Data set(s) description (where you describe in detail the problem you want to solve and its significance)
- Methods (where you shortly describe the machine learning methods and/or other methods employed to solve the problem)
- Experimental setup (including data pre-processing, feature selection and extraction)
- Results
- Discussion and Conclusions
- References

These are generic section titles, which you may adapt appropriately to the application/problem that is being investigated. You may include sections describing modifications of algorithms or developments that are novel and specific to your work. You may include figures, tables, pseudo-code, and appendices with the actual code that has been developed.

More information of how to write a paper is available at the following link: "[Crafting Papers on Machine Learning](#)", by Pat Langley.

You will need to follow the formatting guidelines of the [IEEE Manuscript Template for Conference Proceedings](#) (A4)

The group project general guidelines and milestones:

Please note, the following guidelines are good practice and should lead to better result, but you have the freedom to pick whatever is suitable for your style:

- Working in groups of 1, 2 (or 3), you have to select a real-world classification/clustering problem and one or more appropriate dataset(s) as suggested above. You may also use the following links, which have numerous problems and datasets:
 - UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/>;
 - ICML 2019 accepted papers: <https://icml.cc/Conferences/2019/Schedule?type=Poster>;
 - Kaggle competitions: <http://www.kaggle.com/competitions>;
 - Stanford machine learning projects: <http://cs229.stanford.edu/projects2013.html>, <http://cs229.stanford.edu/projects2012.html>, <http://cs229.stanford.edu/projects2011.html>, <http://cs229.stanford.edu/projects2016.html>.
- You will write a proposal (maximum of 1 A4 page), giving the title of the project, the names of all group members, the description of the problem and the plan of the work. You will need to submit this proposal to your tutor for formative feedback by Wednesday afternoon week 1.
- You will need to investigate and read related work in the next 2 days. By Friday afternoon, week 1, you have to submit an individually written short literature review of your findings in order to get formative feedback.
- In the following 2 weeks you have to select, implement and apply appropriate machine learning algorithms to the selected problem, performing data pre-processing, if needed, and record the results from the experiments.
- You will receive regular feedback on your progress from the module leader or tutor in the labs in week 1, and in the following 2 weeks via some scheduled meetings with the module leader.
- In week 3, you have to write up your final paper, and submit it by Friday 5th October 6.00pm.

MARKING CRITERIA FOR THE PAPER

Criterion	Mark
Technical quality (rigour of the experiments, data preparation, correct application of the selected algorithms and suitability of the selection). You have to provide in appendices evidence of running the experiments.	35%
Evaluation and discussion of the significance of the results (Why the results are important? How does the paper advance the state of the art? How would the results be useful to other researchers or practitioners? Is this a “real” problem or a small “toy” problem?)	25%
Social, ethical, legal and professional considerations related to the problem in question.	5%
Clarity of the writing (Is there sufficient information for the reader to reproduce the results?), references and presentation (Are results clearly presented, with appropriate visualisations?).	15%
Originality (Is there some novel approach to the problem, novel use of techniques? Is there difference from previous contributions?)	20%