

UNIVERSITY OF CAPE TOWN
DEPARTMENT OF STATISTICAL SCIENCES
MASTERS IN DATA SCIENCE STA5077Z
UNSUPERVISED LEARNING
Due date: Monday, 2 September 2019 at 4pm

Late submissions will not be accepted.

Instructions:

- Prepare reports for both topics below. All computer code must be submitted as an appendix and should include comments that clearly explain the purpose of the code.
- You are expected to work on this on your own. Please attach a plagiarism declaration to your report. This can be collected from stats reception. Plagiarism of any form will be reported to the university court. ASSIGNMENT 2: Insert Your Own Problem Here (for SOMs and CLUSTERING) You are required to source an interesting unsupervised learning problem of your own. Your dataset could come from your organisation or from an online repository such as www.kaggle.com/ <https://www.datafirst.uct.ac.za/> . You should model these data using Self Organising Maps and any other Clustering method and identify the best method for your problem. You must submit a neat report with your findings. In your report, you should:
 - Very clearly explain your problem. Remember that your marker will not be familiar with your specific problem
 - Clearly describe your dataset, both in writing and with graphs and/or tables ,
 - Clearly motivate your model choices and hyperparameter settings.
 - Clearly explain your findings.
 - Use appropriate graphs and tables to summarise your findings.

Each student should work on a different dataset. If you intend to use a dataset from a public repository, please let your lecturer know which dataset you are using so that it can be reserved for you only (first come, first served!). If you intend on working with confidential data from your organisation, note that only the course lecturers will have access to your final report and that this will be returned to you once it is marked. The data and findings will be treated as confidential and will not be disclosed to anyone else at any time.