Adaptive Computation and Machine Learning
COMS 4030A and COMS7047A

PROJECT

The project for ACML requires you to analyse a dataset using machine learning techniques and write a report describing your methods and results.
This is a **group** project. You should organise yourselves into groups of size 4 or less.

1) Source a dataset of sufficient complexity and interest.
See, e.g., UCI machine learning laboratory, or kaggle.com, but you are free to obtain datasets from elsewhere, as long as no ethics clearance is required.
You must state clearly where the dataset was obtained.

2) Describe the contents of the dataset and your objectives with respect to the dataset.
(Do not do too much data exploration ...)

3) Describe the required preprocessing of the dataset and the split of data into train/validate/test.

4) Describe the choice of model and machine learning algorithm used. Describe the implementation of the model. Python libraries such as Keras and Pytorch are allowed. The model should be sufficiently complex.
You may use CNN's, RNN's, autoencoders, GANS, or other complex architectures, if desired.
Please **do not** just use an existing model from the internet.

5) Describe any experiments that were done in fine-tuning the model or the hyperparameters.

6) Present various graphs that were generated during the training of the model.

7) Present the results of the trained model on the test dataset, using both text and visuals, and provide a short analysis of the results.

8) Put all the relevant information from the above into a **report**.

Required submission: Your report and your code.

Due date for submission: Friday 30 May, at 17h00.