

## Group coursework 2

- Please upload your report in **one pdf file** with no more than 5 pages long including plots, of which no more than 3 pages should be text.
- Please also upload the script for the analysis.
- This coursework will be assessed based on the quality of your report, e.g. presentation, quality of plots, details about the experiment settings, motivations to use the methods, analysis of the results, conclusions etc.
- Make sure that you have included **sufficient** comments in the codes to make them **readable** by other people. There should be **no error messages** shown when I run your scripts. You can assume that I have installed all required packages.

### Question 1 [25 marks]

Download the `IMDB-Movies.csv` data from Moodle. This dataset contains information of 50 top-rated movies from IMDb.

1. Apply PCA on the numerical variables excluding Year, and report your findings.  
[10 marks]
2. Cluster the movies based on the numerical variables excluding Year. Justify the reason of why you choose the methods and report your findings.  
[15 marks]