High Dimensional Statistics IV - MATH4287

Mini project report (20% of module's total mark):

Submission deadline via Gradescope: 12noon Friday 19th December 2025

The mini project should cover an aspect of high dimensional statistics, for example on any topic covered in the module such as

1. **High dimensional regularisation**,
   For example: lasso, ridge, elastic net.
2. **High dimensional PCA**,
   For example: standard PCA, sparse PCA.
3. **High dimensional clustering,**
   For example: K-means clustering, hierarchical clustering.

There is flexibility in terms of the "type of project", it can be just an applied report focusing on analysing a high dimensional data set using any of the above methods in the module or any other related topic (see next page). For this, a wide variety of high dimensional data sets would be appropriate, but please note that if you decide to write an applied report then it will be part of the mini project task to search and find a data set. There are many high dimensional data sets available online, as well as in those books in the reading list on Ultra. Of course, you can also use the real data examples in the lecture notes and computer practicals, but not just reproduce the same analysis. Or it can be just a theoretical report without a data analysis part. Obviously, a theoretical report should not be just same material as lecture notes, and instead it can study some other aspects of the topic. For instance, you can work on sparse PCA, or elastic net, or any other method that we would only have to briefly cover in the module due to time limitation. It will certainly be acceptable if a report contains both theory and application with any proportion of both.

The main criteria for marking mini project reports are as follows:

1. *Quality of content and quality of writing* (40% of the report mark),
2. *Correctness, and consistency of the notation and presentation* (30% of the report mark),
3. *Clear conclusions of the project* (see below) (15% of the report mark),
4. *Length of report* (see below) (15% of the report mark).

The report must be **4 pages** (min) **to 6 pages** (max) including any list of references, and any additional pages will be ignored during marking. To write your mini report,

start with a brief introduction and mention the objectives and questions you want to answer in the mini project, and proceed with any literature review and then explain your main methods and if needed add a bit of theory and data analysis or comparisons depending on your mini project topic, and finally interpret your results and findings and provide some conclusions and any references. There is an option to include an **appendix of maximum 4 pages**, but it must be clearly headed as Appendix, so the maximum length of a mini project report containing an appendix is 10 pages. Note that there will be no penalty for not including an appendix and likewise having an appendix does not mean a higher mark. The mini project mark will be given based on the four criteria mentioned above, and an appendix is just to help provide some space for reporting additional results if required. For example, the appendix can help you to report any additional results if you wish to report them or can be used for additional mathematical proofs or codes or anything else. There should be **a short conclusions section** in the end summarising the main conclusions of the mini project. Ideally, conclusions should be presented via some bullet points to make them clear. Failing to provide a conclusions section will affect the mark as above, as the idea is to highlight any findings achieved during the mini project.

For more curious students, it is also acceptable to work on any related topic to the module if the topic concerns high dimensional statistics. For example, you may write a report about other topics such as high dimensional logistic regression, high dimensional classification, high dimensional Bayesian statistics, other regularisation methods such as adaptive lasso and group lasso etc, or any other high dimensional techniques for example related to your MSc dissertation topic if applicable.

If you want some specific references, you can look at the “reading list” on Ultra which all should be freely available through library. Also, sometimes it helps to search online on the internet for additional resources.

Finally, given these details and examples, if you are still not sure about what topic you want to work on or if you have any specific questions, feel free to talk to me after lectures or in office hours (Fridays 1-2pm, Room MSC3079) or just drop me an email.