STATS 780/CSE 780, Winter 2021

ASSIGNMENT 5

Submit to Crowdmark using the link that was emailed to you.

Due at 10.00am on Friday, March 19th.

Your assignment must conform to the Assignment Standards listed below.

Assignments submitted up to 24 hours late will incur a 30% penalty.
Assignments submitted more than 24 hours late will receive a zero grade.

Find a data set that is suitable as a topic of a mixture discriminant analysis; this means, amongst other things, that more than one component should be fitted to at least one of the known classes. You may <u>not</u> use a data set that was discussed in class on March 12th. Briefly describe your chosen data set and clearly explain where it was sourced. For each of ten different random 75/25 labelled/unlabelled splits, apply mixture discriminant analysis as well as one other classification technique that was discussed during this course (e.g., a classification tree, random forest, etc.). Compare and contrast the classification performance of this other technique and mixture discriminant analysis. Clearly state what conclusions can be drawn from your analysis — both in the context of the two classification methods and your chosen data set.

The marking scheme for the assignment is as follows:

- Description of data including source, details on variables, relationships between variables, descriptive analyses, etc. [5 marks]
- Description of the problem being addressed. [3 marks]
- Description of the techniques used. [6 marks]
- Descriptions of results, both in technical terms and in the context of the data. [6 marks]
- Conclusions, in the context of the data and the problem being addressed. [5 marks]

Assignment Standards

- LATEX is strongly recommended but not strictly required.
- Eleven-point font (times or similar) must be used with 1.5 line spacing and margins of at least 1 inch all around.
- Your report may not exceed **three pages**, inclusive of tables, figures, and bibliography.

- Do not include a title page. The title and your name should be printed at the top of the first of the two pages.
- The writing and referencing should be appropriate to the graduate level.
- Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.