

MDM-2024 Homework N

Cuong Nguyen (101559968), Petteri Raita (ID2), and Raihan Gafur (ID3)

September 21, 2024

1 General instructions

In the beginning, tell always the course name, homework number and the authors' names and student identifiers. You can use a separate cover page for these if you like.

The report structure (sections) is described in the task. Typically, the first section is “Methods”, where you should describe briefly how you solved the task. Mention here if you used some tool or made a script (given in the appendix) and give references if you used any external sources (i.e., other than the textbook and other course material).

The results are usually presented in several sections. In addition to text, there may be tables or figures. Remember to give table row and column labels and add required labels to plots (e.g., axis names in scatter-plots) so that they are comprehensible. Mention important observations in tables or figures in the text. Please try to make a good layout. A table or figure in the middle of the page surrounded by blank is not reader-friendly. Often, you are asked to interpret or discuss the results. Try to explain *why* the things happen instead of only describing what has happened.

The last section is usually Appendix, where you should include the code you used to produce your results. (In some tasks, there may be multiple appendices.) If you used multiple external sources, give the list of references in the end under caption “References”. (One reference can be included in “Methods” as well.)

Remember all good habits that make presentation easily readable to another person (e.g., rounding numbers to suitable precision, introducing your own abbreviations). When using library functions (e.g., for goodness measures), remember to check that the function really calculates what is asked (the name of the function does not guarantee this). If there are alternative ways to do something, tell which one you chose (e.g., did you use biased or

corrected estimates of standard deviation), because they can lead to different results.

2 Examples of tables and figures

Table 1 and Figure 1 show examples of including tables and figures.

Table 1: Comparison of three clustering methods for $K = 9$. SI=Silhouette index, CH=Calinski-Harabasz index, DB=Davies-Bouldin index.

method	SI	CH	DB
K -means	0.541	182.0	0.692
single-link	0.389	48.9	0.510
complete-link	0.469	129.5	0.731

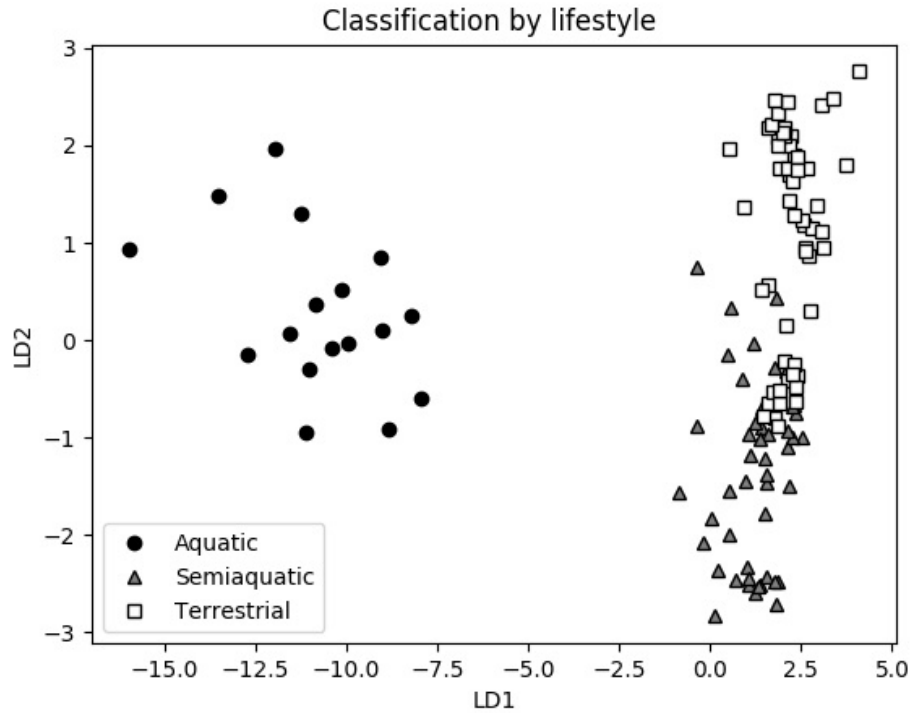


Figure 1: Scatter plot of the data points and their classes along the first linear discriminants (LD1 and LD2).

3 Examples of references

Here is an example of referring to the original papers where Silhouette [3] and Calinski-Harabasz [1] indices were introduced. Use references also to give credit to your sources. E.g., if you wanted to explain the intuition of Calinski-Harabasz index and found useful information in CrossValidated and wikipedia, you can write like this

“Glen_b in CrossVlidated [2] gave a useful hint that CH is actually analogous to the F -ratio in ANOVA...” “According to Wikipedia [4], the ANOVA F -test statistic is...

4 Some math examples and macros

Typing math is probably the funniest thing with latex! Furthermore, you can define new commands for notations you need often. In the prememable, we have given new commands for **M**

$$\mathbf{M} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

M

and two alternative definitions for the vector, \bar{x} and \mathbf{x} (to choose from). Here is one example equation (cumulative hypergeometric probability):

$$p_F = \sum_{i=0}^J \frac{\binom{fr(\mathbf{X})}{fr(\mathbf{X}C)+i} \binom{fr(\neg\mathbf{X})}{fr(\neg\mathbf{X}\neg C)+i}}{\binom{n}{fr(C)}},$$

where $J = \min\{fr(\mathbf{X}\neg C), fr(\neg\mathbf{X}, C)\}$.

References

- [1] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27, 1974.
- [2] Glen_b. Intuition behind the calinski-harabasz index. Answer in CrossValidated 2014. <https://stats.stackexchange.com/questions/97429/intuition-behind-the-calinski-harabasz-index>, accessed 10th Sep 2024.

- [3] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
- [4] Wikipedia contributors. F-test, 2024. <https://en.wikipedia.org/wiki/F-test>, Accessed 29th Aug 2024.