Geometric Foundations of Data Analysis I (CS4102)

Joshua Maglione

Semester 1 (2025)

Module information:

Meeting Times:

Mondays	3:00pm – 3:50pm
Wednesdays	10:00am – 10:50am

Room: Áras de Brún 1020,

Contact: joshua.maglione@universityofgalway.ie

Website: Canvas and https://joshmaglione.com/2025CS4102.html

Topics: We will cover four key methods of data analysis:

1. Least Squares Fitting,
2. Principal Component Analysis,
3. Hierarchical Clustering and Persistence,
4. Nearest Neighbors and the Johnson–Lindenstrauss Theorem

We will also explore these topics in Python using Jupyter Notebooks.

Assessment: The total assessment of the module comprises three components: three in-class quizzes (20%), an in-class exam (40%), and a final exam (40%). We will have a quiz every other week, for a total of six quizzes. The three lowest scores will not be counted. The in-class exam is tentatively scheduled for the 22nd of October (Wednesday in week 7); confirmed by the second week.

Reading list: The following list is not required but could be useful.

1. Blum, Avrim; Hopcroft, John; and Kannan Ravindran. *Foundations of Data Science*. 2018. https://www.cs.cornell.edu/jeh/book.pdf.
2. Hastie, Trevor; Tibsshirani, Robert, and Friedman, Jerome. *The Elements of Statistical Learning*. Second edition – 12th printing, Springer Ser. Statist. Springer, New York, 2009. https://hastie.su.domains/ElemStatLearn/printings/ESLII_print12_toc.pdf.

3. Jolliffe, I. T. *Principal Component Analysis*. Second edition, Springer Ser. Statist. Springer-Verlag, New York, 2002.
4. Margalit, Dan and Rabinoff, Joseph. *Interactive Linear Algebra*. 2019. https://textbooks.math.gatech.edu/ila/.
5. Phillips, Jeff M. *Mathematical Foundations for Data Analysis*. Springer Ser. Data Sci. Springer, Cham, 2021. https://mathfordata.github.io/versions/M4D-v0.6.pdf.
6. Shlens, Jonathon. *A tutorial on principal component analysis*. Preprint (2014). https://arxiv.org/abs/1404.1100.