

METRIC CLUSTER ANALYSIS OF MOLECULAR BIOLOGY DATA FOR SYSTEMS IDENTIFICATION OF GENE REGULATORY NETWORKS

JIM E. BLEVINS

ABSTRACT. Our talk develops metrics for cluster analysis—with a view towards molecular systems-biology. We discuss negative type and other notions from “the sociology of metric spaces” (Gromov).

Cluster analysis uses dissimilarities, distances, and metrics. Metrics facilitate the analysis of clustering methods.

The mathematical theory of metric spaces has been used to characterize the applicability of statistical procedures. For example, least-squares methods rely on the geometric properties of inner-product spaces. In Banach spaces, the strong law of large numbers depends on Beck convexity. In metric spaces, the distance-covariance test of dependence relies on the negative type (Lyons 2013; Janson 2021); characterizing metric spaces that isometrically embed into Euclidean spaces, Schoenberg’s negative type is used in computer science for the design & analysis of approximation algorithms (Deza & Laurent 1997).

REFERENCES

- [1] Benyamini, Y., and J. Lindenstrauss, *Geometric Nonlinear Functional Analysis. Vol. 1*, 2000
- [2] Blumenthal, L.M., *Theory and Applications of Distance Geometry*, 1953 [Chelsea 1970]
- [3] Chipman, H., Trevor J. Hastie, and Robert Tibshirani 2003. Clustering microarray data, pp. 159–200. In Terry Speed, ed. *Statistical Analysis of Gene Expression Microarray Data*. 2003
- [4] Deza, Michel M. and Monique Laurent. *Geometry of Cuts and Metrics*. Berlin: Springer 1997
- [5] Diaconis, Persi, *Group Representations in Probability and Statistics*, IMS 1988
- [6] Gatto, Francesco, Intawat Nookaew, and Jens Nielsen. Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proceedings of the National Academy of Sciences* **111**, no. 9 (2014): E866–E875.
- [7] Grenander, Ulf. *Probabilities on algebraic structures*. Wiley/Almqvist & Wiksell 1963
- [8] Ishchuk, Olena P., Iván Domenzain, Benjamín J. Sánchez, Facundo Muñoz-Paredes, José L. Martínez, Jens Nielsen, & Dina Petranovic, Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences* **119**, no. 30 (2022): e2108245119
- [9] Janson, Svante, On distance covariance in metric and Hilbert spaces, *Revista Latino-Americana de Probabilidade e Estatística* **XVIII** (2021): 1353–1393
- [10] Lyons, Russell, Distance covariance in metric spaces, *Ann. Prob.* **41**,5 (2013): 3284–3305.
- [11] Peirce, Charles Sanders. *The Cambridge Conferences Lectures of 1898*. Harvard UP 1992

STATISTICS, UPPSALA (HUMAN METABOLIC ATLAS & ELYPTA TEAMS, JENS NIELSEN GROUP, SYSTEMS AND SYNTHETIC BIOLOGY, BIOLOGY AND BIOLOGICAL ENGINEERING, CHALMERS)
Email address: James.Blevins.0@gmail.com

2020 *Mathematics Subject Classification*. Primary 46B20, 62H11, 62H20, 62H30; Secondary 46B85, 92C40, 92C42, 93B11, 93B30.

Key words and phrases. Correlation of Data, Distanced based clustering algorithms, Similarity Measurements, Metric embeddings, Recognition of genes and regulatory elements.