

Algorithmic Data Analysis

Esther Galbrun

Spring 2024



UNIVERSITY OF
EASTERN FINLAND

References

E. Lehman, F. T. Leighton and A. R. Meyer, *Mathematics for Computer Science*. MIT OpenCourseWare, 2015. URL: <https://courses.csail.mit.edu/6.042/spring15/mcs.pdf>.

P. N. Klein. *Coding The Matrix: Linear Algebra Through Computer Science Applications*. 2017. URL: <http://codingthematrix.com/>.

D. Prasad. *Elementary linear algebra*. Second edition. Alpha Science International Ltd., 2012.

C. C. Aggarwal. *Data mining: the textbook*. Springer, 2015.

J. Ellenberg. *How Not to Be Wrong: The Power of Mathematical Thinking*. (Quizz Q0a.10). Penguin Press, 2015.
URL: <https://medium.com/@penguinpress/an-excerpt-from-how-not-to-be-wrong-by-jordan-ellenberg-664e708cfc3d>.

S. Goldberg. *CAS CS 237: Probability in Computing*. (Exercise 3.3). Boston University, 2014.

C. C. Aggarwal. *Outlier analysis*. Springer, 2017.

J. Friedman, T. Hastie and R. Tibshirani, *The elements of statistical learning*. Springer, 2001.

M. Stamp. *A Revealing Introduction to Hidden Markov Models*. 2018. URL:

<https://www.cs.sjsu.edu/~stamp/RUA/HMM.pdf>.

J. Leskovec, A. Rajaraman and J. Ullman, *Mining of Massive Datasets, Chapter 4: Mining Data Streams*. (Exercises 3.2 and 3.4). 2014. URL: <http://www.mmds.org/>.

H.-P. Kriegel, P. Kröger and A. Zimek, *Outlier Detection Techniques*. 2010. URL: <https://archive.siam.org/meetings/sdm10/tutorial3.pdf>.

Elements of AI, online open course. URL: <https://course.elementsofai.com/>.

R. Berwick. *An Idiot's guide to Support vector machines (SVMs)*. URL: <http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf>.

I. Batal. *Introduction to Temporal Data Mining*. URL: <https://people.cs.pitt.edu/~iyad/temp.pdf>.

NIST/SEMATECH. *Handbook of Statistical Methods, Chapter 6.2: Introduction to Time Series Analysis*. 1998. URL:
<https://itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm>.

G. Hulten, L. Spencer and P. Domingos, “Mining Time-changing Data Streams.” In: *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '01. ACM, 2001, pp. 97–106. DOI: [10.1145/502512.502529](https://doi.org/10.1145/502512.502529).

P. Domingos and G. Hulten, “Mining High-speed Data Streams.” In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '00. ACM, 2000, pp. 71–80. DOI: [10.1145/347090.347107](https://doi.org/10.1145/347090.347107).

M. Datar et al., “Maintaining Stream Statistics over Sliding Windows.” In: *SIAM Journal on Computing* 31.6 (2002), pp. 1794–1813. DOI: [10.1137/S0097539701398363](https://doi.org/10.1137/S0097539701398363).

P. Patel et al., “Mining motifs in massive time series databases.” In: *Proceedings of the 2002 IEEE International Conference on Data Mining*. 2002, pp. 370–377. DOI: [10.1109/ICDM.2002.1183925](https://doi.org/10.1109/ICDM.2002.1183925).

J. Pei et al., “PrefixSpan;: mining sequential patterns efficiently by prefix-projected pattern growth.” In: *Proceedings of the 17th International Conference on Data Engineering*. 2001, pp. 215–224. DOI: [10.1109/ICDE.2001.914830](https://doi.org/10.1109/ICDE.2001.914830).

M. Vlachos et al., “Indexing Multi-dimensional Time-series with Support for Multiple Distance Measures.” In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '03. ACM, 2003, pp. 216–225. DOI: [10.1145/956750.956777](https://doi.org/10.1145/956750.956777).

T. de Vries, S. Chawla and M. E. Houle, “Finding Local Anomalies in Very High Dimensional Space.” In: *Proceedings of the 2010 IEEE International Conference on Data Mining*. 2010, pp. 128–137. DOI: [10.1109/ICDM.2010.151](https://doi.org/10.1109/ICDM.2010.151).

E. Keogh, J. Lin and A. Fu, “HOT SAX: efficiently finding the most unusual time series subsequence.” In: *Proceedings of the Fifth IEEE International Conference on Data Mining*. 2005. DOI: [10.1109/ICDM.2005.79](https://doi.org/10.1109/ICDM.2005.79).

M. Gupta et al., “Outlier Detection for Temporal Data: A Survey.” In: *IEEE Transactions on Knowledge and Data Engineering* 26.9 (2014), pp. 2250–2267. DOI: [10.1109/TKDE.2013.184](https://doi.org/10.1109/TKDE.2013.184).

F. T. Liu, K. M. Ting and Z.-H. Zhou, “Isolation-Based Anomaly Detection.” In: *ACM Trans. Knowl. Discov. Data* 6.1 (2012), 3:1–3:39. DOI: [10.1145/2133360.2133363](https://doi.org/10.1145/2133360.2133363).

F. T. Liu, K. M. Ting and Z.-H. Zhou, “Isolation Forest.” In: *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*. ICDM '08. IEEE Computer Society, 2008, pp. 413–422. DOI: [10.1109/ICDM.2008.17](https://doi.org/10.1109/ICDM.2008.17).

J. O. Lumbroso. “How Flajolet Processed Streams with Coin Flips.” In: *CoRR* abs/1805.00612 (2018). URL: <http://arxiv.org/abs/1805.00612>.

S. Fortmann-Roe. *Understanding the Bias-Variance Tradeoff*.

URL: <http://scott.fortmann-roe.com/docs/BiasVariance.html>.

X. Burret Sicotte. *Support Vector Machine: Python implementation using CVXOPT*. URL:

https://xavierbourretsicotte.github.io/SVM_implementation.html.

T. Fletcher. *Support Vector Machines Explained*. URL:
<https://static1.squarespace.com/static/58851af9ebbd1a30e98fb283/t/58902fbae4fcb5398aeb7505/1485844411772/SVM+Explained.pdf>.

A. Tulloch. *Support Vector Machine python implementation*. URL: <https://github.com/ajtulloch/svmpy>.

J. Brownlee. *Machine Learning Mastery: Machine Learning Algorithms From Scratch*. URL:
<https://machinelearningmastery.com/category/algorithms-from-scratch/>.

An Interactive Guide To The Fourier Transform. URL:
<https://betterexplained.com/articles/an-interactive-guide-to-the-fourier-transform/>.

J. Jordan. *Evaluating a machine learning model.* URL:
<https://www.jeremyjordan.me/evaluating-a-machine-learning-model/>.

Voronoi diagram python implementation. URL:
https://github.com/nlpenguin/py_geo_voronoi.

Matplotlib. URL: <https://matplotlib.org/>.

Basemap. URL: <https://matplotlib.org/basemap/>.

Scikit-learn. URL: <https://scikit-learn.org>.

UCI ML Repository. URL: <https://archive.ics.uci.edu>.

Kaggle. URL: <https://www.kaggle.com/datasets>.

Finnish Meteorological Institute. *Weather observations*. URL: <https://en.ilmatieteenlaitos.fi/download-observations#!/>.

Stony Brook University. *Outlier Detection DataSets (ODDS)*. URL: <http://odds.cs.stonybrook.edu/>.

GPSseuranta. *GPS-tracking for outdoor events*. URL: <https://www.tulospalvelu.fi/gps/>.

E. Keogh. *Time-series datasets*. URL:

<http://www.cs.ucr.edu/~eamonn/discords/>.

Department of Statistical Science, Duke University.

Time-series datasets. URL: [http:](http://www2.stat.duke.edu/~mw/ts_data_sets.html)

[//www2.stat.duke.edu/~mw/ts_data_sets.html](http://www2.stat.duke.edu/~mw/ts_data_sets.html).