

Machine Learning for Time Series

References

Laurent Oudre
laurent.oudre@ens-paris-saclay.fr

Master MVA
2023-2024

Session 1: Pattern Recognition and Detection

- [1.1] Berndt, D. J., & Clifford, J. (1994, July). Using dynamic time warping to find patterns in time series. In KDD workshop (Vol. 10, No. 16, pp. 359-370).
- [1.2] Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43-49.
- [1.3] Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on acoustics, speech, and signal processing*, 23(1), 67-72.
- [1.4] Mueen, A., Viswanathan, K., Gupta, C. K., & Keogh, E. (2015). The fastest similarity search algorithm for time series subsequences under Euclidean distance. <http://www.cs.unm.edu/~mueen/FastestSimilaritySearch.html>
- [1.5] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., ... & Keogh, E. (2012, August). Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 262-270).
- [1.6] Kim, S. W., Park, S., & Chu, W. W. (2001, April). An index-based approach for similarity search supporting time warping in large sequence databases. In *Proceedings 17th International Conference on Data Engineering* (pp. 607-614). IEEE.
- [1.7] Keogh, E., & Ratanamahatana, C. A. (2004). Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3), 358-386.
- [1.8] Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... & Keogh, E. (2016, December). Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 1317-1322). IEEE.
- [1.9] Grosse, R., Raina, R., Kwong, H., & Ng, A. Y. (2007, July). Shift-invariant sparse coding for audio classification. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* (pp. 149-158).
- [1.10] Wohlberg, B. (2014, May). Efficient convolutional sparse coding. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7173-7177). IEEE.
- [1.11] Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(1).
- [1.12] Daubechies, I., Defrise, M., & De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11), 1413-1457.

Session 2: Feature Extraction and Selection

- [2.1] Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6), 2297-2301.
- [2.2] Chung, Y. A., Wu, C. C., Shen, C. H., Lee, H. Y., & Lee, L. S. (2016). Audio word2vec: Unsupervised learning of audio segment representations using sequence-to-sequence autoencoder. *arXiv preprint arXiv:1603.00982*.
- [2.3] Vincent, L. E., & Thome, N. (2019). Shape and Time Distortion Loss for Training Deep Time Series Forecasting Models. In *Advances in Neural Information Processing Systems* (pp. 4189-4201).
- [2.4] Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.
- [2.5] Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.
- [2.6] Song, F., Guo, Z., & Mei, D. (2010, November). Feature selection using principal component analysis. In *2010 international conference on system science, engineering design and manufacturing informatization* (Vol. 1, pp. 27-30). IEEE.
- [2.7] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6), 1-45.
- [2.8] Gu, Q., Li, Z., & Han, J. (2011, July). Generalized Fisher score for feature selection. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence* (pp. 266-273).
- [2.9] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Session 3: Models and Representation Learning

- [3.1] Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2011). Time series analysis: forecasting and control (Vol. 734). John Wiley & Sons.
- [3.2] Hamilton, J. D. (1994). Time series analysis. Princeton University Press.
- [3.3] Priestley, M. B. (1981). Spectral analysis and time series: probability and mathematical statistics (No. 04; QA280, P7.).
- [3.4] Pankratz, A. (2009). Forecasting with univariate Box-Jenkins models: Concepts and cases (Vol. 224). John Wiley & Sons.
- [3.5] Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 233-244.
- [3.6] Bisgaard, S., & Kulahci, M. (2011). Time series analysis and forecasting by example. John Wiley & Sons.
- [3.7] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.
- [3.8] Bilmes, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510), 126.
- [3.9] Blumensath, T., & Davies, M. E. (2008). Iterative thresholding for sparse approximations. *Journal of Fourier analysis and Applications*, 14(5-6), 629-654.
- [3.10] Mallat, S. G., & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12), 3397-3415.
- [3.11] Daubechies, I., Defrise, M., & De Mol, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11), 1413-1457.
- [3.12] Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1), 183-202.
- [3.13] Boyd, S., Parikh, N., & Chu, E. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc.
- [3.14] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, 32(2), 407-499.
- [3.15] Tomic, I., & Frossard, P. (2011). Dictionary learning. *IEEE Signal Processing Magazine*, 28(2), 27-38.
- [3.16] Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2009, June). Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning* (pp. 689-696).

Session 4: Data Enhancement and Preprocessings

- [4.1] Rubinstein, R., Bruckstein, A. M., & Elad, M. (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6), 1045-1057.
- [4.2] Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis (Vol. 4)*. Cambridge university press.
- [4.3] Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.
- [4.4] Vautard, R., Yiou, P., & Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*, 58(1-4), 95-126.
- [4.5] Flandrin, P., Goncalves, P., & Rilling, G. (2004, September). Detrending and denoising with empirical mode decompositions. In *2004 12th European Signal Processing Conference* (pp. 1581-1584). IEEE.
- [4.6] Boudraa, A. O., & Cexus, J. C. (2006). Denoising via empirical mode decomposition. *Proc. IEEE ISCCSP*, 4(2006).
- [4.7] Comon, P. (1994). Independent component analysis, a new concept?. *Signal processing*, 36(3), 287-314.
- [4.8] Lepot, M., Aubin, J. B., & Clemens, F. H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water*, 9(10), 796.
- [4.9] McKinley, S., & Levine, M. (1998). Cubic spline interpolation. *College of the Redwoods*, 45(1), 1049-1060.
- [4.10] Srebro, N., & Jaakkola, T. (2003). Weighted low-rank approximations. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* (pp. 720-727).
- [4.11] Janssen, A. J. E. M., Veldhuis, R., & Vries, L. (1986). Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(2), 317-330.
- [4.12] Oudre, L. (2015). Automatic detection and removal of impulsive noise in audio signals. *Image Processing On Line*, 5, 267-281.

Session 5: Change-Point and Anomaly Detection

- [5.1] Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167, 107299.
- [5.2] Basseville, M., & Nikiforov, I. V. (1993). *Detection of abrupt changes: theory and application* (Vol. 104). Englewood Cliffs: prentice Hall.
- [5.3] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4), 523-527.
- [5.4] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their applications*, 83(1), 79-102.
- [5.5] Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47-78.
- [5.6] Lung-Yut-Fong, A., Lévy-Leduc, C., & Cappé, O. (2015). Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la Société Française de Statistique*, 156(4), 133-162.
- [5.7] Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1), 1-22.
- [5.8] Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590-1598.
- [5.9] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- [5.10] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- [5.11] Truong, C., Oudre, L., & Vayatis, N. (2017). Penalty learning for changepoint detection. In *2017 25th European Signal Processing Conference (EUSIPCO)* (pp. 1569-1573). IEEE.
- [5.12] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3), 1-58.
- [5.13] Roberts, S. W. (2000). Control chart tests based on geometric moving averages. *Technometrics*, 42(1), 97-101.
- [5.14] Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764-766.
- [5.15] Yamanishi, K., & Takeuchi, J. I. (2002, July). A unifying framework for detecting outliers and change points from non-stationary time series data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 676-681).
- [5.16] Hill, D. J., & Minsker, B. S. (2010). Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environmental Modelling & Software*, 25(9), 1014-1022.
- [5.17] Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... & Keogh, E. (2016, December). Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 1317-1322). IEEE.

Session 6: Multivariate Time Series

- [6.1] Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS*, 385-429.
- [6.2] Lütkepohl, H. (2013). Vector autoregressive models. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*. Edward Elgar Publishing.
- [6.3] Hurvich, C. M., & Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of time series analysis*, 14(3), 271-279.
- [6.4] Garcia-Cardona, C., & Wohlberg, B. (2018, October). Convolutional dictionary learning for multi-channel signals. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers* (pp. 335-342). IEEE.
- [6.5] Moreau, T., Oudre, L., & Vayatis, N. (2018, July). Dicod: Distributed convolutional coordinate descent for convolutional sparse coding. In *International Conference on Machine Learning* (pp. 3626-3634). PMLR.
- [6.6] Tropp, J. A., Gilbert, A. C., & Strauss, M. J. (2006). Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal processing*, 86(3), 572-588.
- [6.7] Tropp, J. A. (2006). Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. *Signal Processing*, 86(3), 589-602.
- [6.8] Cotter, S. F., Rao, B. D., Engan, K., & Kreutz-Delgado, K. (2005). Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53(7), 2477-2488.
- [6.9] Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., & Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5), 808-828.
- [6.10] Chen, S., Sandryhaila, A., Moura, J. M., & Kovacevic, J. (2014, December). Signal denoising on graphs via graph filtering. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (pp. 872-876). IEEE.
- [6.11] Narang, S. K., Gadde, A., & Ortega, A. (2013, May). Signal processing techniques for interpolation in graph structured data. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5445-5449). IEEE.
- [6.12] Mateos, G., Segarra, S., Marques, A. G., & Ribeiro, A. (2019). Connecting the dots: Identifying network structure via graph signal processing. *IEEE Signal Processing Magazine*, 36(3), 16-43.
- [6.13] Stankovic, L., Mandic, D. P., Dakovic, M., Ksil, I., Sejdic, E., & Constantinides, A. G. (2019). Understanding the basis of graph signal processing via an intuitive example-driven approach [lecture notes]. *IEEE Signal Processing Magazine*, 36(6), 133-145.
- [6.14] Dong, X., Thanou, D., Frossard, P., & Vandergheynst, P. (2016). Learning Laplacian matrix in smooth graph signal representations. *IEEE Transactions on Signal Processing*, 64(23), 6160-6173.