Curriculum Vitae: G.J.J. van den Burg

PERSONAL DETAILS

Gerrit Jan Johannes (Gertjan) van den Burg
London, UK

Email: gertjanvandenburg@gmail.com
Website: gertjanvandenburg.com
GitHub: github.com/GjjvdBurg

SUMMARY

– Machine learning researcher with extensive experience in algorithm design and implementation
– Skilled programmer familiar with Python, C, and R, as well as with tools such as PyTorch
– Author of publications in top peer-reviewed journals and conferences (including JMLR and NeurIPS)

PROFESSIONAL EXPERIENCE

Applied Scientist II, Amazon Alexa, UK   Dec. 2021–Present

– Developing methods to improve the accuracy of speech recognition systems

Postdoctoral Researcher, The Alan Turing Institute, UK   Feb. 2018–May 2021

– Working with Prof. Chris Williams and Prof. Charles Sutton
– Topics: probabilistic deep generative models, online matrix
  factorization, change point detection, automated data wrangling

EDUCATION

Ph.D. in Machine Learning, Erasmus University Rotterdam, NL   Jan. 2018

– Working with Prof. Patrick Groenen and Dr. Andreas Alfons
– Topics: multiclass support vector machines, hierarchical classification,
  sparse regularization
– Visiting researcher, University of Michigan, USA   Feb.–Apr. 2016
  Topic: meta-learning for hierarchical classification   & Feb. 2017
  Supervisor: Prof. Alfred Hero
– Visiting student, Stanford University, USA   May 2014
  Topic: generalized multiclass SVMs
  Supervisor: Prof. Patrick Groenen

M.Sc. in Econometrics, Erasmus University Rotterdam, NL   Aug. 2012

– Thesis topic: multiclass support vector machines
– Supervisor: Prof. Patrick Groenen

M.Sc. in Applied Physics, Delft University of Technology, NL   Oct. 2012

– Thesis topic: Monte Carlo simulations of contact processes on the GPU
– Supervisor: Prof. Gerard Barkema, Utrecht University, NL

B.Sc. in Applied Physics, Delft University of Technology, NL   Oct. 2009

- Thesis topic: image processing algorithms for quantifying protein growth
- Supervisor: Dr. Bernd Rieger

PUBLICATIONS

Peer-Reviewed Journals

- **G.J.J. van den Burg**, A. Nazábal, C. Sutton. Wrangling Messy CSV Files by Detecting Row and Type Patterns. *Data Mining and Knowledge Discovery*, 33(6):1799–1820, 2019.
- **G.J.J. van den Burg** and P.J.F. Groenen. GenSVM: A Generalized Multiclass Support Vector Machine. *Journal of Machine Learning Research*, 17(225):1–42, 2016.

Peer-Reviewed International Conferences

- **G.J.J. van den Burg** and C.K.I. Williams. On Memorization in Probabilistic Deep Generative Models. *35th Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- Ö.D. Akyildiz*, **G.J.J. van den Burg***, T. Damoulas, M.J.F. Steel. Probabilistic Sequential Matrix Factorization. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021. (* = joint first author)

ArXiv Preprints

- **G.J.J. van den Burg** and C.K.I. Williams. An Evaluation of Change Point Detection Algorithms. *arXiv:2003.06222*, 2020.
- **G.J.J. van den Burg** and A.O. Hero. Fast Meta-Learning for Adaptive Hierarchical Classifier Design. *arXiv:1711.03512*, 2017.
- **G.J.J. van den Burg**, P.J.F. Groenen, A. Alfons. SparseStep: Approximating the Counting Norm for Sparse Regularization. *arXiv:1701.06967*, 2017.

Thesis

- **G.J.J. van den Burg**. Algorithms for Multiclass Classification and Regularized Regression. Erasmus University Rotterdam, 2018.

Book chapters

- **G.J.J. van den Burg**. Reproducible Research with Make. In *The Turing Way – A Handbook for Reproducible Data Science*, available online, 2019.

AWARDS & GRANTS

- **Best Reviewer Award**, Neural Information Processing Systems Conference   2020
- **Top 33% Reviewer**, International Conference on Machine Learning   2020
- **Best Reviewer Award**, Neural Information Processing Systems Conference   2019
- **Top Educator Award**, Erasmus School of Economics   2016
- **Research Grant**, Erasmus Research Institute of Management   2016
- **Research Grant**, Erasmus Trustfonds   2016

SERVICE TO PROFESSION

Reviewing: 2019–2022

- Journals: Journal of Machine Learning Research, IEEE Transactions on Information Theory
- Conferences: Advances in Neural Information Processing Systems, International Conference on Machine Learning (*Expert Reviewer in 2021*), NewInML workshop at NeurIPS 2020, and Automated Data Science workshop at ECML-PKDD 2019

Other:

– PhD Council Member, Erasmus Research Institute in Management, NL	2014–2015
– Organizer for reading groups, Econometric Institute, NL Topics: machine learning, probabilistic inference	2013–2014

TEACHING EXPERIENCE

The Alan Turing Institute, UK

– Thesis co-supervisor for three M.Sc. students in Machine Learning Topic: automated data wrangling	2020–2021

Erasmus University Rotterdam, NL

– Lecturer and course developer for MATLAB module of Programming course First course at university to use automatically-graded exercises	2015–2016
– Thesis supervisor for two M.Sc. and four B.Sc. students in Econometrics Topics: clustering, classification, and recommender systems	2016–2017
– Teaching assistant for several M.Sc. and B.Sc. courses Topics: applied econometrics, mathematical methods, data analysis	2014–2016

Stichting Studiebegeleiding Leiden, NL

– Mathematics tutor for high-school exam students	2007–2012

OTHER TRAINING

– Machine Learning Summer School, Max Planck Institute for Intelligent Systems Tübingen, Germany (2 weeks)	2013
– R Summer School, International Association for Statistical Computing Vorau, Austria (1 week)	2013

SELECTED TALKS & CONFERENCE PRESENTATIONS

- **G.J.J. van den Burg** and C.K.I. Williams. Memorization in Probabilistic Deep Generative Models. At Google & DeepMind's *Privacy Testing Research* internal seminar series, virtual, Jul. 2021.
- **G.J.J. van den Burg**. Building useful tools for data scientists: The case of CleverCSV. At The Alan Turing Institute's *Tools, Practices, and Systems* seminar series, virtual, Jun. 2020.
- **G.J.J. van den Burg** and C.K.I. Williams. Change Point Detection. At The Alan Turing Institute's *Lunchtime Tech Talks* seminar series, London, UK, Dec. 2019.
- **G.J.J. van den Burg**. Python Packaging Tutorial. At The Alan Turing Institute's *Reproducible Research* seminar, London, UK, Nov. 2019.
- **G.J.J. van den Burg**. Hands-on Tutorial on Make. At The Alan Turing Institute's *Lunchtime Tech Talks* seminar series, London, UK, Mar. 2019.

- **G.J.J. van den Burg** and A.O. Hero. Fast Meta-Learning for Automatic Feature Transformations. In: *Artificial Intelligence for Data Analytics Workshop*, Alan Turing Institute, London, UK, Mar. 2018.
- **G.J.J. van den Burg** and A.O. Hero. Multiclass Classification and Meta-Learning with Bayes Error Estimates. In: *Conf. of the Intl. Federation of Classification Societies*, Tokyo, Japan, Aug. 2017.
- **G.J.J. van den Burg** and P.J.F. Groenen. SparseStep: Approximating the Counting Norm for Sparse Regularization. In: *22nd Intl. Conf. on Computational Statistics*, Oviedo, Spain, Aug. 2016.
- **G.J.J. van den Burg.** Benchmarking Machine Learning Methods on LISA. In: *Surfsara Super D Event*, Amsterdam, The Netherlands, Dec. 2016.
- **G.J.J. van den Burg** and P.J.F. Groenen. SparseStep: Approximating the Counting Norm for Sparse Regularization. In: *Conf. of the Intl. Federation of Classification Societies*, Bologna, Italy, Jul. 2015.
- **G.J.J. van den Burg** and P.J.F. Groenen. An Extended Comparison of Multiclass Support Vector Machines. In: *21st Intl. Conf. on Computational Statistics*, Geneva, Switzerland, Aug. 2014.
- **G.J.J. van den Burg** and P.J.F. Groenen. Flexible Multiclass Support Vector Machines. In: *6th Intl. Conf. of the ERCIM WG on Computational and Methodological Statistics*, London, UK, Dec. 2013.

ACADEMIC SOFTWARE PACKAGES

- CleverCSV: Python package for *Wrangling Messy CSV Files by Detecting Row and Type Patterns*. Received 900+ stars on GitHub and over 1,000,000 downloads.
- GenSVM: Python and R packages for *GenSVM: A Generalized Multiclass Support Vector Machine*.
- SmartSVM: Python package for *Fast Meta-Learning for Adaptive Hierarchical Classifier Design*.
- SparseStep: R package for *SparseStep: Approximating the Counting Norm for Sparse Regularization*.
- Abed: Python package for easy benchmarking of machine learning methods on a compute cluster.
- SyncRNG: Python and R package for synchronized random number generation.

TECHNICAL SKILLS

- Familiar with modern machine learning paradigms (deep learning, variational autoencoders, Gaussian processes, convolutional neural networks, etc.)
- Author of twelve Python packages and three R packages (1M+ downloads combined)
- Proficient in software engineering tools and practices (Git, Scrum, CI/CD)
- Programming languages: Python, C, R, MATLAB, Javascript
- Software: PyTorch, Scikit-Learn, TensorFlow, NumPy, Cython, Docker, Linux, Make, Bash

ADDITIONAL INFORMATION

- Languages: English (fluent), Dutch (native)
- Interests: Climbing, Running, Coding