# Masashi Sugiyama

## List of Publications by Year
### in descending order

| | | | |
|---|---|---|---|
| 139 papers | 3,967 citations | 159573 30 h-index | 149686 56 g-index |
| 141 all docs | 141 docs citations | 141 times ranked | 3492 citing authors |

| # | Article | IF | Citations |
|---|---------|-----|-----------|
| 1 | Change-point detection in time-series data by relative density-ratio estimation. Neural Networks, 2013, 43, 72-83. | 5.9 | 313 |
| 2 | Direct importance estimation for covariate shift adaptation. Annals of the Institute of Statistical Mathematics, 2008, 60, 699-746. | 0.8 | 273 |
| 3 | Semi-supervised local Fisher discriminant analysis forÂdimensionality reduction. Machine Learning, 2010, 78, 35-61. | 5.4 | 232 |
| 4 | High-Dimensional Feature Selection by Feature-Wise Kernelized Lasso. Neural Computation, 2014, 26, 185-207. | 2.2 | 222 |
| 5 | Statistical outlier detection using direct density ratio estimation. Knowledge and Information Systems, 2011, 26, 309-336. | 3.2 | 135 |
| 6 | Sequential changeâ€point detection based on direct densityâ€ratio estimation. Statistical Analysis and Data Mining, 2012, 5, 114-127. | 2.8 | 131 |
| 7 | Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 2 Applications and Future Perspectives. Foundations and Trends in Machine Learning, 2017, 9, 249-429. | 69.0 | 128 |
| 8 | Application of Covariate Shift Adaptation Techniques in Brainâ€"Computer Interfaces. IEEE Transactions on Biomedical Engineering, 2010, 57, 1318-1324. | 4.2 | 108 |
| 9 | Information-Theoretic Semi-Supervised Metric Learning via Entropy Regularization. Neural Computation, 2014, 26, 1717-1762. | 2.2 | 92 |
| 10 | The Degrees of Freedom of Partial Least Squares Regression. Journal of the American Statistical Association, 2011, 106, 697-705. | 3.1 | 74 |
| 11 | Relative Density-Ratio Estimation for Robust Distribution Comparison. Neural Computation, 2013, 25, 1324-1370. | 2.2 | 74 |
| 12 | Density-ratio matching under the Bregman divergence: a unified framework of density-ratio estimation. Annals of the Institute of Statistical Mathematics, 2012, 64, 1009-1044. | 0.8 | 72 |
| 13 | Subspace Information Criterion for Model Selection. Neural Computation, 2001, 13, 1863-1889. | 2.2 | 61 |
| 14 | Pool-based active learning in approximate linear regression. Machine Learning, 2009, 75, 249-274. | 5.4 | 60 |
| 15 | Dual-Augmented Lagrangian Method for Efficient Sparse Reconstruction. IEEE Signal Processing Letters, 2009, 16, 1067-1070. | 3.6 | 58 |
| 16 | Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition. Neurocomputing, 2012, 80, 93-101. | 5.9 | 55 |
| 17 | Class Prior Estimation from Positive and Unlabeled Data. IEICE Transactions on Information and Systems, 2014, E97.D, 1358-1362. | 0.7 | 51 |
| 18 | Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression. IEICE Transactions on Information and Systems, 2014, E97.D, 1677-1681. | 0.7 | 50 |

| # | Article | IF | Citations |
|---|---------|-----|-----------|
| 19 | Density-Difference Estimation. Neural Computation, 2013, 25, 2734-2775. | 2.2 | 49 |
| 20 | Semi-supervised learning of class balance under class-prior change by distribution matching. Neural Networks, 2014, 50, 110-119. | 5.9 | 48 |
| 21 | Dimensionality reduction for density ratio estimation in high-dimensional spaces. Neural Networks, 2010, 23, 44-59. | 5.9 | 47 |
| 22 | Statistical analysis of kernel-based least-squares density-ratio estimation. Machine Learning, 2012, 86, 335-367. | 5.4 | 47 |
| 23 | Class-prior estimation for learning from positive and unlabeled data. Machine Learning, 2017, 106, 463-492. | 5.4 | 47 |
| 24 | Sufficient Dimension Reduction via Squared-Loss Mutual Information Estimation. Neural Computation, 2013, 25, 725-758. | 2.2 | 45 |
| 25 | Semi-supervised speaker identification under covariate shift. Signal Processing, 2010, 90, 2353-2361. | 3.7 | 40 |
| 26 | Variational Bayesian sparse additive matrix factorization. Machine Learning, 2013, 92, 319-347. | 5.4 | 37 |
| 27 | Learning Efficient Tensor Representations with Ring-structured Networks. , 2019, , . | | 36 |
| 28 | A least-squares approach to anomaly detection in static and sequential data. Pattern Recognition Letters, 2014, 40, 36-40. | 4.2 | 33 |
| 29 | Direct density-ratio estimation with dimensionality reduction via least-squares hetero-distributional subspace search. Neural Networks, 2011, 24, 183-198. | 5.9 | 32 |
| 30 | Artist Agent: A Reinforcement Learning Approach to Automatic Stroke Generation in Oriental Ink Painting. IEICE Transactions on Information and Systems, 2013, E96.D, 1134-1144. | 0.7 | 32 |
| 31 | Direct Divergence Approximation between Probability Distributions and Its Applications in Machine Learning. Journal of Computing Science and Engineering, 2013, 7, 99-111. | 0.6 | 32 |
| 32 | Analysis and improvement of policy gradient estimation. Neural Networks, 2012, 26, 118-129. | 5.9 | 31 |
| 33 | Machine Learning with Squared-Loss Mutual Information. Entropy, 2013, 15, 80-112. | 2.2 | 31 |
| 34 | Least-Squares Independent Component Analysis. Neural Computation, 2011, 23, 284-301. | 2.2 | 29 |
| 35 | Health Checkup and Telemedical Intervention Program for Preventive Medicine in Developing Countries: Verification Study. Journal of Medical Internet Research, 2015, 17, e2. | 4.3 | 29 |
| 36 | Single-shot surface profiling by local model fitting. Applied Optics, 2006, 45, 7999. | 2.1 | 26 |

| # | Article | IF | Citations |
|---|---------|----|-----------|
| 37 | Averaging covariance matrices for EEG signal classification based on the CSP: An empirical study. , 2015, , . | | 26 |
| 38 | $f$-Divergence Estimation and Two-Sample Homogeneity Test Under Semiparametric Density-Ratio Models. IEEE Transactions on Information Theory, 2012, 58, 708-720. | 2.4 | 25 |
| 39 | Information-Maximization Clustering Based on Squared-Loss Mutual Information. Neural Computation, 2014, 26, 84-131. | 2.2 | 25 |
| 40 | Semi-supervised AUC optimization based on positive-unlabeled learning. Machine Learning, 2018, 107, 767-794. | 5.4 | 25 |
| 41 | Geodesic Gaussian kernels for value function approximation. Autonomous Robots, 2008, 25, 287-304. | 4.8 | 23 |
| 42 | Canonical dependency analysis based on squared-loss mutual information. Neural Networks, 2012, 34, 46-55. | 5.9 | 23 |
| 43 | Geometry-aware principal component analysis for symmetric positive definite matrices. Machine Learning, 2017, 106, 493-522. | 5.4 | 23 |
| 44 | Artificial Neural Variability for Deep Learning: On Overfitting, Noise Memorization, and Catastrophic Forgetting. Neural Computation, 2021, 33, 2163-2192. | 2.2 | 23 |
| 45 | Trading Variance Reduction with Unbiasedness: The Regularized Subspace Information Criterion for Robust Model Selection in Kernel Regression. Neural Computation, 2004, 16, 1077-1104. | 2.2 | 22 |
| 46 | Theoretical and Experimental Analyses of Tensor-Based Regression and Classification. Neural Computation, 2016, 28, 686-715. | 2.2 | 21 |
| 47 | Incremental Active Learning for Optimal Generalization. Neural Computation, 2000, 12, 2909-2940. | 2.2 | 20 |
| 48 | Multi-parametric solution-path algorithm for instance-weighted support vector machines. Machine Learning, 2012, 88, 297-330. | 5.4 | 20 |
| 49 | Efficient Sample Reuse in Policy Gradients with Parameter-Based Exploration. Neural Computation, 2013, 25, 1512-1547. | 2.2 | 20 |
| 50 | Active Feature Acquisition with Supervised Matrix Completion. , 2018, , . | | 20 |
| 51 | A new algorithm of non-Gaussian component analysis with radial kernel functions. Annals of the Institute of Statistical Mathematics, 2007, 59, 57-75. | 0.8 | 19 |
| 52 | A batch ensemble approach to active learning with model selection. Neural Networks, 2008, 21, 1278-1286. | 5.9 | 19 |
| 53 | Least-squares two-sample test. Neural Networks, 2011, 24, 735-751. | 5.9 | 17 |
| 54 | Reward-Weighted Regression with Sample Reuse for Direct Policy Search in Reinforcement Learning. Neural Computation, 2011, 23, 2798-2832. | 2.2 | 17 |

| # | Article | IF | Citations |
|---|---------|-----|-----------|
| 55 | Target-less camera-LiDAR extrinsic calibration using a bagged dependence estimator. , 2016, , . | | 17 |
| 56 | Using the variogram for vector outlier screening: application to feature-based image registration. International Journal of Computer Assisted Radiology and Surgery, 2018, 13, 1871-1880. | 2.8 | 17 |
| 57 | Direct Learning of Sparse Changes in Markov Networks by Density Ratio Estimation. Neural Computation, 2014, 26, 1169-1197. | 2.2 | 16 |
| 58 | A unified method for optimizing linear image restoration filters. Signal Processing, 2002, 82, 1773-1787. | 3.7 | 15 |
| 59 | Learning under nonstationarity: covariate shift and class‑balance change. Wiley Interdisciplinary Reviews: Computational Statistics, 2013, 5, 465-477. | 3.9 | 15 |
| 60 | Statistical outlier detection for diagnosis of cyber attacks in power state estimation. , 2016, , . | | 15 |
| 61 | Spatiotemporal dynamics of odor representations in the human brain revealed by EEG decoding. Proceedings of the National Academy of Sciences of the United States of America, 2022, 119, e2114966119. | 7.1 | 15 |
| 62 | Bias Reduction and Metric Learning for Nearest-Neighbor Estimation of Kullback-Leibler Divergence. Neural Computation, 2018, 30, 1930-1960. | 2.2 | 14 |
| 63 | Model-based policy gradients with parameter-based exploration by least-squares conditional density estimation. Neural Networks, 2014, 57, 128-140. | 5.9 | 13 |
| 64 | Bandit-Based Task Assignment for Heterogeneous Crowdsourcing. Neural Computation, 2015, 27, 2447-2475. | 2.2 | 13 |
| 65 | Semi-supervised information-maximization clustering. Neural Networks, 2014, 57, 103-111. | 5.9 | 12 |
| 66 | Convex formulation of multiple instance learning from positive and unlabeled bags. Neural Networks, 2018, 105, 132-141. | 5.9 | 12 |
| 67 | Theoretical and Experimental Evaluation of the Subspace Information Criterion. Machine Learning, 2002, 48, 25-50. | 5.4 | 11 |
| 68 | Computational complexity of kernel-based density-ratio estimation: a condition number analysis. Machine Learning, 2013, 90, 431-460. | 5.4 | 11 |
| 69 | Feature Selection via $l_1$-Penalized Squared-Loss Mutual Information. IEICE Transactions on Information and Systems, 2013, E96.D, 1513-1524. | 0.7 | 11 |
| 70 | Computationally Efficient Estimation of Squared-Loss Mutual Information with Multiplicative Kernel Models. IEICE Transactions on Information and Systems, 2014, E97.D, 968-971. | 0.7 | 11 |
| 71 | Computationally Efficient Class-Prior Estimation under Class Balance Change Using Energy Distance. IEICE Transactions on Information and Systems, 2016, E99.D, 176-186. | 0.7 | 11 |
| 72 | Active deep Q-learning with demonstration. Machine Learning, 2020, 109, 1699-1725. | 5.4 | 11 |

| # | Article | IF | Citations |
|---|---------|----|-----------|
| 73 | A Feature-Driven Active Framework for Ultrasound-Based Brain Shift Compensation. Lecture Notes in Computer Science, 2018, , 30-38. | 1.3 | 11 |
| 74 | SemiCCA: Efficient Semi-supervised Learning of Canonical Correlations. IPSJ Online Transactions, 2013, 6, 37-44. | 0.1 | 10 |
| 75 | Importance-weighted covariance estimation for robust common spatial pattern. Pattern Recognition Letters, 2015, 68, 139-145. | 4.2 | 10 |
| 76 | Model-based reinforcement learning with dimension reduction. Neural Networks, 2016, 84, 1-16. | 5.9 | 10 |
| 77 | ECG-Based Concentration Recognition With Multi-Task Regression. IEEE Transactions on Biomedical Engineering, 2019, 66, 101-110. | 4.2 | 10 |
| 78 | Classification From Pairwise Similarities/Dissimilarities and Unlabeled Data via Empirical Risk Minimization. Neural Computation, 2021, 33, 1234-1268. | 2.2 | 10 |
| 79 | Theory and Algorithm for Learning with Dissimilarity Functions. Neural Computation, 2009, 21, 1459-1484. | 2.2 | 9 |
| 80 | On Kernel Parameter Selection in Hilbert-Schmidt Independence Criterion. IEICE Transactions on Information and Systems, 2012, E95.D, 2564-2567. | 0.7 | 9 |
| 81 | Modal Regression via Direct Log-Density Derivative Estimation. Lecture Notes in Computer Science, 2016, , 108-116. | 1.3 | 9 |
| 82 | Support consistency of direct sparse-change learning in Markov networks. Annals of Statistics, 2017, 45, . | 2.6 | 9 |
| 83 | Classification from Triplet Comparison Data. Neural Computation, 2020, 32, 659-681. | 2.2 | 9 |
| 84 | On the Applicability of Registration Uncertainty. Lecture Notes in Computer Science, 2019, , 410-419. | 1.3 | 9 |
| 85 | Direct Approximation of Quadratic Mutual Information and Its Application to Dependence-Maximization Clustering. IEICE Transactions on Information and Systems, 2013, E96.D, 2282-2285. | 0.7 | 8 |
| 86 | Direct Density Ratio Estimation with Convolutional Neural Networks with Application in Outlier Detection. IEICE Transactions on Information and Systems, 2015, E98.D, 1073-1079. | 0.7 | 8 |
| 87 | Cross-Domain Matching with Squared-Loss Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37, 1764-1776. | 13.9 | 8 |
| 88 | Conditional Density Estimation with Dimensionality Reduction via Squared-Loss Conditional Entropy Minimization. Neural Computation, 2015, 27, 228-254. | 2.2 | 8 |
| 89 | Least-squares independence regression for non-linear causal inference under non-Gaussian noise. Machine Learning, 2014, 96, 249-267. | 5.4 | 7 |
| 90 | Direct Estimation of the Derivative of Quadratic Mutual Information with Application in Supervised Dimension Reduction. Neural Computation, 2017, 29, 2076-2122. | 2.2 | 7 |

| # | Article | IF | Citations |
|---|---------|----|-----------|
| 91 | Homotopy continuation approaches for robust SV classification and regression. Machine Learning, 2017, 106, 1009-1038. | 5.4 | 7 |
| 92 | Boosting and margin theory. Frontiers of Electrical and Electronic Engineering, 2012, 7, 127-133. | 0.5 | 6 |
| 93 | Clustering Unclustered Data: Unsupervised Binary Labeling of Two Datasets Having Different Class Balances. , 2013, , . | | 6 |
| 94 | Direct Density Derivative Estimation. Neural Computation, 2016, 28, 1101-1140. | 2.2 | 6 |
| 95 | Good arm identification via bandit feedback. Machine Learning, 2019, 108, 721-745. | 5.4 | 6 |
| 96 | Principled analytic classifier for positive-unlabeled learning via weighted integral probability metric. Machine Learning, 2020, 109, 513-532. | 5.4 | 6 |
| 97 | Constraint learning for control tasks with limited duration barrier functions. Automatica, 2021, 127, 109504. | 5.0 | 6 |
| 98 | Constrained Least-Squares Density-Difference Estimation. IEICE Transactions on Information and Systems, 2014, E97.D, 1822-1829. | 0.7 | 5 |
| 99 | Trial and Error: Using Previous Experiences as Simulation Models in Humanoid Motor Learning. IEEE Robotics and Automation Magazine, 2016, 23, 96-105. | 2.0 | 5 |
| 100 | Binary classification with ambiguous training data. Machine Learning, 2020, 109, 2369-2388. | 5.4 | 5 |
| 101 | Predictive Approaches for Low-Cost Preventive Medicine Program in Developing Countries. , 2015, , . | | 5 |
| 102 | A Multipurpose Linear Component Analysis Method Based on Modulated Hebb-Oja Learning Rule. IEEE Signal Processing Letters, 2008, 15, 677-680. | 3.6 | 4 |
| 103 | Winning the Kaggle Algorithmic Trading Challenge with the Composition of Many Models and Feature Engineering. IEICE Transactions on Information and Systems, 2013, E96.D, 742-745. | 0.7 | 4 |
| 104 | Statistical Analysis of Distance Estimators with Density Differences and Density Ratios. Entropy, 2014, 16, 921-942. | 2.2 | 4 |
| 105 | Direct conditional probability density estimation with sparse feature selection. Machine Learning, 2015, 100, 161-182. | 5.4 | 4 |
| 106 | Regularized Multitask Learning for Multidimensional Log-Density Gradient Estimation. Neural Computation, 2016, 28, 1388-1410. | 2.2 | 4 |
| 107 | On Generalization Performance and Non-Convex Optimization of Extended $\nu$-Support Vector Machine. New Generation Computing, 2009, 27, 259-279. | 3.3 | 3 |
| 108 | Noise adaptive optimization of matrix initialization for frequency-domain independent component analysis. , 2013, 23, 1-8. | | 3 |

| # | Article | IF | Citations |
|---|---------|-----|-----------|
| 109 | Lung lesion detection in FDG-PET/CT with Gaussian process regression. , 2017, , . | | 3 |
| 110 | Stochastic Multichannel Ranking with Brain Dynamics Preferences. Neural Computation, 2020, 32, 1499-1530. | 2.2 | 3 |
| 111 | Semisupervised Ordinal Regression Based on Empirical Risk Minimization. Neural Computation, 2021, 33, 3361-3412. | 2.2 | 3 |
| 112 | Multiwavelength-integrated local model fitting method for interferometric surface profiling. Applied Optics, 2012, 51, 6700. | 1.8 | 2 |
| 113 | Improved algorithm for multiwavelength single-shot interferometric surface profiling: speeding up the multiwavelength-integrated local model fitting method by local information sharing. Applied Optics, 2013, 52, 4042. | 1.8 | 2 |
| 114 | Salient Object Detection Based on Direct Density-ratio Estimation. IPSJ Online Transactions, 2013, 6, 96-103. | 0.1 | 2 |
| 115 | Unsupervised Dimension Reduction via Least-Squares Quadratic Mutual Information. IEICE Transactions on Information and Systems, 2014, E97.D, 2806-2809. | 0.7 | 2 |
| 116 | A fault detection technique for the steel manufacturing process based on a normal pattern library. IFAC-PapersOnLine, 2015, 48, 871-876. | 0.9 | 2 |
| 117 | Registration of infrared transmission images using squared-loss mutual information. Precision Engineering, 2015, 39, 187-193. | 3.4 | 2 |
| 118 | Dependence maximization localization: a novel approach to 2D street-map-based robot localization. Advanced Robotics, 2016, 30, 1431-1445. | 1.8 | 2 |
| 119 | An Online Policy Gradient Algorithm for Markov Decision Processes with Continuous States and Actions. Neural Computation, 2016, 28, 563-593. | 2.2 | 2 |
| 120 | Clipped Matrix Completion: A Remedy for Ceiling Effects. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33, 5151-5158. | 4.9 | 2 |
| 121 | Early Stopping Heuristics in Pool-Based Incremental Active Learning for Least-Squares Probabilistic Classifier. IEICE Transactions on Information and Systems, 2012, E95.D, 2065-2073. | 0.7 | 1 |
| 122 | Improving importance estimation in pool-based batch active learning for approximate linear regression. Neural Networks, 2012, 36, 73-82. | 5.9 | 1 |
| 123 | Designing Various Multivariate Analysis at Will via Generalized Pairwise Expression. IPSJ Online Transactions, 2013, 6, 45-54. | 0.1 | 1 |
| 124 | Sufficient Dimension Reduction via Direct Estimation of the Gradients of Logarithmic Conditional Densities. Neural Computation, 2018, 30, 477-504. | 2.2 | 1 |
| 125 | Millionaire: a hint-guided approach for crowdsourcing. Machine Learning, 2019, 108, 831-858. | 5.4 | 1 |
| 126 | Information-Theoretic Representation Learning for Positive-Unlabeled Classification. Neural Computation, 2021, 33, 244-268. | 2.2 | 1 |

| # | Article | IF | Citations |
|---|---------|----|-----------|
| 127 | Direction Matters: On Influence-Preserving Graph Summarization and Max-Cut Principle for Directed Graphs. Neural Computation, 2021, 33, 2128-2162. | 2.2 | 1 |
| 128 | Positive-Unlabeled Classification under Class Prior Shift and Asymmetric Error. , 2019, , 271-279. | | 1 |
| 129 | Introduction to the Special Section on the 2nd Asia Conference on Machine Learning (ACML 2010). ACM Transactions on Intelligent Systems and Technology, 2012, 3, 1-1. | 4.5 | 0 |
| 130 | Multi-Task Approach to Reinforcement Learning for Factored-State Markov Decision Problems. IEICE Transactions on Information and Systems, 2012, E95.D, 2426-2437. | 0.7 | 0 |
| 131 | Computationally Efficient Multi-Label Classification by Least-Squares Probabilistic Classifiers. IEICE Transactions on Information and Systems, 2013, E96.D, 1871-1874. | 0.7 | 0 |
| 132 | Introduction: special issue of selected papers of ACML 2013. Machine Learning, 2015, 99, 165-167. | 5.4 | 0 |
| 133 | Online Direct Density-Ratio Estimation Applied to Inlier-Based Outlier Detection. Neural Computation, 2015, 27, 1899-1914. | 2.2 | 0 |
| 134 | Foreword: special issue for the journal track of the 8th Asian conference on machine learning (ACML) Tj ETQq0 0 0 rgBT /Overlock 10 Tf | 5.4 | 0 |
| 135 | Introduction: special issue of selected papers from ACML 2015. Machine Learning, 2017, 106, 459-461. | 5.4 | 0 |
| 136 | Stochastic Divergence Minimization for Biterm Topic Models. IEICE Transactions on Information and Systems, 2018, E101.D, 668-677. | 0.7 | 0 |
| 137 | Foreword: special issue for the journal track of the 10th Asian Conference on Machine Learning (ACML 2018). Machine Learning, 2019, 108, 717-719. | 5.4 | 0 |
| 138 | Unsupervised key frame selection using information theory and colour histogram difference. International Journal of Business Intelligence and Data Mining, 2020, 16, 324. | 0.2 | 0 |
| 139 | Polynomial-Time Algorithms for Multiple-Arm Identification with Full-Bandit Feedback. Neural Computation, 2020, 32, 1733-1773. | 2.2 | 0 |