

# Michael I Jordan

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/1885839/publications.pdf>

Version: 2024-02-01

182  
papers

32,724  
citations

26626

56  
h-index

14208

128  
g-index

196  
all docs

196  
docs citations

196  
times ranked

27110  
citing authors

#	ARTICLE	IF	CITATIONS
1	Machine learning: Trends, perspectives, and prospects. <i>Science</i> , 2015, 349, 255-260.	12.6	4,904
2	Hierarchical Dirichlet Processes. <i>Journal of the American Statistical Association</i> , 2006, 101, 1566-1581.	3.1	2,215
3	An Introduction to Variational Methods for Graphical Models. <i>Machine Learning</i> , 1999, 37, 183-233.	5.4	1,889
4	An Introduction to MCMC for Machine Learning. <i>Machine Learning</i> , 2003, 50, 5-43.	5.4	1,641
5	Deep generative modeling for single-cell transcriptomics. <i>Nature Methods</i> , 2018, 15, 1053-1058.	19.0	1,227
6	Graphical Models, Exponential Families, and Variational Inference. <i>Foundations and Trends in Machine Learning</i> , 2007, 1, 1-305.	69.0	1,196
7	Convex and Semi-Nonnegative Matrix Factorizations. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2010, 32, 45-55.	13.9	951
8	Variational inference for Dirichlet process mixtures. <i>Bayesian Analysis</i> , 2006, 1, 121.	3.0	939
9	Multiple kernel learning, conic duality, and the SMO algorithm. , 2004, , .		872
10	Detecting large-scale system problems by mining console logs. , 2009, , .		737
11	Factorial Hidden Markov Models. <i>Machine Learning</i> , 1997, 29, 245-273.	5.4	674
12	Modeling annotated data. , 2003, , .		597
13	A statistical framework for genomic data fusion. <i>Bioinformatics</i> , 2004, 20, 2626-2635.	4.1	568
14	Convexity, Classification, and Risk Bounds. <i>Journal of the American Statistical Association</i> , 2006, 101, 138-156.	3.1	529
15	A Direct Formulation for Sparse PCA Using Semidefinite Programming. <i>SIAM Review</i> , 2007, 49, 434-448.	8.4	524
16	Chemogenomic profiling: Identifying the functional interactions of small molecules in yeast. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2004, 101, 793-798.	7.1	460
17	The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. <i>Journal of the ACM</i> , 2010, 57, 1-30.	2.2	390
18	Bayesian parameter estimation via variational methods. <i>Statistics and Computing</i> , 2000, 10, 25-37.	1.5	383

#	ARTICLE	IF	CITATIONS
19	Graphical Models. <i>Statistical Science</i> , 2004, 19, 140.	2.8	375
20	Transferable Representation Learning with Deep Adaptation Networks. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2019, 41, 3071-3085.	13.9	345
21	Joint covariate selection and joint subspace selection for multiple classification problems. <i>Statistics and Computing</i> , 2010, 20, 231-252.	1.5	325
22	Local Privacy and Statistical Minimax Rates. , 2013, , .		310
23	Managing data transfers in computer clusters with orchestra. , 2011, , .		301
24	Fast approximate spectral clustering. , 2009, , .		295
25	Learning Dependency-Based Compositional Semantics. <i>Computational Linguistics</i> , 2013, 39, 389-446.	3.3	268
26	Estimating Divergence Functionals and the Likelihood Ratio by Convex Risk Minimization. <i>IEEE Transactions on Information Theory</i> , 2010, 56, 5847-5861.	2.4	264
27	Partial Transfer Learning with Selective Adversarial Networks. , 2018, , .		249
28	A Scalable Bootstrap for Massive Data. <i>Journal of the Royal Statistical Society Series B: Statistical Methodology</i> , 2014, 76, 795-816.	2.2	237
29	Scalable statistical bug isolation. <i>ACM SIGPLAN Notices</i> , 2005, 40, 15-26.	0.2	234
30	HopSkipJumpAttack: A Query-Efficient Decision-Based Attack. , 2020, , .		229
31	A sticky HDP-HMM with application to speaker diarization. <i>Annals of Applied Statistics</i> , 2011, 5, .	1.1	225
32	Support union recovery in high-dimensional multivariate regression. <i>Annals of Statistics</i> , 2011, 39, .	2.6	217
33	Universal Domain Adaptation. , 2019, , .		216
34	A Python library for probabilistic analysis of single-cell omics data. <i>Nature Biotechnology</i> , 2022, 40, 163-166.	17.5	216
35	A critical assessment of <i>Mus musculus</i> gene function prediction using integrated genomic evidence. <i>Genome Biology</i> , 2008, 9, S2.	9.6	214
36	Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. <i>Molecular Systems Biology</i> , 2021, 17, e9620.	7.2	211

#	ARTICLE	IF	CITATIONS
37	A variational perspective on accelerated methods in optimization. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113, E7351-E7358.	7.1	193
38	Predicting Multiple Metrics for Queries: Better Decisions Enabled by Machine Learning. Proceedings - International Conference on Data Engineering, 2009, , .	0.0	189
39	Communication-Efficient Distributed Statistical Inference. Journal of the American Statistical Association, 2019, 114, 668-681.	3.1	189
40	Probabilistic Independence Networks for Hidden Markov Probability Models. Neural Computation, 1997, 9, 227-269.	2.2	168
41	Optimal Rates for Zero-Order Convex Optimization: The Power of Two Function Evaluations. IEEE Transactions on Information Theory, 2015, 61, 2788-2806.	2.4	168
42	Minimax Optimal Procedures for Locally Private Estimation. Journal of the American Statistical Association, 2018, 113, 182-201.	3.1	167
43	Protein Molecular Function Prediction by Bayesian Phylogenomics. PLoS Computational Biology, 2005, 1, e45.	3.2	162
44	Kernel dimension reduction in regression. Annals of Statistics, 2009, 37, .	2.6	162
45	Bayesian Nonparametric Inference of Switching Dynamic Linear Models. IEEE Transactions on Signal Processing, 2011, 59, 1569-1585.	5.3	162
46	Genomic privacy and limits of individual detection in a pool. Nature Genetics, 2009, 41, 965-967.	21.4	153
47	Learning semantic correspondences with less supervision. , 2009, , .		152
48	Hierarchical Bayesian nonparametric models with applications. , 2010, , 158-207.		149
49	A kernel-based learning approach to ad hoc sensor network localization. ACM Transactions on Sensor Networks, 2005, 1, 134-152.	3.6	145
50	Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. PLoS Computational Biology, 2010, 6, e1000763.	3.2	145
51	Characterizing, modeling, and generating workload spikes for stateful services. , 2010, , .		145
52	Genome-Wide Requirements for Resistance to Functionally Distinct DNA-Damaging Agents. PLoS Genetics, 2005, 1, e24.	3.5	144
53	Computational and statistical tradeoffs via convex relaxation. Proceedings of the National Academy of Sciences of the United States of America, 2013, 110, E1181-90.	7.1	143
54	Bug isolation via remote program sampling. ACM SIGPLAN Notices, 2003, 38, 141-154.	0.2	138

#	ARTICLE	IF	CITATIONS
55	An HDP-HMM for systems with state persistence. , 2008, , .		137
56	Solving Consensus and Semi-supervised Clustering Problems Using Nonnegative Matrix Factorization. , 2007, , .		136
57	Online System Problem Detection by Mining Patterns of Console Logs. , 2009, , .		134
58	Nested Hierarchical Dirichlet Processes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37, 256-270.	13.9	111
59	Distributed optimization with arbitrary local solvers. Optimization Methods and Software, 2017, 32, 813-848.	2.4	111
60	Scaling up crowd-sourcing to very large datasets. Proceedings of the VLDB Endowment, 2014, 8, 125-136.	3.8	109
61	Knowing when you're wrong. , 2014, , .		106
62	MLI: An API for Distributed Machine Learning. , 2013, , .		105
63	Statistical debugging. , 2006, , .		103
64	Mixed Memory Markov Models: Decomposing Complex Stochastic Processes as Mixtures of Simpler Ones. Machine Learning, 1999, 37, 75-87.	5.4	102
65	Sulfur and Nitrogen Limitation in Escherichia coli K-12: Specific Homeostatic Responses. Journal of Bacteriology, 2005, 187, 1074-1090.	2.2	96
66	Privacy Aware Learning. Journal of the ACM, 2014, 61, 1-57.	2.2	96
67	Automating model search for large scale machine learning. , 2015, , .		91
68	Nonparametric empirical Bayes for the Dirichlet process mixture model. Statistics and Computing, 2006, 16, 5-14.	1.5	78
69	DestVI identifies continuums of cell types in spatial transcriptomics data. Nature Biotechnology, 2022, 40, 1360-1369.	17.5	75
70	Nonnegative Matrix Factorization for Combinatorial Optimization: Spectral Clustering, Graph Matching, and Clique Finding. , 2008, , .		73
71	Matrix concentration inequalities via the method of exchangeable pairs. Annals of Probability, 2014, 42, .	1.8	71
72	Consistent probabilistic outputs for protein function prediction. Genome Biology, 2008, 9, S6.	9.6	68

#	ARTICLE	IF	CITATIONS
73	On the computational complexity of high-dimensional Bayesian variable selection. <i>Annals of Statistics</i> , 2016, 44, .	2.6	68
74	Active site prediction using evolutionary and structural information. <i>Bioinformatics</i> , 2010, 26, 617-624.	4.1	67
75	On statistics, computation and scalability. <i>Bernoulli</i> , 2013, 19, .	1.3	65
76	Phylogenetic Inference via Sequential Monte Carlo. <i>Systematic Biology</i> , 2012, 61, 579-593.	5.6	64
77	Joint modeling of multiple time series via the beta process with application to motion capture segmentation. <i>Annals of Applied Statistics</i> , 2014, 8, .	1.1	62
78	First-order methods almost always avoid strict saddle points. <i>Mathematical Programming</i> , 2019, 176, 311-337.	2.4	61
79	Regression on manifolds using kernel dimension reduction. , 2007, , .		59
80	Bayesian semiparametric Wiener system identification. <i>Automatica</i> , 2013, 49, 2053-2063.	5.0	55
81	Genome-scale phylogenetic function annotation of large and diverse protein families. <i>Genome Research</i> , 2011, 21, 1969-1980.	5.5	54
82	A latent variable model for chemogenomic profiling. <i>Bioinformatics</i> , 2005, 21, 3286-3293.	4.1	53
83	Nonparametric decentralized detection using kernel methods. <i>IEEE Transactions on Signal Processing</i> , 2005, 53, 4053-4066.	5.3	53
84	On surrogate loss functions and f-divergences. <i>Annals of Statistics</i> , 2009, 37, .	2.6	50
85	Perturbed Iterate Analysis for Asynchronous Stochastic Optimization. <i>SIAM Journal on Optimization</i> , 2017, 27, 2202-2229.	2.0	50
86	Sampling can be faster than optimization. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2019, 116, 20881-20885.	7.1	50
87	A Randomization Test for Controlling Population Stratification in Whole-Genome Association Studies. <i>American Journal of Human Genetics</i> , 2007, 81, 895-905.	6.2	48
88	Iterative Discovery of Multiple Alternative Clustering Views. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2014, 36, 1340-1353.	13.9	47
89	Multiple-sequence functional annotation and the generalized hidden Markov phylogeny. <i>Bioinformatics</i> , 2004, 20, 1850-1860.	4.1	44
90	Learning Multiscale Representations of Natural Scenes Using Dirichlet Processes. , 2007, , .		44

#	ARTICLE	IF	CITATIONS
91	On the inference of ancestries in admixed populations. <i>Genome Research</i> , 2008, 18, 668-675.	5.5	44
92	An asymptotic analysis of generative, discriminative, and pseudolikelihood estimators. , 2008, , .		43
93	Bayesian Nonparametric Methods for Learning Markov Switching Processes. <i>IEEE Signal Processing Magazine</i> , 2010, , .	5.6	43
94	Log-determinant relaxation for approximate inference in discrete Markov random fields. <i>IEEE Transactions on Signal Processing</i> , 2006, 54, 2099-2109.	5.3	40
95	Beta Processes, Stick-Breaking and Power Laws. <i>Bayesian Analysis</i> , 2012, 7, .	3.0	40
96	SM a SH: a benchmarking toolkit for human genome variant calling. <i>Bioinformatics</i> , 2014, 30, 2787-2795.	4.1	40
97	Lessons from <i>Escherichia coli</i> genes similarly regulated in response to nitrogen and sulfur limitation. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2005, 102, 3453-3458.	7.1	39
98	A Dual Receptor Crosstalk Model of G-Protein-Coupled Signal Transduction. <i>PLoS Computational Biology</i> , 2008, 4, e1000185.	3.2	38
99	Ergodic Mirror Descent. <i>SIAM Journal on Optimization</i> , 2012, 22, 1549-1578.	2.0	37
100	Cluster Forests. <i>Computational Statistics and Data Analysis</i> , 2013, 66, 178-192.	1.2	37
101	Learning from measurements in exponential families. , 2009, , .		36
102	Evolutionary inference via the Poisson Indel Process. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2013, 110, 1160-1166.	7.1	34
103	Understanding the acceleration phenomenon via high-resolution differential equations. <i>Mathematical Programming</i> , 2022, 195, 79-148.	2.4	34
104	Bayesian Haplotype Inference via the Dirichlet Process. <i>Journal of Computational Biology</i> , 2007, 14, 267-284.	1.6	32
105	Bayesian inference for queueing networks and modeling of internet services. <i>Annals of Applied Statistics</i> , 2011, 5, .	1.1	32
106	Real-Time Machine Learning. , 2017, , .		31
107	Distributed Low-Rank Subspace Segmentation. , 2013, , .		30
108	Feature Allocations, Probability Functions, and Paintboxes. <i>Bayesian Analysis</i> , 2013, 8, .	3.0	30

#	ARTICLE	IF	CITATIONS
109	A unified treatment of multiple testing with prior knowledge using the p-filter. <i>Annals of Statistics</i> , 2019, 47, .	2.6	30
110	Multiway Spectral Clustering: A Margin-Based Perspective. <i>Statistical Science</i> , 2008, 23, .	2.8	29
111	Automatic exploration of datacenter performance regimes. , 2009, , .		29
112	Active spectral clustering via iterative uncertainty reduction. , 2012, , .		29
113	Optimal prediction for sparse linear models? Lower bounds for coordinate-separable M-estimators. <i>Electronic Journal of Statistics</i> , 2017, 11, .	0.7	29
114	Word alignment via quadratic assignment. , 2006, , .		29
115	Cluster and Feature Modeling from Combinatorial Stochastic Processes. <i>Statistical Science</i> , 2013, 28, .	2.8	25
116	Combinatorial Clustering and the Beta Negative Binomial Process. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> , 2015, 37, 290-306.	13.9	23
117	Learning Low-Dimensional Signal Models. <i>IEEE Signal Processing Magazine</i> , 2011, 28, 39-51.	5.6	22
118	Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data. <i>Bioinformatics</i> , 2009, 25, i231-i239.	4.1	20
119	Subtree power analysis and species selection for comparative genomics. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2005, 102, 7900-7905.	7.1	19
120	Extensions of the Informative Vector Machine. <i>Lecture Notes in Computer Science</i> , 2005, , 56-87.	1.3	19
121	Bayesian multi-population haplotype inference via a hierarchical dirichlet process mixture. , 2006, , .		18
122	Nonparametric link prediction in large scale dynamic networks. <i>Electronic Journal of Statistics</i> , 2014, 8, .	0.7	18
123	Union support recovery in high-dimensional multivariate regression. , 2008, , .		17
124	Nonparametric estimation of the likelihood ratio and divergence functionals. , 2007, , .		15
125	Feature space resampling for protein conformational search. <i>Proteins: Structure, Function and Bioinformatics</i> , 2010, 78, 1583-1593.	2.6	15
126	Sufficient dimension reduction for visual sequence classification. , 2010, , .		15



#	ARTICLE	IF	CITATIONS
127	Optimism-driven exploration for nonlinear systems. , 2015, , .		15
128	Posteriors, conjugacy, and exponential families for completely random measures. Bernoulli, 2018, 24, .	1.3	14
129	A sequential algorithm for false discovery rate control on directed acyclic graphs. Biometrika, 2019, 106, 69-86.	2.4	14
130	Is there an analog of Nesterov acceleration for gradient-based MCMC?. Bernoulli, 2021, 27, .	1.3	13
131	Phylogenetic molecular function annotation. Journal of Physics: Conference Series, 2009, 180, 012024.	0.4	12
132	Major Advances and Emerging Developments of Graphical Models [From the Guest Editors]. IEEE Signal Processing Magazine, 2010, 27, 17-138.	5.6	12
133	A general bootstrap performance diagnostic. , 2013, , .		12
134	A Flexible and Efficient Algorithm for Regularized Fisher Discriminant Analysis. Lecture Notes in Computer Science, 2009, , 632-647.	1.3	12
135	Association Mapping and Significance Estimation via the Coalescent. American Journal of Human Genetics, 2008, 83, 675-683.	6.2	11
136	Local privacy and statistical minimax rates. , 2013, , .		11
137	A Marked Poisson Process Driven Latent Shape Model for 3D Segmentation of Reflectance Confocal Microscopy Image Stacks of Human Skin. IEEE Transactions on Image Processing, 2017, 26, 172-184.	9.8	11
138	Generalized Momentum-Based Methods: A Hamiltonian Perspective. SIAM Journal on Optimization, 2021, 31, 915-944.	2.0	11
139	On dissipative symplectic integration with applications to gradient-based optimization. Journal of Statistical Mechanics: Theory and Experiment, 2021, 2021, 043402.	2.3	11
140	Image Denoising with Nonparametric Hidden Markov Trees. , 2007, , .		10
141	Supervised hierarchical Pitman-Yor process for natural scene segmentation. , 2011, , .		10
142	The asymptotics of ranking algorithms. Annals of Statistics, 2013, 41, .	2.6	9
143	Machine Learning and Databases. , 2015, , .		9
144	A control-theoretic perspective on optimal high-order optimization. Mathematical Programming, 2022, 195, 929-975.	2.4	9

#	ARTICLE	IF	CITATIONS
145	SOUL: An Energy-Efficient Unsupervised Online Learning Seizure Detection Classifier. IEEE Journal of Solid-State Circuits, 2022, 57, 2532-2544.	5.4	9
146	On optimal quantization rules for sequential decision problems. , 2006, , .		8
147	A semiparametric Bayesian approach to Wiener system identification*. IFAC Postprint Volumes IPPV / International Federation of Automatic Control, 2012, 45, 1137-1142.	0.4	8
148	Decoding From Pooled Data: Phase Transitions of Message Passing. IEEE Transactions on Information Theory, 2019, 65, 572-585.	2.4	8
149	On kernel methods for covariates that are rankings. Electronic Journal of Statistics, 2018, 12, .	0.7	8
150	A graphical model for predicting protein molecular function. , 2006, , .		7
151	On Optimal Quantization Rules for Some Problems in Sequential Decentralized Detection. IEEE Transactions on Information Theory, 2008, 54, 3285-3295.	2.4	7
152	Is Temporal Difference Learning Optimal? An Instance-Dependent Analysis. SIAM Journal on Mathematics of Data Science, 2021, 3, 1013-1040.	1.8	7
153	Qualcomm Context-Awareness Symposium Sets Research Agenda for Context-Aware Smartphones. IEEE Pervasive Computing, 2012, 11, 76-79.	1.3	6
154	Molecular function prediction for a family exhibiting evolutionary tendencies toward substrate specificity swapping: Recurrence of tyrosine aminotransferase activity in the Î± subfamily. Proteins: Structure, Function and Bioinformatics, 2013, 81, 1593-1609.	2.6	6
155	On the Adaptivity of Stochastic Gradient-Based Optimization. SIAM Journal on Optimization, 2020, 30, 1473-1500.	2.0	6
156	A 1.5n/cfs Unsupervised Online Learning Classifier for Seizure Detection. , 2021, , .		6
157	QuTE: Decentralized multiple testing on sensor networks with false discovery rate control. , 2017, , .		5
158	Ergodic mirror descent. , 2011, , .		4
159	Matrix-Variate Dirichlet Process Priors with Applications. Bayesian Analysis, 2014, 9, .	3.0	4
160	Decoding from Pooled Data: Sharp Information-Theoretic Bounds. SIAM Journal on Mathematics of Data Science, 2019, 1, 161-188.	1.8	4
161	Universal Domain Adaptation. , 2020, , 195-211.		4
162	Adaptivity of Stochastic Gradient Methods for Nonconvex Optimization. SIAM Journal on Mathematics of Data Science, 2022, 4, 634-648.	1.8	4

#	ARTICLE	IF	CITATIONS
163	Geometric methods for sampling, optimization, inference, and adaptive agents. Handbook of Statistics, 2022, , 21-78.	0.6	4
164	A permutation-augmented sampler for DP mixture models. , 2007, , .		3
165	Mining Massive Amounts of Genomic Data: A Semiparametric Topic Modeling Approach. Journal of the American Statistical Association, 2017, 112, 921-932.	3.1	3
166	Evaluating Sensitivity to the Stick-Breaking Prior in Bayesian Nonparametrics (with Discussion). Bayesian Analysis, 2023, 18, .	3.0	3
167	Nonparametric Bayesian Identification of Jump Systems with Sparse Dependencies. IFAC Postprint Volumes IPPV / International Federation of Automatic Control, 2009, 42, 1591-1596.	0.4	2
168	Computational Thinking, Inferential Thinking and "Big Data". , 2015, , .		2
169	Latent Marked Poisson Process with Applications to Object Segmentation. Bayesian Analysis, 2018, 13, .	3.0	2
170	Function-Specific Mixing Times and Concentration Away from Equilibrium. Bayesian Analysis, 2020, 15, .	3.0	2
171	A Bayesian nonparametric approach to super-resolution single-molecule localization. Annals of Applied Statistics, 2021, 15, .	1.1	2
172	Leo Breiman. Annals of Applied Statistics, 2010, 4, .	1.1	1
173	Nonparametric Combinatorial Sequence Models. Journal of Computational Biology, 2011, 18, 1649-1660.	1.6	1
174	Decoding from pooled data: Phase transitions of message passing. , 2017, , .		1
175	Unsupervised Online Learning for Long-Term High Sensitivity Seizure Detection. , 2020, 2020, 528-531.		1
176	Nonparametric Combinatorial Sequence Models. Lecture Notes in Computer Science, 2011, , 516-530.	1.3	1
177	On the Inference of Ancestries in Admixed Populations. , 2008, , 424-433.		1
178	Private Prediction Sets. , 0, , .		1
179	ICMLA 2008 Invited Speakers. , 2008, , .		0
180	Visually Relating Gene Expression and in vivo DNA Binding Data. , 2011, , .		0

#	ARTICLE	IF	CITATIONS
181	Changepoint Analysis for Efficient Variant Calling. Lecture Notes in Computer Science, 2014, , 20-34.	1.3	0
182	Saturating Splines and Feature Selection. Journal of Machine Learning Research, 2018, 18, .	62.4	0