

# References

- [1] Gabriella Csurka *et al.* Visual categorization with bags of keypoints. *Workshop on Statistical Learning in Computer Vision*, ECCV, volume 1, no. 1–22, 2004.
- [2] Jianguo Zhang *et al.* Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision* **73**(2) 213–238, 2007.
- [3] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, **60**(2) 91–110, 2004.
- [4] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2. IEEE, 2006.
- [5] Jianchao Yang *et al.* Linear spatial pyramid matching using sparse coding for image classification. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.
- [6] Geoffrey Hinton *et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE* **29**(6) 82–97, 2012.
- [7] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*. An MIT Press book in preparation. Draft chapters available at <http://www.iro.umontreal.ca/~bengioy/dlbook> (2014).
- [8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*(2014).
- [9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature* **521**(7553) 436–444, 2015.
- [10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 2012.
- [11] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.
- [12] World economic outlook database, <https://www.imf.org/external/pubs/ft/weo/2013/02/weodata/index.aspx>.
- [13] Anelia Angelova, Yaser Abu-Mostafa, and Pietro Perona. Pruning training sets for learning of object categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pp. 494–501. IEEE, 2005.
- [14] Sitaram Asur and Bernardo A Huberman. Predicting the future with social media. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010*, volume 1, pp. 492–499. IEEE, 2010.

- [15] Horace Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, **12**(3) 241–253, 2001.
- [16] Horace B Barlow. The coding of sensory messages. In *Current Problems in Animal Behaviour*, pp. 331–360, 1961.
- [17] Yoshua Bengio, Yann LeCun, *et al.* Scaling learning algorithms towards AI. *Large-scale Kernel Machines*, **34**(5), 2007.
- [18] Dimitri P Bertsekas. Incremental gradient, subgradient, and proximal methods for convex optimization: A survey. In *Optimization for Machine Learning*, 2010, 1–38, MIT Press, 2011.
- [19] Christopher M Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [20] Christopher M Bishop *et al.* *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.
- [21] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pp. 177–186. Springer, 2010.
- [22] Léon Bottou and Chih-Jen Lin. Support vector machine solvers. *Large Scale Kernel Machines*, pp. 301–320, MIT Press, 2007.
- [23] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, **3**(1) 1–122, 2011.
- [24] Stephen Poythress Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [25] Hilton Bristow and Simon Lucey. Why do linear svms trained on hog features perform so well? *arXiv preprint arXiv:1406.2419*, 2014.
- [26] Paul R Burton, David G Clayton, Lon R Cardon, *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**(7145) 661–678, 2007.
- [27] Olivier Chapelle. Training a support vector machine in the primal. *Neural Computation*, **19**(5) 1155–1178, 2007.
- [28] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, **2**(4) 303–314, 1989.
- [29] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pp. 886–893. IEEE, 2005.
- [30] Richard O Duda, Peter E Hart, and David G Stork. *Pattern Classification*. John Wiley & Sons, 2012.
- [31] Jeremy Elson, John R Douceur, Jon Howell, and Jared Saul. Asirra: a captcha that exploits interest-aligned manual image categorization. In *ACM Conference on Computer and Communications Security*, pp. 366–374. Citeseer, 2007.
- [32] Markus Enzweiler and Darius M Gavrilă. Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **31**(12) 2179–2195, 2009.
- [33] Carmen Fernandez, Eduardo Ley, and Mark FJ Steel. Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, **16**(5) 563–576, 2001.
- [34] Jerome Friedman, Trevor Hastie, Robert Tibshirani, *et al.* Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, **28**(2) 337–407, 2000.

- [35] Galileo Galilei. *Dialogues Concerning Two New Sciences*. Dover, 1914.
- [36] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. JMLR W&CP Volume*, volume 15, pp. 315–323, 2011.
- [37] James Douglas Hamilton. *Time Series Analysis*, volume 2. Princeton University Press, 1994.
- [38] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**(5) 359–366, 1989.
- [39] Dilawar (<http://math.stackexchange.com/users/1674/dilawar>). Largest eigenvalue of a positive semi-definite matrix is less than or equal to sum of eigenvalues of its diagonal blocks. Mathematics Stack Exchange. URL:<http://math.stackexchange.com/q/144890> (version: 2012-05-14).
- [40] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, *et al.* *Spoken Language Processing*, volume 18. Prentice Hall, 2001.
- [41] Judson P Jones and Larry A Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, **58**(6) 1233–1258, 1987.
- [42] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 1097–1105, NIPS, 2012.
- [43] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10), MIT Press, 1995.
- [44] Yann LeCun, Koray Kavukcuoglu, and Clément Farabet. Convolutional networks and applications in vision. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pp. 253–256. IEEE, 2010.
- [45] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, pp. 556–562, MIT Press, 2001.
- [46] Donghoon Lee, Wilbert Van der Klaauw, Andrew Haughwout, Meta Brown, and Joelle Scally. Measuring student debt and its performance. *FRB of New York Staff Report*, (668), 2014.
- [47] Moshe Lichman. UCI Machine Learning Repository, [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
- [48] Jianqiang Lin, Sang-Mok Lee, Ho-Joon Lee, and Yoon-Mo Koo. Modeling of typical microbial cell growth in batch culture. *Biotechnology and Bioprocess Engineering*, **5**(5) 382–385, 2000.
- [49] Zhiyun Lu, Avner May, Kuan Liu, *et al.* How to scale up kernel methods to be as good as deep neural nets. *arXiv preprint arXiv:1411.4000*, 2014.
- [50] David G Luenberger. *Linear and Nonlinear Programming*. Springer, 2003.
- [51] David J C MacKay. Introduction to gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, **168** 133–166, 1998.
- [52] David J C MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
- [53] Saturnino Maldonado-Bascon, Sergio Lafuente-Arroyo, Pedro Gil-Jimenez, Hilario Gomez-Moreno, and Francisco López-Ferreras. Road-sign detection and recognition based on support vector machines. *Intelligent Transportation Systems, IEEE Transactions on*, **8**(2):264–278, 2007.
- [54] Christopher D Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

- [55] Stjepan Marčelja. Mathematical description of the responses of simple cortical cells. *JOSA*, **70**(11) 1297–1300, 1980.
- [56] Valerii Mayer and Ekaterina Varaksina. Modern analogue of ohm's historical experiment. *Physics Education*, **49**(6) 689, 2014.
- [57] Gordon E Moore. Cramming more components onto integrated circuits. *Proceedings of the IEEE*, **86**(1): 82–85, 1998.
- [58] Isaac Newton. *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press, 1999.
- [59] Jorge Nocedal and S Wright. *Numerical Optimization, Series in Operations Research and Financial Engineering*. Springer-Verlag, 2006.
- [60] Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, **37**(23) 3311–3325, 1997.
- [61] Brad Osgood. The Fourier transform and its applications. Electrical Engineering Department, Stanford University, 2009.
- [62] Reggie Panaligan and Andrea Chen. Quantifying movie magic with google search. *Google Whitepaper–Industry Perspectives+ User Insights*, 2013.
- [63] Jooyoung Park and Irwin W Sandberg. Universal approximation using radial-basis-function networks. *Neural Computation*, **3**(2) 246–257, 1991.
- [64] Jeffrey Pennington, Felix Yu, and Sanjiv Kumar. Spherical random features for polynomial kernels. In *Advances in Neural Information Processing Systems*, pages 1837–1845, NIPS, 2015.
- [65] Simon J D Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, 2012.
- [66] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, **12**(1) 145–151, 1999.
- [67] Lawrence R Rabiner and Bing-Hwang Juang. *Fundamentals of Speech Recognition*, volume 14, Prentice-Hall, 1993.
- [68] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, pp. 1177–1184, NIPS, 2007.
- [69] Ali Rahimi and Benjamin Recht. Uniform approximation of functions with random bases. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pp. 555–561. IEEE, 2008.
- [70] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *The Journal of Machine Learning Research*, **5** 101–141, 2004.
- [71] Walter Rudin. *Principles of Mathematical Analysis*, volume 3. McGraw-Hill, 1964.
- [72] Xavier X Sala-i Martin. I just ran two million regressions. *The American Economic Review*, pp. 178–183, 1997.
- [73] Jonathan Shewchuk. An introduction to the conjugate gradient method without the agonizing pain, <http://www-2.cs.cmu.edu/jrs/jrspapers>, 1994.
- [74] Elias M Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*, volume 1. Princeton University Press, 2011.
- [75] Samuele Straulino. Reconstruction of Galileo Galilei's experiment: the inclined plane. *Physics Education*, **43**(3) 316, 2008.
- [76] Silke Szymczak, Joanna M Biernacka, Heather J Cordell, *et al.* Machine learning in genome-wide association studies. *Genetic Epidemiology*, **33**(S1) S51–S57, 2009.
- [77] Yichuan Tang. Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239, 2013.

- [78] Andrea Vedaldi and Brian Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the International Conference on Multimedia*, pp. 1469–1472. ACM, 2010.
- [79] Pierre Verhulst. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique* **10**: 113–121. Technical report, Retrieved 09/08, 2009.
- [80] Patrik Waldmann, Gábor Mészáros, Birgit Gredler, Christian Fürst, and Johann Sölkner. Evaluation of the lasso and the elastic net in genome-wide association studies. *Frontiers in Genetics*, **4**, 2013.
- [81] Horn, Roger A and Johnson, Charles R. *Matrix analysis*. Cambridge University Press, 2012.