**PIP Summer School 2016  
Unsupervised & deep learning references**

**Books about machine learning**

David Barber's [Bayesian Reasoning and Machine Learning](http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage" \t "_blank)

Bishop's [Pattern Recognition and Machine Learning](http://research.microsoft.com/en-us/um/people/cmbishop/prml/" \t "_blank)

Hastie, Tibshirani, and Friedman's [The Elements of Statistical Learning](http://statweb.stanford.edu/~tibs/ElemStatLearn/" \t "_blank)

Kevin Murphy's [Machine learning: a Probabilistic Perspective](http://www.cs.ubc.ca/~murphyk/MLbook/" \t "_blank)

Foundations of Machine Learning,[Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar](http://www.amazon.com/dp/026201825X" \t "_blank)

Learning From Data,[Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin](http://www.amazon.com/gp/product/1600490069" \t "_blank)

Information Theory, Inference, and Learning Algorithms,[David J. C. MacKay](http://www.inference.phy.cam.ac.uk/itila/)[

All of Statistics,[Larry Wasserman](http://www.amazon.com/All-Statistics-Statistical-Inference-Springer/dp/0387402721" \t "_blank)

Probabilistic Graphical Models: Principles and Techniques, [Daphne Koller, Nir Friedman](http://www.amazon.com/Probabilistic-Graphical-Models-Principles-Computation/dp/0262013193" \t "_blank)

[Machine Learning with R](http://bit.ly/16bxAum" \t "_blank)

[Building Machine Learning Systems with Python](http://bit.ly/1vHzOWS" \t "_blank)

[Matrix Computations (Johns Hopkins Studies in the Mathematical Sciences): Gene H. Golub, Charles F. Van Loan: 9781421407944: Amazon.com: Books](http://www.amazon.com/Computations-Hopkins-Studies-Mathematical-Sciences/dp/1421407949/ref=sr_1_1?ie=UTF8&qid=1456795538&sr=8-1&keywords=matrix+computations" \t "_blank)

**Clustering**

A1-Sultan K. S., A tabu search approach to the clustering problem, Pattern Recognition, 28:1443-1451,1995.

Al-Sultan K. S. , Khan M. M. : Computational experience on four algorithms for the hard clustering problem. Pattern Recognition Letters 17(3): 295-308, 1996.

Banfield J. D. and Raftery A. E. . Model-based Gaussian and non-Gaussian clustering. Biometrics, 49:803-821, 1993.

Bentley J. L. and Friedman J. H., Fast algorithms for constructing minimal spanning trees in coordinate spaces. IEEE Transactions on Computers, C- 27(2):97-105, February 1978. 275

Bonner, R., On Some Clustering Techniques. IBM journal of research and development, 8:22-32, 1964.

Can F. , Incremental clustering for dynamic information processing, in ACM Transactions on Information Systems, no. 11, pp 143-164, 1993.

Cheeseman P., Stutz J.: Bayesian Classification (AutoClass): Theory and Results. Advances in Knowledge Discovery and Data Mining 1996: 153-180

Dhillon I. and Modha D., Concept Decomposition for Large Sparse Text Data Using Clustering. Machine Learning. 42, pp.143-175. (2001).

Dempster A.P., Laird N.M., and Rubin D.B., Maximum likelihood from incomplete data using the EM algorithm. Journal of the Royal Statistical Society, 39(B), 1977.

Duda, P. E. Hart and D. G. Stork, Pattern Classification, Wiley, New York, 2001.

Ester M., Kriegel H.P., Sander S., and Xu X., A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and U. Fayyad, editors, Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), pages 226-231, Menlo Park, CA, 1996. AAAI, AAAI Press.

Estivill-Castro, V. and Yang, J. A Fast and robust general purpose clustering algorithm. Pacific Rim International Conference on Artificial Intelligence, pp. 208-218, 2000.

Fraley C. and Raftery A.E., “How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis”, Technical Report No. 329. Department of Statistics University of Washington, 1998.

B.S. Everitt, Unsolved Problems in Cluster Analysis, Biometrics, vol. 35, 169-182, 1979.

Fisher, D., 1987, Knowledge acquisition via incremental conceptual clustering, in machine learning 2, pp. 139-172.

Fortier, J.J. and Solomon, H. 1996. Clustering procedures. In proceedings of the Multivariate Analysis, ’66, P.R. Krishnaiah (Ed.), pp. 493-506.

Gluck, M. and Corter, J., 1985. Information, uncertainty, and the utility of categories. Proceedings of the Seventh Annual Conference of the Cognitive Science Society (pp. 283-287). Irvine, California: Lawrence Erlbaum Associates.

Guha, S., Rastogi, R. and Shim, K. CURE: An efficient clustering algorithm for large databases. In Proceedings of ACM SIGMOD International Conference on Management of Data, pages 73-84, New York, 1998.

Han, J. and Kamber, M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2001.

Hartigan, J. A. Clustering algorithms. John Wiley and Sons., 1975.

Huang, Z., Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery, 2(3), 1998.

Hoppner F. , Klawonn F., Kruse R., Runkler T., Fuzzy Cluster Analysis, Wiley, 2000.

Hubert, L. and Arabie, P., 1985 Comparing partitions. Journal of Classification, 5. 193-218.

Jain, A.K. Murty, M.N. and Flynn, P.J. Data Clustering: A Survey. ACM Computing Surveys, Vol. 31, No. 3, September 1999.

A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, New Jersey, 1988.

A.K. Jain, Data clustering : 50 years beyond K-means. Pattern Recognition Letters, vol.31, no.8, 651-666, 2010.

Kaufman, L. and Rousseeuw, P.J., 1987, Clustering by Means of Medoids, In Y. Dodge, editor, Statistical Data Analysis, based on the L1 Norm, pp. 405- 416, Elsevier/North Holland, Amsterdam.

Kim, D.J., Park, Y.W. and Park,. A novel validity index for determination of the optimal number of clusters. IEICE Trans. Inf., Vol. E84-D, no.2, 2001, 281-285.

King, B. Step-wise Clustering Procedures, J. Am. Stat. Assoc. 69, pp. 86-101, 1967.

Larsen, B. and Aone, C. 1999. Fast and effective text mining using linear-time document clustering. In Proceedings of the 5th ACM SIGKDD, 16-22, San Diego, CA.

Marcotorchino, J.F. and Michaud, P. Optimisation en Analyse Ordinale des Donns. Masson, Paris.

Mishra, S. K. and Raghavan, V. V., An empirical study of the performance of heuristic methods for clustering. In Pattern Recognition in Practice, E. S. Gelsema and L. N. Kanal, Eds. 425436, 1994.

Murtagh, F. A survey of recent advances in hierarchical clustering algorithms which use cluster centers. Comput. J. 26 354-359, 1984.

Ng, R. and Han, J. 1994. Very large data bases. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB94, Santiago, Chile, Sept.), VLDB Endowment, Berkeley, CA, 144155.

A.S. Pandya and R.B. Macy, Pattern Recognition with Neural Networks in C++, IEEE Press, 1995.

Rand, W. M., Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66: 846–850, 1971.

Ray, S., and Turi, R.H. Determination of Number of Clusters in K-Means Clustering and Application in Color Image Segmentation. Monash university, 1999.

Selim, S.Z., and Ismail, M.A. K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. In IEEE transactions on pattern analysis and machine learning, vol. PAMI-6, no. 1, January, 1984.

Selim, S. Z. AND Al-Sultan, K. 1991. A simulated annealing algorithm for the clustering problem. Pattern Recogn. 24, 10 (1991), 10031008.

Sneath, P., and Sokal, R. Numerical Taxonomy. W.H. Freeman Co., San Francisco, CA, 1973.

Strehl A. and Ghosh J., Clustering Guidance and Quality Evaluation Using Relationship-based Visualization, Proceedings of Intelligent Engineering Systems Through Artificial Neural Networks, 2000, St. Louis, Missouri, USA, pp 483-488.

**Dimensionality Reduction**

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15(6):1373–1396, 2003.

V. Cherkassky and F. Mulier. Learning from data. Wiley, New York, 1998.

T. Cox and M. Cox. Multidimensional Scaling. Chapman Hall, Boca Raton, 2nd edition, 2001.

B. Frey. Graphical models for machine learning and digital communication. MIT Press, Cambridge, Mass, 1998.

J. Ham, D. Lee, S. Mika, and Sch¨olkopf B. A kernel view of the dimensionality reduction of manifolds. In International Conference on Machine Learning, 2004.

H. Hotelling. Analysis of a complex of statistical variables into components. J. of Educational Psychology, 24:417–441, 1933.

A. Hyv¨arinen. Survey on independent component analysis. Neural Computing Surveys, 2:94–128, 1999.

I. Jolliffe. Principal Component Analysis. Springer-Verlag, New York, 1986.

V. Kumar, A. Grama, A. Gupta, and G. Karypis. Introduction to parallel computing. Benjamin, Cummings, 1994.

S. Mika, B. Sch¨olkopf, A. Smola, K.-R. M¨uller, M. Scholz, and G. R¨atsch. Kernel PCA and de-noising in feature spaces. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, Proceedings NIPS 11. MIT Press, 1999.

K. Pearson. On lines and planes of closest fit to systems of points in space. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, Sixth Series 2:559–572, 1901

S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290:2323–2326, 2000.

L. Saul and S. Roweis. Think globally, fit locally: Unsupervised learning of nonlinear manifolds. JMLR, 2003.

B. Sch¨olkopf, S. Mika, A. Smola, G. R¨atsch, and K.-R. M¨uller. Kernel PCA pattern reconstruction via approximate pre-images. In L. Niklasson, M. Bod´en, and T. Ziemke, editors, Proceedings of the 8th International Conference on Artificial Neural Networks, pages 147–152, 1998.

B. Sch¨olkopf and A. Smola. Learning with Kernels. MIT Press, Cambridge, Massachusetts, 2002.

R. Rivest T. Cormen, C. Leiserson and C. Stein. Introduction to algorithms. MIT Press, Cambridge, Massachusetts, 2001.

J. Friedman T. Hastie, R. Tibshirani. The elements of statistical learning. Springer, New York, 2002.

J. Tenenbaum. Mapping a manifold of perceptual observations. In Advances in Neural Information Processing Systems 10, pages 682–687, 1998.

J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. Science, 290:2319–2323, 2000.

K. Weinberger and L. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In Proceedings of the International Conference on Machine Learning, pages 839–846, 2004.

K. Weinberger and L. Saul. Unsupervised learning of image manifolds by semidefinite programing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 988–995, 2004.

C. K. I. Williams. On a connection between kernel PCA and metric multidimensional scaling. Machine Learning, 46(1-3):11–19, 2002.

P. Vincent Y. Bengio and J.-F. Paiement. Learning eigenfunctions of similarity: Linking spectral clustering and kernel pca. Technical Report 1232, Universite de Montreal, 2003.

**Useful links**

General machine learning: <https://github.com/josephmisiti/awesome-machine-learning/>

Website for competitions, datasets, etc. <https://www.kaggle.com/>

t-SNE website: <https://lvdmaaten.github.io/tsne/>

**Deep Learning**

B. Olshausen, D. Field. Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. Nature, 1996.

•H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. NIPS, 2007.

•R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng. Self-taught learning: Transfer learning from unlabeled data. ICML, 2007.

•H. Lee, R. Raina, A. Teichman, and A. Y. Ng. Exponential Family Sparse Coding with Application to Self-taught Learning. IJCAI, 2009.

•J. Yang, K. Yu, Y. Gong, and T. Huang. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. CVPR, 2009.

•Y. Bengio. Learning Deep Architectures for AI, Foundations and Trends in Machine Learning, 2009.

•Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. NIPS, 2007.

•P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol. Extracting and composing robust features with denoising autoencoders. ICML, 2008.

•H. Lee, C. Ekanadham, and A. Y. Ng. Sparse deep belief net model for visual area V2. NIPS, 2008.

•Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural Computation, 1:541–551, 1989.

•H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. ICML, 2009.

•H. Lee, Y. Largman, P. Pham, and A. Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. NIPS, 2009.

•A. R. Mohamed, G. Dahl, and G. E. Hinton, Deep belief networks for phone recognition. NIPS 2009 workshop on deep learning for speech recognition.

•G. Dahl, M. Ranzato, A. Mohamed, G. Hinton, Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine, NIPS 2010

•M. Ranzato, A. Krizhevsky, G. E. Hinton, Factored 3-Way Restricted Boltzmann Machines for Modeling Natural Images. AISTATS, 2010.

M. Ranzato, G. E. Hinton. Modeling Pixel Means and Covariances Using Factorized Third-Order Boltzmann Machines. CVPR, 2010.

•G. Taylor, G. E. Hinton, and S. Roweis. Modeling Human Motion Using Binary Latent Variables. NIPS, 2007.

•G. Taylor and G. E. Hinton. Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style. ICML, 2009.

•G. Taylor, R. Fergus, Y. LeCun and C. Bregler. Convolutional Learning of Spatio-temporal Features. ECCV, 2010.

•K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. LeCun, Learning Invariant Features through Topographic Filter Maps. CVPR, 2009.

•K. Kavukcuoglu, M. Ranzato, and Y. LeCun, Fast Inference in Sparse Coding Algorithms with Applications to Object Recognition. CBLL-TR-2008-12-01, 2008.

•K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, What is the Best Multi-Stage Architecture for Object Recognition? ICML, 2009.

•R. Salakhutdinov and I. Murray. On the Quantitative Analysis of Deep Belief Networks. ICML, 2008.

•R. Salakhutdinov and G. E. Hinton. Deep Boltzmann machines. AISTATS, 2009.

•K. Yu, T. Zhang, and Y. Gong. Nonlinear Learning using Local Coordinate Coding, NIPS, 2009.

•J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Learning Locality-constrained Linear Coding for Image Classification. CVPR, 2010.

•H. Larochelle, Y. Bengio, J. Louradour and P. Lamblin, Exploring Strategies for Training Deep Neural Networks, JMLR, 2009.

•D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent and S. Bengio, Why Does Unsupervised Pre-training Help Deep Learning? JMLR, 2010.

•J. Yang, K. Yu, and T. Huang. Supervised Translation-Invariant Sparse Coding. CVPR, 2010.

•Y. Boureau, F. Bach, Y. LeCun and J. Ponce: Learning Mid-Level Features for Recognition. CVPR, 2010.

•I. J. Goodfellow, Q. V. Le, A. M. Saxe, H. Lee, and A. Y. Ng. Measuring invariances in deep networks. NIPS, 2009.

**Useful links:**

<http://deeplearning.net>

- Detailed list of books, tutorials, code, etc.

<https://github.com/ChristosChristofidis/awesome-deep-learning>

Introductory references:

<http://deeploria.gforge.inria.fr/stories/introductory-references/>

- Deep vision: <https://github.com/kjw0612/awesome-deep-vision>

- Recurrent neural networks literature: <https://github.com/kjw0612/awesome-rnn>