Yoshua Bengio

Département d'informatique
et recherche opérationnelle,
Université de Montréal

Phone : 514-343-6804
Fax : 514-343-5834
Yoshua.Bengio@umontreal.ca
www.iro.umontreal.ca/∼bengioy

Canada Research Chair
on Statistical Learning Algorithms

Titles and Distinctions

• **Full Professor**, Université de Montréal, since 2002. Previously associate professor (1997-2002) and assistant professor (1993-1997).

• **Canada Research Chair** on *Statistical Learning Algorithms* since 2000 (tier 2 : 2000-2005, tier 1 since 2006).

• **NSERC-Ubisoft Industrial Research Chair**, since 2011. Previously NSERC-CGI Industrial Chair, 2005-2010.

• **Recipient** of the ACFAS Urgel-Archambault 2009 prize (covering physics, mathematics, computer science, and engineering).

• **Fellow**, CIFAR (Canadian Institute For Advanced Research), since 2004. Co-director of the CIFAR NCAP program since 2014.

• Member of the board of the **Neural Information Processing Systems (NIPS) Foundation**, since 2010.

• **Action Editor**, *Journal of Machine Learning Research* (JMLR), *Neural Computation*, *Foundations and Trends in Machine Learning*, and *Computational Intelligence*. Member of the 2012 editor-in-chief nominating committee for JMLR.

• **Fellow**, CIRANO (Centre Inter-universitaire de Recherche en Analyse des Organisations), since 1997.

• Previously **Associate Editor**, *Machine Learning*, *IEEE Trans. on Neural Networks*.

• **Founder (1993) and head** of the *Laboratoire d'Informatique des Systèmes Adaptatifs* (LISA), and the *Montreal Institute for Learning Algorithms* (MILA), currently including 5 faculty, 40 students, 5 post-docs, 5 researchers on staff, and numerous visitors. This is probably the largest concentration of deep learning researchers in the world (around 60 researchers).

• Member of the board of the Centre de Recherches Mathématiques, UdeM, 1999-2009. Member of the Awards Committee of the Canadian Association for Computer Science (2012-2013). Member of the NIPS'2012 committee for best paper awards.

• *Program Co-Chair*, NIPS'2008, *General Chair*, NIPS'2009. Note that *Advances in Neural Information Processing Systems* (NIPS) is a very high level conference, the most important in my field ($>$1000 submissions), with reviewing and acceptance criteria comparing favorably to best journals (acceptance rate between 20% and 25%). Over 48 papers published in NIPS over the years puts me among the most important contributors to the NIPS community.

• Area chair or member of the program committee (I don't mean reviewer, but meta-reviewer) for numerous other conferences, including NIPS'95, NIPS'2004, ICML'2003, ICML'2004, ICML'2006, ICML'2008, AISTATS'2001, ICPR'2002, ICONIP'1996, IJCNN'2000, CAp'2004, CAp'2006, CAp'2010, CAp'2011, ICML'2012, ICML'2013.

• Member of *grant selection committees* for Quebec's FQRNT (1999-2000), and Canada's NSERC (2000-2003, 2006-2007).

Academic Studies and Diplomas

1992–1993	**Post-doctoral Fellow, Learning and vision algorithms,** *Yann LeCun's group,* AT&T Bell Laboratories, New-Jersey.
1991–1992	**Post-doctoral Fellow,** *NSERC scholarship,* **Statistical learning / sequential data,** *Michael I. Jordan's group,* Brain and Cognitive Sciences Dept., MIT, Massachusetts.
1988–1991	**Ph.D. in computer science,** *NSERC scholarship,* **Neural Networks and Markovian Models,** Computer Science Dept., McGill University.
1986–1988	**M.Sc. in computer science** , CGPA 4.0/4.0, *Speech recognition with statistical methods,* Computer Science Dept., McGill University.
1982–1986	**B.Eng. in computer engineering,** *Honours class, CGPA 3.82/4.0* Electrical Engineering Dept., McGill University.

Research Areas

• Improving the generalization power of learning algorithms.
- Deep Learning : learning abstractions in deep architectures.
- Non-local learning of complex functions.
- Learning from high-dimensional data.
- Learning generative and structured output models.
- {Transfer, multi-task, semi-supervised, multi-modal} learning.

• Temporal dependencies.
- Recurrent neural networks
- Neural Language Models, representation-learning for NLP

• Brain-inspired learning algorithms : deterministic and stochastic neural networks, high-order neuron models, biologically inspired learning algorithms.

• Analysis of data with temporal structure and non-stationarity :
- Applications involving audio and visual data.
- Learning long-term dependencies and sequential structure.
- Learning to model natural language.
- Applications to video games, marketing, finance, insurance, and advertising.

• Computational efficiency, parallel implementation, online learning & Big Data.

• Statistical analysis of simulations experiments of learning algorithms on real data, hyper-parameter optimization / model selection.

Major Scientific and Technical Impact

• More than 26 500 citations to scientific publications authored by Yoshua Bengio found by *Google Scholar* in June 2015, with an H-index of 65.

• 70 scientific journal articles, 146 conference proceeding papers, 2 books (a 3rd on the way), 19 book chapters, 3 patents, 35 technical reports.

• Co-created the rapidly expanding area of deep learning in 2006, with Hinton's (Toronto), LeCun's (NYU) and Ng's labs, attracting much attention from the NIPS/ICML community, funding agencies (NSF, DARPA), the media and high-tech industry (Google, Microsoft, Facebook, Samsung, Intel...).

- — Deep learning allowed us to win the final phase of the Unsupervised and Transfer Learning Challenge (presented at ICML 2011) and the Transfer Learning Challenge (presented at the NIPS 2011 Workshops).
- — Major contributions to the theoretical justification and analysis of deep learning and distributed representations.
- — The first use of rectifiers in very deep purely supervised networks (AISTATS'2011), and of the improved initialization of deep nets without pretraining (AISTATS'2010), both core to recent success of deep nets in speech and object recognition.
- — The first asynchronous stochastic gradient descent for parallelizing neural networks (JMLR'2003), a technique now used at Google on a very large scale.
- — The first and highly cited neural probabilistic language models, based on learning distributed word embeddings (JMLR'2003) and associated hierarchical structured output to speed up training and sampling (AISTATS'2005).
- — The first book on deep learning (2009), currently writing a more comprehensive book, for MIT Press.
- — Bengio et al 1994 (IEEE Transactions on Neural Networks) has overturned a whole research field (gradient-based learning in recurrent neural networks) by showing a fundamental challenge.

• 175 invited talks, at places such as NIPS workshops, ICML workshops, Stanford, Oxford, Google, CMU, Microsoft Research, Google Research, the Fields Institute, MIT, University College London, New York University, Facebook AI Research, Johns Hopkins U., U. Toronto, IBM Research, Yahoo, Qualcomm, Samsung, Nuance, the Gatsby Unit, Deep Mind, the ICONIP conference, DARPA, summer schools. Tutorials at ACL 2012, AAAI 2013, IPAM, SSTiC 2013.

• Created the ICLR (Int. Conf. on Learning Representations) in 2013 and organized or co-organized more than 21 workshops in career, including the Learning Workshop every year since 1998, the Deep Learning Workshop at NIPS in 2007, the ICML'2009 workshop on Learning Feature Hierarchies, the NIPS'2010, 2011, 2012 and 2013 workshops on Deep Learning and Unsupervised Feature Learning, and the ICML 2012 and 2013 workshop on Representation Learning.

2 Research Contributions

2.1 Refereed Journal Publications

[1] G. Mesnil, Y. Dauphin, K. Yao, Y. Bengio, L. Deng, D. Hakkani-Tur, X. He, L. Heck, G. Tur, D. Yu, and G. Zweig, "Using recurrent neural networks for slot filling in spoken language understanding," *IEEE Tr. ASSP*, 2015.

[2] G. Alain and Y. Bengio, "What regularized auto-encoders learn from the data-generating distribution," *Journal of Machine Learning Research*, vol. 15, pp. 3563–3593, 2014.

[3] A. Courville, G. Desjardins, J. Bergstra, and Y. Bengio, "The spike-and-slab RBM and extensions to discrete and sparse data distributions," *IEEE Tr. PAMI*, vol. 36, no. 9, pp. 1874–1887, 2014.

[4] Y. Bengio, A. Courville, and P. Vincent, "Representation learning : A review and new perspectives," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 35, no. 8, pp. 1798–1828, 2013.

[5] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "A semantic matching energy function for learning with multi-relational data," *Machine Learning : Special Issue on Learning Semantics*, 2013.

[6] G. Mesnil, A. Bordes, J. Weston, G. Chechik, and Y. Bengio, "Learning semantic representations of objects and their parts," *Machine Learning : Special Issue on Learning Semantics*, 2013.

[7] O. Delalleau, E. Contal, E. Thibodeau-Laufer, R. Chandias Ferrari, Y. Bengio, and F. Zhang, "Beyond skill rating : Advanced matchmaking in ghost recon online," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, pp. 167–177, Sept. 2012.

[8] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, Feb. 2012.

[9] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning algorithms for the classification restricted Boltzmann machine," *JMLR*, vol. 13, pp. 643–669, Mar. 2012.

[10] Y. Bengio, N. Chapados, O. Delalleau, H. Larochelle, and X. Saint-Mleux, "Detonation classification from acoustic signature with the restricted Boltzmann machine," *Computational Intelligence*, vol. 28, no. 2, 2012.

[11] O. Breuleux, Y. Bengio, and P. Vincent, "Quickly generating representative samples from an RBM-derived process," *Neural Computation*, vol. 23, pp. 2053–2073, Aug. 2011.

[12] J. Bergstra, Y. Bengio, and J. Louradour, "Suitability of V1 energy models for object classification," *Neural Computation*, vol. 23, p. 774–790, Mar. 2011.

[13] M. Mandel, R. Pascanu, D. Eck, Y. Bengio, L. M. Aeillo, R. Schifanella, and F. Menczer, "Contextual tag inference," *ACM T. Multimedia Comp., Comm. & Appl.*, vol. 7S, p. 1–32, Oct. 2011.

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders : Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, Dec. 2010.

[15] H. Larochelle, Y. Bengio, and J. Turian, "Tractable multivariate binary density estimation and the restricted Boltzmann forest," *Neural Computation*, vol. 22, pp. 2285–2307, Sept. 2010.

[16] N. Le Roux and Y. Bengio, "Deep belief networks are compact universal approximators," *Neural Computation*, vol. 22, pp. 2192–2207, Aug. 2010.

[17] Y. Bengio, O. Delalleau, and C. Simard, "Decision trees do not generalize to new variations," *Computational Intelligence*, vol. 26, pp. 449–467, Nov. 2010.

[18] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning ?," *Journal of Machine Learning Research*, vol. 11, pp. 625–660, Feb. 2010.

[19] F. Rivest, J. Kalaska, and Y. Bengio, "Alternative time representations in dopamine models," *Journal of Computational Neuroscience*, vol. 28, no. 1, pp. 107–130, 2009.

[20] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009. Also published as a book. Now Publishers, 2009.

[21] C. Dugas, Y. Bengio, F. Belisle, C. Nadeau, and R. Garcia, "Incorporating functional knowledge in neural networks," *The Journal of Machine Learning Research*, vol. 10, pp. 1239–1262, June 2009.

[22] J. Carreau and Y. Bengio, "A hybrid Pareto mixture for conditional asymmetric fat-tailed distribution," *IEEE Transactions on Neural Networks*, vol. 20, no. 7, pp. 1087–1101, 2009.

[23] J. Carreau and Y. Bengio, "A hybrid Pareto model for asymmetric fat-tailed data : the univariate case," *Extremes*, vol. 12, no. 1, pp. 53–76, 2009.

[24] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *Journal of Machine Learning Research*, vol. 10, pp. 1–40, Jan. 2009.

[25] Y. Bengio and O. Delalleau, "Justifying and generalizing contrastive divergence," *Neural Computation*, vol. 21, pp. 1601–1621, June 2009.

[26] N. Le Roux and Y. Bengio, "Representational power of restricted Boltzmann machines and deep belief networks," *Neural Computation*, vol. 20, pp. 1631–1649, June 2008.

[27] Y. Bengio and J.-S. Sénécal, "Adaptive importance sampling to accelerate training of a neural probabilistic language model," *IEEE Trans. Neural Networks*, vol. 19, no. 4, pp. 713–722, 2008.

[28] Y. Bengio, "Neural net language models," *Scholarpedia*, vol. 3, no. 1, p. 3881, 2008.

[29] Y. Bengio, "On the challenge of learning complex functions," in *Computational Neuroscience : Theoretical Insights into Brain Function* (P. Cisek, J. Kalaska, and T. Drew, eds.), Progress in Brain Research, Elsevier, 2007.

[30] N. Chapados and Y. Bengio, "Noisy k best-paths for approximate dynamic programming with application to portfolio optimization," *Journal of Computers*, vol. 2, no. 1, pp. 12–19, 2007.

[31] Y. Bengio, M. Monperrus, and H. Larochelle, "Nonlocal estimation of manifold structure," *Neural Computation*, vol. 18, no. 10, pp. 2509–2528, 2006.

[32] D. Erhan, P.-J. L'Heureux, S. Y. Yue, and Y. Bengio, "Collaborative filtering on a family of biological targets.," *Journal of Chemical Information and Modeling*, vol. 46, no. 2, pp. 626–635, 2006.

[33] M. C. Zaccaro, H. Boon, M. Pattarawarapan, Z. Xia, A. Caron, P.-J. L'Heureux, Y. Bengio, K. Burgess, and H. U. Saragori, "Selective small molecule peptidomimetic ligands of trkc and trka receptors afford discrete or complete neurotrophic activities," *Chemistry & Biology*, vol. 12, no. 9, pp. 1015–1028, 2005.

[34] P.-J. L'Heureux, J. Carreau, Y. Bengio, O. Delalleau, and S. Y. Yue, "Locally linear

embedding for dimensionality reduction in QSAR," *Journal of Computer-Aided Molecular Design*, vol. 18, pp. 475–482, 2004.

[35] Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, and M. Ouimet, "Learning eigenfunctions links spectral embedding and kernel PCA," *Neural Computation*, vol. 16, no. 10, pp. 2197–2219, 2004.

[36] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of K-fold cross-validation," *Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, 2004.

[37] R. Collobert, Y. Bengio, and S. Bengio., "Scaling large learning problems with hard parallel mixtures," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 3, pp. 349–365, 2003.

[38] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Machine Learning*, vol. 52, no. 3, pp. 239–281, 2003.

[39] J. Ghosn and Y. Bengio, "Bias learning, knowledge sharing," *IEEE Transactions on Neural Networks*, vol. 14, pp. 748–765, July 2003.

[40] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.

[41] Y. Bengio and N. Chapados, "Extensions to metric-based model selection," *Journal of Machine Learning Research*, vol. 3, pp. 1209–1227, Mar. 2003. Special Issue on Feature Selection.

[42] I. Takeuchi, Y. Bengio, and T. Kanamori, "Robust regression with asymmetric heavy-tail noise distributions," *Neural Computation*, vol. 14, no. 10, pp. 2469–2496, 2002.

[43] R. Collobert, S. Bengio, and Y. Bengio, "Parallel mixture of SVMs for very large scale problems," *Neural Computation*, vol. 14, no. 5, pp. 1105–1114, 2002.

[44] O. Chapelle, V. Vapnik, and Y. Bengio, "Model selection for small-sample regression," *Machine Learning Journal*, vol. 48, no. 1, pp. 9–23, 2002.

[45] P. Vincent and Y. Bengio, "Kernel matching pursuit," *Machine Learning*, vol. 48, pp. 165–187, 2002.

[46] N. Chapados and Y. Bengio, "Cost functions and model combination for var–based asset allocation using neural networks," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 890–906, 2001.

[47] Y. Bengio, V.-P. Lauzon, and R. Ducharme, "Experiments on the application of IOHMMs to model financial returns series," *IEEE Transactions on Neural Networks*, vol. 12, no. 1, pp. 113–123, 2001.

[48] H. Schwenk and Y. Bengio, "Boosting neural networks," *Neural Computation*, vol. 12, no. 8, pp. 1869–1887, 2000.

[49] Y. Bengio, "Gradient-based optimization of hyperparameters," *Neural Computation*, vol. 12, no. 8, pp. 1889–1900, 2000.

[50] S. Bengio and Y. Bengio, "Taking on the curse of dimensionality in joint distributions using neural networks," *IEEE Transactions on Neural Networks, special issue on Data Mining and Knowledge Discovery*, vol. 11, no. 3, pp. 550–557, 2000.

[51] Y. Bengio, "Markovian models for sequential data," *Neural Computing Surveys*, vol. 2, pp. 129–162, 1999.

[52] S. Bengio, Y. Bengio, J. Robert, and G. Bélanger, "Stochastic learning of strategic equilibria for auctions," *Neural Computation*, vol. 11, no. 5, pp. 1199–1209, 1999.

[53] L. Bottou, P. Haffner, P. Howard, P. Simard, and Y. Bengio, "High quality document image compression with DjVu," *Journal of Electronic Imaging*, vol. 7, no. 3,

pp. 410–425, 1998.

[54] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Nov. 1998.

[55] Y. Bengio, F. Gingras, B. Goulard, and J.-M. Lina, "Gaussian mixture densities for classification of nuclear power plant data," *Computers and Artificial Intelligence, special issue on Intelligent Technologies for Electric and Nuclear Power Plants*, vol. 17, no. 2–3, pp. 189–209, 1998.

[56] F. Gingras and Y. Bengio, "Handling asynchronous or missing financial data with recurrent networks," *International Journal of Computational Intelligence and Organizations*, vol. 1, no. 3, pp. 154–163, 1998.

[57] Y. Bengio, "Using a financial training criterion rather than a prediction criterion," *International Journal of Neural Systems*, vol. 8, no. 4, pp. 433–443, 1997. Special issue on noisy time-series.

[58] Y. Bengio and P. Frasconi, "Input/Output HMMs for sequence processing," *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1231–1249, 1996.

[59] Y. Bengio and P. Frasconi, "Diffusion of context and credit information in Markovian models," *Journal of Artificial Intelligence Research*, vol. 3, pp. 249–270, 1995.

[60] Y. Bengio, Y. LeCun, C. Nohl, and C. Burges, "Lerec : A NN/HMM hybrid for on-line handwriting recognition," *Neural Computation*, vol. 7, no. 6, pp. 1289–1303, 1995.

[61] S. Bengio, Y. Bengio, and J. Cloutier, "On the search for new learning rules for ANNs," *Neural Processing Letters*, vol. 2, no. 4, pp. 26–30, 1995.

[62] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994. Special Issue on Recurrent Neural Networks, March 94.

[63] Y. Bengio, "A connectionist approach to speech recognition," *International Journal on Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 647–668, 1993. special issue entitled Advances in Pattern Recognition Systems using Neural Networks.

[64] Y. Bengio, M. Gori, and R. De Mori, "Learning the dynamic nature of speech with back-propagation for sequences," *Pattern Recognition Letters*, vol. 13, no. 5, pp. 375–385, 1992. (Special issue on Artificial Neural Networks).

[65] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "Phonetically motivated acoustic parameters for continuous speech recognition using artificial neural networks," *Speech Communication*, vol. 11, no. 2–3, pp. 261–271, 1992. Special issue on neurospeech.

[66] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "Global optimization of a neural network-hidden Markov model hybrid," *IEEE Transactions on Neural Networks*, vol. 3, no. 2, pp. 252–259, 1992.

[67] Y. Bengio and Y. Pouliot, "Efficient recognition of immunoglobulin domains from amino-acid sequences using a neural network," *Computer Applications in the Biosciences*, vol. 6, no. 2, pp. 319–324, 1990.

[68] P. Cosi, Y. Bengio, and R. De Mori, "Phonetically-based multi-layered networks for acoustic property extraction and automatic speech recognition," *Speech Communication*, vol. 9, no. 1, pp. 15–30, 1990.

[69] Y. Bengio and R. D. Mori, "Use of multilayer networks for the recognition of phonetic features and phonemes," *Computational Intelligence*, vol. 5, pp. 134–141, 1989.

[70] Y. Bengio, R. Cardin, R. De Mori, and E. Merlo, "Programmable execution of multi-layered networks for automatic speech recognition," *Communications of the*

Association for Computing Machinery, vol. 32, no. 2, pp. 195–199, 1989.

2.2 Articles in Refereed Conference Proceedings

[71] J. Bornschein and Y. Bengio, "Reweighted wake-sleep," in *ICLR'2015, arXiv :1406.2751*, 2015.

[72] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *ICLR'2015, arXiv :1409.0473*, 2015.

[73] A. Romero, N. Ballas, S. Ebrahimi Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets : Hints for thin deep nets," in *ICLR'2015, arXiv :1412.6550*, 2015.

[74] G. Mesnil, S. Rifai, A. Bordes, X. Glorot, Y. Bengio, and P. Vincent, "Unsupervised learning of semantics of object detections for scene categorization," *Pattern Recognition Applications and Methods, Advances in Intelligent Systems and Computing*, pp. 209–224, 2015.

[75] T. Raiko, L. Yao, K. Cho, and Y. Bengio, "Iterative neural autoregressive distribution estimator (NADE-k)," in *NIPS'2014*, 2014.

[76] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks ?," in *NIPS'2014*, 2014.

[77] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *NIPS'2014*, 2014.

[78] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *NIPS'2014*, 2014.

[79] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization," in *NIPS'2014*, 2014.

[80] Y. Bengio, E. Thibodeau-Laufer, and J. Yosinski, "Deep generative stochastic networks trainable by backprop," in *ICML'2014*, 2014.

[81] M. Chen, K. Weinberger, F. Sha, and Y. Bengio, "Marginalized denoising auto-encoders for nonlinear representations," in *ICML'2014*, 2014.

[82] D. Warde-Farley, I. J. Goodfellow, A. Courville, and Y. Bengio, "An empirical analysis of dropout in piecewise linear networks," in *International Conference on Learning Representations 2014(Conference Track)*, Apr. 2014.

[83] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgeting in gradient-based neural networks," in *International Conference on Learning Representations 2014(Conference Track)*, Apr. 2014.

[84] Y. Bengio, L. Yao, and K. Cho, "Bounding the test log-likelihood of generative models," in *International Conference on Learning Representations 2014(Conference Track)*, Apr. 2014.

[85] R. Pascanu, Ç. Gülçehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," in *International Conference on Learning Representations 2014(Conference Track)*, Apr. 2014.

[86] R. Pascanu, G. Montufar, and Y. Bengio, "On the number of inference regions of deep feed forward networks with piece-wise linear activations," in *International Conference on Learning Representations 2014(Conference Track)*, Apr. 2014.

[87] R. Pascanu and Y. Bengio, "Revisiting natural gradient for deep networks," in *International Conference on Learning Representations 2014(Conference Track)*, Apr. 2014.

[88] Y. Dauphin and Y. Bengio, "Stochastic ratio matching of RBMs for sparse high-dimensional inputs," in *Advances in Neural Information Processing Systems 26 (NIPS'13)*, NIPS Foundation (http ://books.nips.cc), 2013.

[89] Y. Bengio, L. Yao, G. Alain, and P. Vincent, "Generalized denoising auto-encoders as generative models," in *Advances in Neural Information Processing Systems 26 (NIPS'13)*, NIPS Foundation (http ://books.nips.cc), 2013.

[90] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "High-dimensional sequence transduction," in *Proc. ICASSP 3*, 2013.

[91] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," in *ICML'2013*, 2013.

[92] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *ICML'2013*, 2013.

[93] Y. Bengio, G. Mesnil, Y. Dauphin, and S. Rifai, "Better mixing via deep representations," in *Proceedings of the 30th International Conference on Machine Learning (ICML'13)*, ACM, 2013.

[94] H. Luo, P. L. Carrier, A. Courville, and Y. Bengio, "Texture modeling with convolutional spike-and-slab RBMs and deep extensions," in *AISTATS'2013*, 2013.

[95] G. Mesnil, S. Rifai, A. Bordes, X. Glorot, Y. Bengio, and P. Vincent, "Unsupervised and transfer learning under uncertainty : from object detections to scene categorization," in *ICPRAM*, 2013.

[96] S. Rifai, Y. Bengio, Y. Dauphin, and P. Vincent, "A generative process for sampling contractive auto-encoders," in *Proceedings of the Twenty-nine International Conference on Machine Learning (ICML'12)*, ACM, 2012.

[97] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Discriminative non-negative matrix factorization for multiple pitch estimation," in *ISMIR*, 2012.

[98] S. Rifai, Y. Bengio, A. Courville, P. Vincent, and M. Mirza, "Disentangling factors of variation for facial expression recognition," in *European Conference on Computer Vision*, 2012.

[99] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," *AISTATS'2012*, 2012.

[100] I. J. Goodfellow, A. Courville, and Y. Bengio, "Large-scale feature learning with spike-and-slab sparse coding," in *Proceedings of the Twenty-nine International Conference on Machine Learning (ICML'12)*, ACM, 2012.

[101] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences : Application to polyphonic music generation and transcription," in *ICML'2012*, 2012.

[102] S. Rifai, Y. Dauphin, P. Vincent, Y. Bengio, and X. Muller, "The manifold tangent classifier," in *NIPS'2011*, 2011. Student paper award.

[103] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *NIPS'2011*, 2011.

[104] G. Desjardins, A. Courville, and Y. Bengio, "On tracking the partition function," in *NIPS'2011*, 2011.

[105] O. Delalleau and Y. Bengio, "Shallow vs. deep sum-product networks," in *NIPS*, 2011.

[106] Y. Bengio and O. Delalleau, "On the expressive power of deep architectures," in *ALT'2011*, 2011.

[107] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning,"

in *JMLR W&CP : Proc. Unsupervised and Transfer Learning*, 2011.

[108] G. Mesnil, Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley, P. Vincent, A. Courville, and J. Bergstra, "Unsupervised and transfer learning challenge : a deep learning approach," in *JMLR W&CP : Proc. Unsupervised and Transfer Learning*, vol. 7, 2011.

[109] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, "Higher order contractive auto-encoder," in *ECML PKDD*, 2011.

[110] A. Bordes, J. Weston, R. Collobert, and Y. Bengio, "Learning structured embeddings of knowledge bases," in *AAAI 2011*, 2011.

[111] A. Courville, J. Bergstra, and Y. Bengio, "Unsupervised models of images by spike-and-slab RBMs," in *ICML'2011*, 2011.

[112] Y. Dauphin, X. Glorot, and Y. Bengio, "Large-scale learning of embeddings with reconstruction sampling," in *Proceedings of theTwenty-eight International Conference on Machine Learning (ICML'11)*, June 2011.

[113] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders : Explicit invariance during feature extraction," in *ICML'2011*, 2011.

[114] P. Hamel, S. Lemieux, Y. Bengio, and D. Eck, "Temporal pooling and multiscale learning for automatic annotation and ranking of music audio," in *ISMIR*, 2011.

[115] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification : A deep learning approach," in *ICML'2011*, 2011.

[116] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *AISTATS'2011*, 2011.

[117] A. Courville, J. Bergstra, and Y. Bengio, "A spike and slab restricted Boltzmann machine," in *JMLR W&CP : Proc. AISTATS'2011*, vol. 15, 2011. Recipient of People's Choice Award.

[118] Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cisse, M. Côté, D. Erhan, J. Eustache, X. Glorot, X. Muller, S. Pannetier Lebeuf, R. Pascanu, S. Rifai, F. Savard, and G. Sicard, "Deep learners benefit more from out-of-distribution examples," in *JMLR W&CP : Proc. AISTATS'2011*, 2011.

[119] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, "Theano : a CPU and GPU math expression compiler," in *Proc. SciPy*, 2010.

[120] M. Mandel, D. Eck, and Y. Bengio, "Learning tags that vary within a song," in *In Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, pp. 399–404, Aug. 2010.

[121] J. Turian, L. Ratinov, and Y. Bengio, "Word representations : A simple and general method for semi-supervised learning," in *Proc. ACL'2010*, pp. 384–394, 2010.

[122] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning ?," in *JMLR W&CP : Proc. AISTATS'2010*, vol. 9, pp. 201–208, 2010.

[123] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *JMLR W&CP : Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, vol. 9, pp. 249–256, May 2010.

[124] G. Desjardins, A. Courville, Y. Bengio, P. Vincent, and O. Delalleau, "Tempered

Markov chain Monte Carlo for training of restricted Boltzmann machine," in *AISTATS*, pp. 145–152, 2010.

[125] J. Bergstra and Y. Bengio, "Slow, decorrelated features for pretraining complex cell-like networks," in *NIPS'2009*, Dec. 2009.

[126] A. Courville, D. Eck, and Y. Bengio, "An infinite factor model hierarchy via a noisy-or mechanism," in *Neural Information Processing Systems Conference (NIPS) 22*, pp. 405–413, 2009.

[127] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *ICML'09*, 2009.

[128] D. Erhan, P.-A. Manzagol, Y. Bengio, S. Bengio, and P. Vincent, "The difficulty of training deep architectures and the effect of unsupervised pre-training," in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS 2009)*, pp. 153–160, Apr. 2009.

[129] J. Turian, J. Bergstra, and Y. Bengio, "Quadratic features and deep architectures for chunking," in *Proc. NAACL-HLT'2009*, pp. 245–248, 2009.

[130] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08)* (W. W. Cohen, A. McCallum, and S. T. Roweis, eds.), pp. 1096–1103, ACM, 2008.

[131] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," in *Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08)* (W. W. Cohen, A. McCallum, and S. T. Roweis, eds.), pp. 536–543, ACM, 2008.

[132] N. Le Roux, P.-A. Manzagol, and Y. Bengio, "Topmoumoute online natural gradient algorithm," in *Advances in Neural Information Processing Systems 20 (NIPS'07)* (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), pp. 849–856, Cambridge, MA : MIT Press, 2008.

[133] N. Le Roux, Y. Bengio, P. Lamblin, M. Joliveau, and B. Kégl, "Learning the 2-d topology of images," in *Advances in Neural Information Processing Systems 20 (NIPS'07)* (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), (Cambridge, MA), pp. 841–848, MIT Press, 2008.

[134] N. Chapados and Y. Bengio, "Augmented functional time series representation and forecasting with gaussian processes," in *Advances in Neural Information Processing Systems 20 (NIPS'07)* (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), (Cambridge, MA), pp. 265–272, MIT Press, 2008.

[135] H. Larochelle, D. Erhan, and Y. Bengio, "Zero-data learning of new tasks," in *AAAI Conference on Artificial Intelligence*, 2008.

[136] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proceedings of the 24th International Conference on Machine Learning (ICML'07)* (Z. Ghahramani, ed.), pp. 473–480, ACM, 2007.

[137] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems 19 (NIPS'06)* (B. Schölkopf, J. Platt, and T. Hoffman, eds.), pp. 153–160, MIT Press, 2007.

[138] N. Le Roux and Y. Bengio, "Continuous neural networks," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*

(AISTATS'07), (San Juan, Porto Rico), Omnipress, Mar. 2007.

[139] N. Chapados and Y. Bengio, "Forecasting commodity contract spreads with Gaussian process," in *13th International Conference on Computing in Economics and Finance*, June 2007.

[140] J. Carreau and Y. Bengio, "A hybrid Pareto model for conditional density estimation of asymmetric fat-tail data," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS'07)*, (San Juan, Porto Rico), Omnipress, Mar. 2007.

[141] Y. Bengio, N. Le Roux, P. Vincent, O. Delalleau, and P. Marcotte, "Convex neural networks," in *Advances in Neural Information Processing Systems 18 (NIPS'05)* (Y. Weiss, B. Schölkopf, and J. Platt, eds.), pp. 123–130, Cambridge, MA : MIT Press, 2006.

[142] Y. Bengio, O. Delalleau, and N. Le Roux, "The curse of highly variable functions for local kernel machines," in *Advances in Neural Information Processing Systems 18 (NIPS'05)* (Y. Weiss, B. Schölkopf, and J. Platt, eds.), pp. 107–114, Cambridge, MA : MIT Press, 2006.

[143] Y. Bengio, H. Larochelle, and P. Vincent, "Non-local manifold Parzen windows," in *Advances in Neural Information Processing Systems 18 (NIPS'05)* (Y. Weiss, B. Schölkopf, and J. Platt, eds.), pp. 115–122, MIT Press, 2006.

[144] N. Chapados and Y. Bengio, "The k best-paths approach to approximate dynamic programming with application to portfolio optimization," in *AI06*, pp. 491–502, 2006.

[145] Y. Grandvalet and Y. Bengio, "Semi-supervised Learning by Entropy Minimization," in *Advances in Neural Information Processing Systems 17 (NIPS'04)* (L. Saul, Y. Weiss, and L. Bottou, eds.), (Cambridge, MA), MIT Press, Dec. 2005.

[146] F. Rivest, Y. Bengio, and J. Kalaska, "Brain inspired reinforcement learning," in *Advances in Neural Information Processing Systems 17 (NIPS'04)* (L. Saul, Y. Weiss, and L. Bottou, eds.), (Cambridge, MA), pp. 1129–1136, MIT Press, Cambridge, 2005.

[147] Y. Bengio and M. Monperrus, "Non-local manifold tangent learning," in *Advances in Neural Information Processing Systems 17 (NIPS'04)* (L. Saul, Y. Weiss, and L. Bottou, eds.), pp. 129–136, MIT Press, 2005.

[148] O. Delalleau, Y. Bengio, and N. Le Roux, "Efficient non-parametric function induction in semi-supervised learning," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS'05)* (R. G. Cowell and Z. Ghahramani, eds.), pp. 96–103, Society for Artificial Intelligence and Statistics, Jan. 2005.

[149] M. Ouimet and Y. Bengio, "Greedy spectral embedding," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS'05)* (R. G. Cowell and Z. Ghahramani, eds.), pp. 253–260, 2005.

[150] F. Morin and Y. Bengio, "Hierarchical probabilistic neural network language model," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS'05)* (R. G. Cowell and Z. Ghahramani, eds.), pp. 246–252, 2005.

[151] I. Bhattacharya, L. Getoor, and Y. Bengio, "Unsupervised sense disambiguation using bilingual probabilistic models," in *Conference of the Association for Computational Linguistics (ACL'04)*, 2004.

[152] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering," in *Advances in Neural Information Processing Systems 16 (NIPS'03)*

(S. Thrun, L. Saul, and B. Schölkopf, eds.), MIT Press, 2004.

[153] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross-validation," in *Advances in Neural Information Processing Systems 16 (NIPS'03)* (S. Thrun, L. Saul, and B. Schölkopf, eds.), (Cambridge, MA), MIT Press, Cambridge, 2004.

[154] N. Boufaden, Y. Bengio, and G. Lapalme, "Approche statistique pour le rep∼A©rage de mots informatifs dans les textes oraux," in TALN'2004*, Traitement Automatique du Langage Naturel.*, 2004.

[155] Y. Bengio and J.-S. Sénécal, "Quick training of probabilistic neural nets by importance sampling," in *Proceedings of the conference on Artificial Intelligence and Statistics (AISTATS)*, 2003.

[156] P. Vincent and Y. Bengio, "Manifold parzen windows," in *Advances in Neural Information Processing Systems 15 (NIPS'02)* (S. Becker and S. Thrun, eds.), (Cambridge, MA), pp. 825–832, MIT Press, 2003.

[157] R. Collobert, Y. Bengio, and S. Bengio, "Scaling large learning problems with hard parallel mixtures," in *Pattern Recognition with Support Vector Machines* (S. Lee and A. Verri, eds.), vol. 2388 of Lecture Notes in Computer Science, pp. 8–23, Springer-Verlag, 2002.

[158] Y. Bengio, I. Takeuchi, and K. Kanamori, "The challenge of non-linear regression on large datasets with asymmetric heavy tails," in *Proceedings of 2002 Joint Statistical Meetings*, pp. 193–205, American Statistical Association publ., 2002.

[159] Y. Bengio and N. Chapados, "Metric-based model selection for time-series forecasting," in *Proc. of 2002 IEEE International Workshop on Neural Networks for Signal Processing*, pp. 13–24, IEEE Press, september 2002.

[160] P. Vincent and Y. Bengio, "K-local hyperplane and convex distance nearest neighbor algorithms," in *Advances in Neural Information Processing Systems 14 (NIPS'01)* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), (Cambridge, MA), pp. 985–992, MIT Press, 2002.

[161] R. Collobert, S. Bengio, and Y. Bengio, "A parallel mixture of SVMs for very large scale problems," in *Advances in Neural Information Processing Systems 14 (NIPS'01)* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), pp. 633–640, 2002.

[162] N. Chapados, Y. Bengio, P. Vincent, J. Ghosn, C. Dugas, I. Takeuchi, and L. Meng, "Estimating car insurance premia : a case study in high-dimensional data inference," in *Advances in Neural Information Processing Systems 14 (NIPS'01)* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), (Cambridge, MA), pp. 1369–1376, MIT Press, 2002.

[163] N. Boufaden, L. G., and Y. Bengio, "Topic segmentation : First stage of dialogue-based information extraction process," in *Proceedings of the Natural Language Pacific Rim Symposium, NLPRS-01*, 2001.

[164] Y. Bengio, R. Ducharme, and P. Vincent, "A neural probabilistic language model," in *Advances in Neural Information Processing Systems 13 (NIPS'00)* (T. Leen, T. Dietterich, and V. Tresp, eds.), pp. 932–938, MIT Press, 2001.

[165] C. Dugas, Y. Bengio, F. Bélisle, and C. Nadeau, "Incorporating second-order functional knowledge for better option pricing," in *Advances in Neural Information Processing Systems 13 (NIPS'00)* (T. Leen, T. Dietterich, and V. Tresp, eds.), pp. 472–478, MIT Press, 2001.

[166] Y. Bengio, "Probabilistic neural network models for sequential data," in *International Joint Conference on Neural Networks (IJCNN)*, vol. 5, pp. 79–84, 2000.

[167] Y. Bengio, "Continuous optimization of hyper-parameters," in *International Joint Conference on Neural Networks (IJCNN)*, vol. V, pp. 305–310, 2000.

[168] J. Ghosn and Y. Bengio, "Bias learning, knowledge sharing," in *International Joint Conference on Neural Networks (IJCNN)*, vol. I, pp. 9–14, 2000.

[169] P. Vincent and Y. Bengio, "A neural support vector network architecture with adaptive kernels," in *International Joint Conference on Neural Networks (IJCNN)*, vol. 5, pp. 5187–5192, 2000.

[170] N. Chapados and Y. Bengio, "VaR-based asset allocation using neural networks," in *Computational Finance 2000*, 2000.

[171] F. Gingras, Y. Bengio, and C. Nadeau, "On out-of-sample statistics for time-series," in *Computational Finance 2000*, 2000.

[172] Y. Bengio and S. Bengio, "Modeling high-dimensional discrete data with multi-layer neural networks," in *Advances in Neural Information Processing Systems 12 (NIPS'99)* (S. Solla, T. Leen, and K.-R. Müller, eds.), pp. 400–406, MIT Press, 2000.

[173] C. Nadeau and Y. Bengio, "Inference for the generalization error," in *Advances in Neural Information Processing Systems 12 (NIPS'99)* (S. Solla, T. Leen, and K.-R. Müller, eds.), pp. 307–313, MIT Press, 2000.

[174] S. Pigeon and Y. Bengio, "Binary pseudowavelets and application to bilevel image processing," in *Proceedings of the Data Compression Conference, DCC'1999*, 1999.

[175] Y. Bengio, S. Bengio, J. F. Isabelle, and Y. Singer, "Shared context probabilistic transducers," in *Advances in Neural Information Processing Systems 10 (NIPS'97)* (M. Jordan, M. Kearns, and S. Solla, eds.), pp. 409–415, MIT Press, 1998.

[176] M. Bonneville, J. Meunier, Y. Bengio, and J. Soucy, "Support vector machines for improving the classification of brain pet images," in *SPIE Medical Imaging 1998*, (San Diego), 1998.

[177] H. Schwenk and Y. Bengio, "Training methods for adaptive boosting of neural networks," in *Advances in Neural Information Processing Systems 10 (NIPS'97)* (M. Jordan, M. Kearns, and S. Solla, eds.), pp. 647–653, MIT Press, 1998.

[178] P. Haffner, L. Bottou, P. G. Howard, P. Simard, Y. Bengio, and Y. L. Cun, "Browsing through high quality document images with DjVu," in *Proceedings of the Advances in Digital Libraries Conference (ADL'98)*, (Washington, DC, USA), p. 309, IEEE Computer Society, 1998.

[179] L. Bottou, P. G. Howard, and Y. Bengio, "The Z-coder adaptive binary coder," in *Proceedings of the Conference on Data Compression (DCC'98)*, (Washington, DC, USA), p. 13, IEEE Computer Society, 1998.

[180] S. Pigeon and Y. Bengio, "A memory-efficient adaptive huffman coding algorithm for very large sets of symbols," in *Proceedings of the Conference on Data Compression (DCC'98)*, p. 568, 1998.

[181] Y. LeCun, L. Bottou, and Y. Bengio, "Reading checks with multilayer graph transformer networks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97)*, vol. 1, pp. 151–154, Apr 1997.

[182] M. Rahim, Y. Bengio, and Y. LeCun, "Discriminative feature and model design for automatic speech recognition," in *In Proc. of Eurospeech*, pp. 75–78, 1997.

[183] L. Bottou, Y. Bengio, and Y. LeCun, "Global training of document processing systems

using graph transformer networks," in *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR'97)*, (Puerto Rico), pp. 490–494, IEEE, 1997.

[184] H. Schwenk and Y. Bengio, "AdaBoosting neural networks : Application to on-line character recognition," in *International Conference on Artificial Neural Networks*, pp. 967–972, Springer Verlag, 1997.

[185] J. Ghosn and Y. Bengio, "Multi-task learning for stock selection," in *Advances in Neural Information Processing Systems 9 (NIPS'96)* (M. Mozer, M. Jordan, and T. Petsche, eds.), pp. 946–952, MIT Press, Cambridge, MA, 1997.

[186] Y. Bengio, "Training a neural network with a financial criterion rather then a prediction criterion," in *Proceedings of the Fourth International Conference on Neural Networks in the Capital Markets (NNCM-96)* (A. Weigend, Y. Abu-Mostafa, and A.-P. Regenes, eds.), pp. 433–443, World Scientific, 1997.

[187] S. Bengio and Y. Bengio, "An EM algorithm for asynchronous input/output hidden Markov models," in *International Conference On Neural Information Processing* (L. Xu, ed.), (Hong-Kong), pp. 328–334, 1996.

[188] Y. Bengio and F. Gingras, "Recurrent neural networks for missing or asynchronous data," in *Advances in Neural Information Processing Systems 8 (NIPS'95)* (D. S. Touretzky, M. Mozer, and M. Hasselmo, eds.), pp. 395–401, MIT Press, 1996.

[189] S. El Hihi and Y. Bengio, "Hierarchical recurrent neural networks for long-term dependencies," in *Advances in Neural Information Processing Systems 8 (NIPS'95)* (D. S. Touretzky, M. Mozer, and M. Hasselmo, eds.), MIT Press, 1996.

[190] Y. Bengio and P. Frasconi, "An input/output HMM architecture," in *Advances in Neural Information Processing Systems 7 (NIPS'94)* (G. Tesauro, D. S. Touretzky, and T. K. Leen, eds.), pp. 427–434, MIT Press, 1995.

[191] Y. Bengio and P. Frasconi, "Diffusion of credit in Markovian models," in *Advances in Neural Information Processing Systems 7 (NIPS'94)* (G. Tesauro, D. S. Touretzky, and T. K. Leen, eds.), pp. 553–560, MIT Press, 1995.

[192] L. Bottou and Y. Bengio, "Convergence properties of the K-means algorithm," in *Advances in Neural Information Processing Systems 7 (NIPS'94)* (G. Tesauro, D. S. Touretzky, and T. K. Leen, eds.), pp. 585–592, MIT Press, 1995.

[193] S. Bengio, Y. Bengio, and J. Cloutier, "Use of genetic programming for the search of a new learning rule for neural networks," in *Proceedings of the First IEEE Conference on Evolutionary Computation*, pp. 324–327 vol.1, Jun 1994.

[194] P. Frasconi and Y. Bengio, "An EM approach to grammatical inference : Input/output HMMs," in *International Conference on Pattern Recognition (ICPR'94)*, (Jerusalem 1994), pp. 289–294, 1994.

[195] Y. LeCun and Y. Bengio, "Word-level training of a handritten word recognizer based on convolutional neural networks," in *International Conference on Pattern Recognition (ICPR'94)* (IEEE, ed.), (Jerusalem 1994), 1994.

[196] Y. Bengio and Y. LeCun, "Word normalization for on-line handwritten word recognition," in *International Conference on Pattern Recognition (ICPR'94)*, pp. 409–413, 1994.

[197] Y. Bengio, Y. LeCun, and D. Henderson, "Globally trained handwritten word recognizer using spatial representation, space displacement neural networks and hidden Markov models," in *Advances in Neural Information Processing Systems 6 (NIPS'93)* (J. D. Cowan, G. Tesauro, and J. Alspector, eds.), pp. 937–944, MIT Press, 1994.

[198] Y. Bengio and P. Frasconi, "Credit assignment through time : Alternatives to backpropagation," in *Advances in Neural Information Processing Systems 6 (NIPS'93)* (J. D. Cowan, G. Tesauro, and J. Alspector, eds.), pp. 75–82, MIT Press, 1994.

[199] Y. LeCun, Y. Bengio, D. Henderson, and A. Weisbuch, "On-line handwriting recognition with neural networks : spatial representation versus temporal representation," in *Proceedings of the International Conference on Handwriting and Drawing*, 1993.

[200] Y. Bengio, P. Frasconi, M. Gori, and G. Soda, "Recurrent neural networks for adaptive temporal processing," in *Proc. of the 6th Italian Workshop on Neural Networks, WIRN-93* (E. Caianello, ed.), (Vietri, Italy), pp. 1183–1195, World Scientific Publ., 1993.

[201] S. Bengio, Y. Bengio, J. Cloutier, and J. Gecsei, "Generalization of a parametric learning rule," in *Proceedings of the International Conference on Artificial Neural Networks 1993* (S. Gielen and B. Kappen, eds.), (Amsterdam, The Netherlands), pp. 502–502, Springer-Verlag, 1993.

[202] Y. Bengio, P. Frasconi, and P. Simard, "The problem of learning long-term dependencies in recurrent networks," in *IEEE International Conference on Neural Networks*, (San Francisco), pp. 1183–1195, IEEE Press, 1993. (invited paper).

[203] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "Neural network - gaussian mixture hybrid for speech recognition or density estimation," in *Advances in Neural Information Processing Systems 4 (NIPS'91)* (J. E. Moody, S. J. Hanson, and R. P. Lipmann, eds.), (Denver, CO), pp. 175–182, Morgan Kaufmann, 1992.

[204] S. Bengio, Y. Bengio, J. Cloutier, and J. Gecsei, "Aspects théoriques de l'optimisation d'une règle d'apprentissage," in *Actes de la conférence Neuro-Nîmes 1992*, (Nîes, France), 1992.

[205] Y. Bengio, S. Bengio, J. Cloutier, and J. Gecsei, "On the optimization of a synaptic learning rule," in *in Conference on Optimality in Biological and Artificial Networks*, 1992.

[206] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "Global optimization of a neural network - hidden markov model hybrid," in *International Joint Conference on Neural Networks (IJCNN)*, vol. 2, pp. 789–794, 1991.

[207] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "A comparative study of hybrid acoustic phonetic decoders based on artificial neural networks," in *Proceedings of EuroSpeech'91*, 1991.

[208] Y. Bengio, S. Bengio, J. Cloutier, and J. Gecsei, "Learning a synaptic learning rule," in *International Joint Conference on Neural Networks (IJCNN)*, pp. II–A969, 1991.

[209] Y. Bengio, R. D. Mori, G. Flammia, and R. Kompe, "A comparative study on hybrid acoustic phonetic decoders based on artificial neural networks," in *Proceeding of EuroSpeech*, 1991.

[210] Y. Bengio, R. De Mori, and M. Gori, "Experiments on automatic speech recognition using bps," in *Parallel Architectures and Neural Networks* (E. Caianello, ed.), pp. 223–232, World Scientific Publ., 1990.

[211] Y. Bengio, R. Cardin, R. De Mori, and Y. Normandin, "A hybrid coder for hidden Markov models using a recurrent neural network," in *International Conference on Acoustics, Speech and Signal Processing*, (Albuquerque, NM), pp. 537–540, 1990.

[212] Y. Bengio, Y. Pouliot, S. Bengio, and P. Agin, "A neural network to detect homologies in proteins," in *Advances in Neural Information Processing Systems 2 (NIPS'89)*

(D. S. Touretzky, ed.), (Denver, CO), pp. 423–430, Morgan Kaufmann, 1990.

[213] Y. Bengio, R. Cardin, and R. De Mori, "Speaker independent speech recognition with neural networks and speech knowledge," in *Advances in Neural Information Processing Systems 2 (NIPS'89)* (D. S. Touretzky, ed.), (Denver, CO), pp. 218–225, Morgan Kaufmann, 1990.

[214] Y. Bengio, P. Cosi, R. Cardin, and R. D. Mori, "Use of multi-layered networks for coding speech with phonetic features," in *Advances in Neural Information Processing Systems 1 (NIPS'88)* (D. Touretzky, ed.), (Denver, CO), pp. 224–231, Morgan Kaufmann, San Mateo, 1989.

[215] R. De Mori, Y. Bengio, and P. Cosi, "On the generalization capability of multilayered networks in the extraction of speech properties," in *Proceedings of the International Joint Conference on Artificial Intelligence*, (Detroit), pp. 1531–1536, IEEE, 1989.

[216] M. Gori, Y. Bengio, and R. De Mori, "BPS : A learning algorithm for capturing the dynamical nature of speech," in *International Joint Conference on Neural Networks (IJCNN)*, (Washington D.C.), pp. 643–644, IEEE, New York, 1989.

[217] Y. Bengio, R. Cardin, P. Cosi, and R. De Mori, "Speech coding with multi-layer networks," in *International Conference on Acoustics, Speech and Signal Processing*, (Glasgow, Scotland), pp. 164–167, 1989.

[218] Y. Bengio and R. De Mori, "Use of neural networks for the recognition of place of articulation," in *International Conference on Acoustics, Speech and Signal Processing*, (New-York, NY), pp. 103–106, 1988.

2.3 Refereed Books and Book Chapters

[219] Y. Bengio, "Evolving culture vs local minima," in *Growing Adaptive Machines : Integrating Development and Learning in Artificial Neural Networks*, no. also as ArXiv 1203.2990v1, pp. T. Kowaliw, N. Bredeche & R. Doursat, eds., Springer-Verlag, Mar. 2013.

[220] Y. Bengio and A. Courville, "Deep learning of representations," in *Handbook on Neural Information Processing*, vol. 49, Springer : Berlin Heidelberg, 2013.

[221] Y. Bengio, *Learning deep architectures for AI*. Now Publishers, 2009.

[222] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," in *Large Scale Kernel Machines* (L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, eds.), MIT Press, 2007.

[223] Y. Grandvalet and Y. Bengio, "Entropy regularization," in *Semi-Supervised Learning* (O. Chapelle, B. Schölkopf, and A. Zien, eds.), pp. 151–168, MIT Press, 2006.

[224] Y. Bengio, O. Delalleau, and N. Le Roux, "Label propagation and quadratic criterion," in *Semi-Supervised Learning* (O. Chapelle, B. Schölkopf, and A. Zien, eds.), pp. 193–216, MIT Press, 2006.

[225] O. Delalleau, Y. Bengio, and N. Le Roux, "Large-scale algorithms," in *Semi-Supervised Learning* (O. Chapelle, B. Schölkopf, and A. Zien, eds.), pp. 333–341, MIT Press, 2006.

[226] Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, and M. Ouimet, "Spectral dimensionality reduction," in *Feature Extraction, Foundations and Applications* (I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, eds.), Springer, 2006.

[227] Y. Bengio and Y. Grandvalet, "Bias in estimating the variance of k-fold cross-validation," in *Statistical Modeling and Analysis for Complex Data Problem* (P. Duchesne and B. Remillard, eds.), pp. 75–95, Kluwer : Lawrence Erlbaum, 2004.

[228] C. Dugas, Y. Bengio, N. Chapados, P. Vincent, G. Denoncourt, and C. Fournier, "Statistical learning algorithms applied to automobile insurance ratemaking," in *Intelligent and Other Computational Techniques in Insurance : Theory and Applications* (L. Jain and A. Shapiro, eds.), World Scientific Publishing Company, 2004.

[229] E. Trentin, F. Brugnara, Y. Bengio, C. Furlanello, and R. D. Mori, "Statistical and neural network models for speech recognition," in *Connectionist Approaches to Clinical Problems in Speech and Language* (R. Daniloff, ed.), pp. 213–264, Lawrence Erlbaum, 2002.

[230] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Intelligent Signal Processing*, pp. 306–351, IEEE Press, 2001. chap. 9.

[231] J. Schmidhuber, S. Hochreiter, and Y. Bengio, "Evaluating benchmark problems by random guessing," in *Field Guide to Dynamical Recurrent Networks* (J. Kolen and S. Kremer, eds.), IEEE Press, 2001.

[232] S. Hochreiter, F. F. Informatik, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets : the difficulty of learning long-term dependencies," in *Field Guide to Dynamical Recurrent Networks* (J. Kolen and S. Kremer, eds.), IEEE Press, 2000.

[233] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, Contour and Grouping in Computer Vision*, pp. 319–345, Springer, 1999.

[234] Y. Bengio, *Neural Networks for Speech and Sequence Recognition*. London, UK : International Thompson Computer Press, 1996.

[235] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time-series," in *The Handbook of Brain Theory and Neural Networks* (M. A. Arbib, ed.), pp. 255–257, MIT Press, 1995.

[236] Y. LeCun and Y. Bengio, "Pattern recognition and neural networks," in *The Handbook of Brain Theory and Neural Networks* (M. A. Arbib, ed.), pp. 711–714, MIT Press, 1995.

[237] Y. Bengio, "Radial basis functions for speech recognition," in *Speech Recognition and Understanding : Recent Advances, Trends and Applications*, pp. 293–298, NATO Advanced Study Institute Series F : Computer and Systems Sciences, 1990.

[238] Y. Bengio and R. De Mori, "Speech coding with multilayer networks," in *Neurocomputing : Algorithms, Architectures and Applications* (F. Fogelman Soulie and J. Herault, eds.), pp. 207–216, NATO Advanced Study Institute Series F : Computer and Systems Sciences, 1990.

[239] R. De Mori, Y. Bengio, and P. Cosi, "On the use of an ear model and multi-layer networks for automatic speech recognition," in *Structural Pattern Analysis* (R. Mohr, T. Pavlidis, and A. Sanfelin, eds.), World Scientific, 1990.

[240] Y. Bengio and R. De Mori, "Connectionist models and their application to automatic speech recognition," in *Artificial Neural Networks and Statistical Pattern Recognition : Old and New Connections* (I. K. Sethi and A. K. Jain, eds.), pp. 175–192, Elsevier, Machine Intelligence and Pattern Recognition Series, 1990.

2.4 Patents

[241] Y. Bengio, L. Bottou, and P. G. Howard, "Z-coder : a fast adaptive binary arithmetic coder." U.S. Patent 6,188,334, February 13, 2001, along with patents 6,225,925, 6,281,817, and 6,476,740, 2001.

[242] Y. Bengio, L. Bottou, and Y. LeCun, "Module for constructing trainable modular network in which each module outputs and inputs data structured as a graph." U.S. Patent 6,128,606, October 3, 2000.

[243] Y. Bengio, Y. LeCun, C. Nohl, and C. Burges, "Visitor registration system using automatic handwriting recognition." Patent submitted in the U.S.A. in October 1994, submission number 1-16-18-1, 1994.