





















- [17] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, pp. 1798–1828, 2013.
- [18] Y. LeCun, “Generalization and network design strategies,” *Connectionism in perspective*, pp. 143–155, 1989.
- [19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning (adaptive computation and machine learning series)*. MIT Press, 2016.
- [20] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of machine learning research*, vol. 15, pp. 1929–1958. 2014.
- [21] Nair, Vinod, and Geoffrey E. Hinton. “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the International Conference on Machine Learning*, pp. 807–814. 2010.
- [22] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” *arXiv preprint arXiv:1607.01759*. 2016.
- [23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [24] T. H. Chang, Y. T. Sung, and Y. T. Lee, “A Chinese word segmentation and POS tagging system for readability research,” in *Proceedings of the Annual Meeting of the Society for Computers in Psychology*, 2012.
- [25] F. Chollet, “Keras: Deep learning library for theano and tensorflow. URL: <https://keras.io>.” 2015.