

## BIBLIOGRAFÍA / RECURSOS EN INTERNET

### Support Vector Machines

- Boser, B., Guyon, I., Vapnik, V. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, 1992
- Cortes, C., Vapnik, V. Support Vector Networks. Machine Learning, 20(3):273-297, 1995
- Hernández Orallo, J., Ramírez Quintana, M. J., Ferri Ramírez, C. Introducción a la Minería de Datos. Pearson - Prentice Hall, 2004
- James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning, pp. 351-372. Springer, 2013
- Joachims, T. Learning to Classify Text Using Support Vector Machines -- Methods, Theory and Algorithms. Kluwer Academic Publishers, 2002
- Joachims, T. Text Categorization with Support Vector Machines: Learning with many Relevant Features. In: Proceedings of the 10th European Conference on Machine Learning, pp. 137-142, 1998
- Vapnik, V. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995
- Vapnik, V. Statistical Learning Theory. Wiley, 1998
- Vapnik, V. An Overview of Statistical Learning Theory. IEEE Transactions on Neural Networks, Vol. 10, No. 5, pp. 988-999, 1999

### Redes Neuronales y Deep Learning

- Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C. A neural probabilistic language model. *Journal of machine learning research* 3 (Feb), 1137–1155, 2003
  - Disponible también en:  
<https://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf>
- Cho, K., Raiko, T., Ilin, A. Enhanced gradient for training restricted Boltzmann machines. *Neural Computation*, 25(3):805–831, 2013
  - Describe el Stochastic Gradient Descent, una de las variantes algorítmicas que más han disminuido el tiempo de entrenamiento de las NN
  - Disponible también en:  
<https://direct.mit.edu/neco/article/25/3/805/7857/Enhanced-Gradient-for-Training-Restricted>
- Chollet, F. Deep Learning with Python. Manning. Ed 1, 2017; Ed 2 prevista oct. 2021
  - Libro de referencia. François Chollet es el autor de Keras, capa que ofrece distintos niveles de abstracción sobre tensorflow, uno de los principales framework de Deep Learning.
- Glorot, X., Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. 2011
  - Artículo de referencia que demuestra cómo una correcta inicialización de los pesos mejora el entrenamiento de la NN.

- Disponible en: [https://www.researchgate.net/publication/215616968\\_Understanding\\_the\\_difficulty\\_of\\_training\\_deep\\_feedforward\\_neural\\_networks](https://www.researchgate.net/publication/215616968_Understanding_the_difficulty_of_training_deep_feedforward_neural_networks)
- Goodfellow, I. Tutorial: Generative Adversarial Networks. NIPS, 2016
  - Tutorial de las Redes Generativas Antagónicas o GAN escrito por el propio autor.
  - Disponible en: <https://dev.arxiv.org/abs/1701.00160v3>
- Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. MIT Press; 2016.
  - Libro de referencia. Liderado por Ian Goodfellow, conocido por ser el autor de las GAN
  - Disponible también en <http://www.deeplearningbook.org>
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. 2012
  - Describe el dropout, que permite disminuir el overfitting y mejorar el entrenamiento
  - Disponible en: <https://arxiv.org/pdf/1207.0580.pdf>
- Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. Proc. of the National Academy of Sciences, 79:2554–2558, 1982
  - Describe un primer modelo de Red Neuronal Recurrente
  - Disponible también en: <https://www.pnas.org/content/pnas/79/8/2554.full.pdf>
- Kingma, D. Ba, J. Adam: A method for stochastic optimization. ICLR, 2015
  - Describe el optimizador Adam, que es uno de los más utilizados actualmente
  - Disponible también en: <https://arxiv.org/pdf/1412.6980.pdf>
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1(4), 541–551, 1989
  - Artículo de referencia en el uso de CNN para el conjunto de datos MNIST
  - Disponible también en: <https://direct.mit.edu/neco/article/1/4/541/5515/Backpropagation-Applied-to-Handwritten-Zip-Code>
- McCulloch, W. S. Pitts, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133, 1943
  - Describe el primer modelo conceptual de Neurona Artificial y posibles conexiones desde el ámbito de la lógica proposicional
  - Disponible también en: <https://web.csulb.edu/~cwallis/382/readings/482/mcculloch.logical.calculus.ideas.1943.pdf>
- Minsky, M. Papert, S. *Perceptrons*. MIT Press, 1969
- Ng, A. Machine Learning. Coursera, 2012
  - MOOC de referencia en Redes Neuronales
  - Disponible en: <https://www.coursera.org/learn/machine-learning>
- Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review* 65(6), 386–408, 1958
  - Describe el modelo perceptron, evolución del modelo de Neurona Artificial

- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. Rumelhart, D. E. and McClelland, J. L. (Eds.), Parallel Distributed Processing, Vol. 2, 318–362. MIT Press, 1986
  - Describe el algoritmo de backpropagation para entrenar las NN
  - Disponible también en: [https://web.stanford.edu/class/psych209a/ReadingsByDate/02\\_06/PDPVolIIChapter8.pdf](https://web.stanford.edu/class/psych209a/ReadingsByDate/02_06/PDPVolIIChapter8.pdf)
- Schmidhuber, J. Deep Learning in Neural Networks: An Overview, 2014
  - Disponible en: <https://arxiv.org/abs/1404.7828>
- Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R. Dropout: a simple way to prevent neural networks from overfitting. JMLR 15(1), 1929–1958, 2014
  - Describe el dropout, que permite disminuir el overfitting y mejorar el entrenamiento
  - Disponible también en: <https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>
- Sutton, R. The bitter lesson, 2019
  - Este post de uno de los padres del Reinforcement Learning demuestra con diversos ejemplos (ajedrez, visión artificial, reconocimiento de voz) que la fuerza bruta de las redes neuronales se impone con frecuencia a modelos específicos.
  - Disponible en: <http://www.incompleteideas.net/InclIdeas/BitterLesson.html>
- Ujjwalkarn. A Quick Introduction to Neural Networks, 2016
  - Disponible en: <https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/>

## Procesado de imágenes

- Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification. Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume 2: 1237–1242, 2012
  - Disponible también en: <https://people.idsia.ch/~juergen/ijcai2011.pdf>
- Fukushima, K. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. Biological Cybernetics 36 (4): 193–202, 1980
  - El Neocognitron es un antecesor de las Redes Neuronales Convolucionales (CNN)
  - Disponible también en: <https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf>
- Hubel, D. H., Wiesel, T. N. Receptive fields of single neurons in the cat's striate cortex. J. Physiol., 148:574–591, 1959
  - Artículo de referencia. Describe un experimento que demostró que la visión de los mamíferos consta de células simples (sensibles a la orientación de bordes) y de complejas (más robustas frente a rotación y traslación).

- Disponible también en:  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1363130/pdf/jphysiol01298-0128.pdf>
- He, K., Zhang, X., Ren, S., Sun, J. Deep Residual Learning for Image Recognition, 2015
  - Artículo en que se define el modelo ResNet
  - Disponible en: <https://arxiv.org/pdf/1512.03385.pdf>
- Krizhevsky, A., Sutskever, I., Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks, 2012
  - Artículo de referencia que describe el modelo de Deep Learning AlexNet que ganó la competición ImageNet Large-Scale Visual Recognition Challenge de 2012 (ILSVRC12). Es considerado como la singularidad de las NN, momento en que superaron a otros modelos de Machine Learning
  - Disponible en:  
[https://www.cs.toronto.edu/~kriz/imagenet\\_classification\\_with\\_deep\\_convolutional.pdf](https://www.cs.toronto.edu/~kriz/imagenet_classification_with_deep_convolutional.pdf)
- Lowe, D. G. Three-Dimensional Object Recognition from Single Two-Dimensional Images. Artificial Intelligence, 31, 3 (March 1987), pp. 355–395, 1987
  - Disponible también en: <https://www.cs.ubc.ca/~lowe/papers/aij87.pdf>
- Ujjwalkarn. An Intuitive Explanation of Convolutional Neural Networks, 2016
  - Disponible en: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
- Zeiler, M., Fergus, R. Visualizing and Understanding Convolutional Networks, 2013
  - Disponible en: <https://arxiv.org/abs/1311.2901>

## Procesado de lenguaje natural

- Bahdanau, D., Cho, K., Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. ICLR, 2015
  - Artículo de referencia en modelos de atención con LSTM
  - Disponible en: <https://arxiv.org/pdf/1409.0473.pdf>
- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623), 2021
- Bloem, P. Transformers from scratch, 2019
  - Disponible en: <http://peterbloem.nl/blog/transformers>
- Britz, D. Understanding Convolutional Neural Networks for NLP, 2015
  - Disponible en: <http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/>
- Brownlee, J. Long Short-Term Memory Networks With Python, 2017
  - Disponible en (previo pago): <https://machinelearningmastery.com/lstms-with-python/>
- Cho, K., Merriënboer, B., Culcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, 2014
  - Disponible en: <https://arxiv.org/pdf/1406.1078v3.pdf>

- Devlin, J., Chang, M. W., Lee, K., Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018
  - Disponible en <https://arxiv.org/pdf/1810.04805.pdf>
- Dimson, T. This word does not exist.
  - Generador de palabras y de su significado, basado en GPT-2.
  - Disponible en: <https://www.thisworddoesnotexist.com>
- Hinton, G., Vinyals, O., Dean, J. (2015). Distilling the knowledge in a neural network, 2015
  - Disponible en: <https://arxiv.org/pdf/1503.02531.pdf>
- Hochreiter, S. Schmidhuber, J. Long Short-Term Memory. Neural Computation, 9(8):1735–1780, 1997
  - Artículo de referencia, en gran parte responsable de la irrupción del Deep Learning en el Procesado de Lenguaje Natural y en otras áreas
  - Disponible también en: [https://www.researchgate.net/publication/13853244\\_Long\\_Short-term\\_Memory](https://www.researchgate.net/publication/13853244_Long_Short-term_Memory)
- Huggingface. Transformers.
  - Web de referencia en Transformers
  - Disponible en: <https://github.com/huggingface/transformers>
- Jurafsky, D., Martin, J. Speech and Language Processing, 2021 (previsto)
  - Libro de referencia en procesado de Voz y Lenguaje Natural
  - Disponible en: <https://web.stanford.edu/~jurafsky/slp3/>
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, 2019
  - Disponible en <https://arxiv.org/pdf/1910.13461.pdf>
- Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., Potts, C. Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies (pp. 142-150), 2011
- Mikolov, T., Chen, K., Corrado, G., Dean, J. Efficient Estimation of Word Representations in Vector Space, 2013
  - Artículo de referencia. Describe word2vec, que es uno de los modelos responsables de la expansión del Deep Learning en el NLP
  - Disponible en: <https://arxiv.org/abs/1301.3781>
- Mikolov, T., Le, Q. V., Sutskever, I. Exploiting similarities among languages for machine translation, 2013
  - Disponible en: <https://arxiv.org/pdf/1309.4168.pdf>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. Distributed representations of words and phrases and their compositionality, 2013
  - Disponible en: <https://arxiv.org/pdf/1310.4546.pdf>
- Olah, C. Understanding LSTM Networks, 2015
  - Tutorial de referencia para comprender cómo funcionan las LSTM
  - Disponible en: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. Language models are unsupervised multitask learners. OpenAI blog, 2019

- Disponible en: [https://d4mucfpksywv.cloudfront.net/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf)
- Salton, G., McGill, MJ. Introduction to modern information retrieval. McGraw-Hill, 1986
- Schick, T., Schütze, H. Exploiting cloze questions for few-shot text classification and natural language inference, 2020
  - Disponible en: <https://arxiv.org/pdf/2001.07676.pdf>
- Schuster, M., Paliwal, K. K. Bidirectional recurrent neural networks. IEEE transactions on Signal Processing, 45(11), 2673-2681, 1997
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I. Attention is all you need. NIPS, 2017
  - Artículo de referencia en modelos de atención (sin LSTM) – Transformers
  - Disponible en: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- Yin, W., Hay, J., Roth, D. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach, 2019
  - Disponible en <https://arxiv.org/pdf/1909.00161.pdf>
- Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., Fidler, S. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In Proceedings of the IEEE international; 2015
  - Disponible en: <https://arxiv.org/pdf/1506.06724.pdf>

## MATERIAL COMPLEMENTARIO

- Fei-Fei Li. ImageNet.
  - Dataset de referencia; fue concebido para servir de banco de pruebas en modelos de clasificación de imágenes. Su versión clásica consta de mil clases y un millón de imágenes
  - Disponible en: <https://image-net.org>
- Russell, S. Norvig, P. Artificial Intelligence: A Modern Approach (2nd Ed.). Prentice Hall, 2002