

UNIVERSIDAD LOYOLA ANDALUCÍA



TESIS DOCTORAL

“Diversidad explícita en modelos de ensembles de Extreme Learning Machine”

Doctorando/a: Carlos Perales González

Directores: David Becerra Alonso
Francisco Fernández Navarro
Mariano Carbonero Ruz

Tutor/a de Doctorado: David Becerra Alonso

Sevilla, 2021



UNIVERSIDAD LOYOLA ANDALUCÍA

Tesis Doctoral

Diversidad explícita en modelos de ensembles de
Extreme Learning Machine

Doctorando en Ciencia de los Datos

Doctorando/a:

Carlos Perales González

Directores:

David Becerra Alonso

Francisco Fernández Navarro

Mariano Carbonero Ruz

Tutor/a:

David Becerra Alonso

Sevilla, 2021

Agradecimientos

A pesar de tener pasión por la Ciencia, a pesar de querer formarme como investigador, el esfuerzo y la determinación de empezar y terminar esta tesis doctoral no habrían suficientes para llegar a buen puerto sin las personas que han marcado cada día, mes y año de este largo proceso.

Gracias a mi familia, a mis padres Maria José y Aquilino, y a mi hermano Joel, por apoyarme tanto en este tan largo camino. Por escucharme y por aconsejarme. Por demostrarme que puedo vivir en Sevilla, y tener un hogar en Córdoba. También quiero agradecer a mis amigos cordobeses, quienes me recuerdan quién soy y donde he llegado por el momento. Jose, Felipe, Carlos, Fran, Ricardo, Cristian, Rafa, Chema, Gonzalo, Sara, Marina, Miriam, Álvaro, y todos los demás, gracias por seguir ahí. Con vosotros, Córdoba es y será siempre mi hogar.

Dentro de la Universidad, hay compañeros de profesión que ahora puedo llamar amigos. Kisko, Carlos, Antonio Durán, Beatriz, Jorge, Carmelina, Álvaro, Alejandro, y muchos otros. De entre todos, quiero destacar especialmente a Javier Pérez, cuyos consejos fueron y todavía son para mí un faro en los momentos de mayor incertidumbre.

Apenas había visitado Sevilla antes de empezar el doctorado. Quién me iba a decir que me iba a encantar vivir aquí. En gran parte, gracias a la gente tan maravillosa que he conocido. Semana a semana, mes a mes, año a año, hemos creado vínculos que nos han permitido infundirnos felicidad mutuamente. Gracias a vosotros, Victoria, Pedro, Javier, Irene, Aurora, Laura, Ana, Elisa, Nuria, Cristina, Jorge, Carlos, todos. Sevilla brilla más con vosotros.

A todos vosotros, una vez más, muchas gracias.

Resumen

Extreme Learning Machine (ELM) ha mostrado ser un rápido algoritmo de aprendizaje automático, adecuado para problemas de regresión y clasificación. Con el fin de generalizar los resultados del ELM estándar, varios métodos de *ensemble* han sido desarrollados. Estos métodos de *ensemble* son meta-algoritmos que generalizan los resultados de los ELMs, generando varios predictores base cuyas predicciones se combinan en una predicción de conjunto final.

La mayoría de estos métodos confían en el muestro de datos para generar predictores diferentes y conseguir así la generalización de los resultados. Estos métodos tienen como hipótesis que los datos de entrenamiento son suficientemente heterogéneos para que los predictores generados sean diversos entre sí. En esta tesis, se proponen métodos de *ensemble* que promueven la diversidad explícitamente, evitando la hipótesis de que los datos deben de muestrearse de manera diversa. Esta promoción de la diversidad se realiza a través de las funciones objetivo de los ELMs, usando ideas del entorno de trabajo de *Negative Correlation Learning* (NCL).

La formulación de la diversidad a través de la función objetivo de ELM permite desarrollar una solución analítica para los parámetros de los ELMs base. Esto reduce significativamente el coste computacional, comparado con la versión clásica de NCL para redes neuronales artificiales. De manera adicional, los métodos *ensemble* propuestos han sido validado mediante estudios experimentales con conjuntos de datos de *benchmark*, comparando con métodos *ensemble* existentes en la literatura ELM.

Abstract

Extreme Learning Machine (ELM) has shown to be a speedy and suitable algorithm for both regression and classification problems. To generalize the results of standard ELM, several ensemble methods have been implemented. These ensemble methods are meta-algorithms that generalize the results of ELMs by generating several base learners whose predictions are combined in a final prediction.

Most of these methods rely on data sampling procedures for achieving generalization, under the assumption that training data is heterogeneously enough to set up base learners diverse among them. In this research, ensemble learning methods that promote diversity explicitly are proposed. This avoids the hypothesis that the data must be heterogeneously enough to generate diverse base learners. This diversity promotion was made using ideas from the Negative Correlation Learning (NCL) framework.

The error function formulation allows us to develop an analytical solution to the parameters of the ELM base learners, which significantly reduces the computational burden of the classical NCL ensemble method for Neural Networks. Additionally, the proposed ensemble methods have been validated by an experimental study with a variety of benchmark datasets, comparing it with existing ensemble methods in ELM.

Índice general

I	Disertación	
1.	Introducción	1
2.	Objetivos	3
3.	Conclusiones	5
II	Artículos científicos publicados	9
1.	Regularized ensemble neural networks models in the Extreme Learning Machine framework	10
2.	Negative Correlation Learning in the Extreme Learning Machine framework	27
3.	A unified framework for global optimization of ensemble models	47

Parte I
Disertación

Capítulo 1

Introducción

Durante los últimos años, *Extreme Learning Machine* (ELM) [24] ha demostrado ser un algoritmo de aprendizaje automático eficiente, considerado entre los mejores predictores del estado del arte [17], junto con SVM [2, 13, 22, 40, 39] y *Random Forests* [29, 23, 4, 5, 43]. Ha sido usado ampliamente, no solo para problemas tradicionales de aprendizaje automático supervisado [24], sino también para problemas como series temporales [44, 54], clasificación de imágenes [10] y reconocimiento de voz [52].

Tanto la versión de red neuronal con una capa oculta (*Single-hidden Layer Feedforward Network* o SLFN) [26] y la versión usando *kernel trick* [24] son ampliamente usadas en este tipo de problemas de aprendizaje automático, debido a su potente capacidad de mapeo no lineal [15]. La versión neuronal del entorno de trabajo de ELM se apoya en la aleatoriedad de los pesos de las neuronas de la capa oculta para permitir una optimización más rápida, solo optimizando la última capa, sin perder su capacidad de representación no lineal [26, 25]. Esto permite procesar datasets muy pesados en tiempos computacionales razonables [21, 30], así como facilita el uso de ELMs como parte de estructuras más grandes como los *ensembles* [44, 32, 47, 41, 49, 53].

Los meta-algoritmos conocidos como métodos de *ensemble* son usados para generalizar los resultados de mezclas de clasificadores o de regresores, conocidos como predictores base en este entorno de trabajo de *ensemble learning*. Puesto que la mejora del rendimiento va asociada a la diversidad entre los predictores base [8, 14], las metodologías de *ensemble* más sencillas generalmente buscan entrenar cada predictor base de manera separada y posteriormente combinando su salida mediante pesos. Existen diversos métodos de votos que han sido presentados para mejorar la eficacia [3, 19, 42, 9, 45]. En este campo, las metodologías de *ensemble* conocidas como *Bagging* [3] y *Boosting* [19] son las aproximaciones más usadas [16, 1, 51], principalmente por su fácil implementación y por su balance compensado entre la diversidad entre los predictores base y el rendimiento del *ensemble*. La clave en estas metodologías de *ensemble* reside en los datos de entrenamiento para generar diversidad. De esta manera, las soluciones diversas a los problemas de optimización asociados a los predictores base son buscadas implícitamente a través del muestreo de datos. Así que, implícitamente, se asume que el muestreo de los datos de entrenamiento es suficiente para generar diversidad entre los predictores base [31].

Sin embargo, si los datos de entrenamiento son muy homogéneos, no se pueden elegir subconjuntos suficientemente diferentes para que los predictores generados con ellos sean diversos. Existen otras maneras de generar predictores base, como modificando las capa-

tidades no lineales de los predictores a través de kernels [28, 48] o con sistemas híbridos, con predictores base de familias de modelos diferentes [50, 20]. Sin embargo, estas metodologías de *ensemble* arrastran el fallo del muestreo de datos para generar predictores: fallan en cuantificar la diversidad entre los predictores base que conforman el *ensemble*.

Motivados por esta peculiaridad de los *ensembles* de promoción de la diversidad implícita, Yao et al. [34, 35, 33] y Brown et al. [6, 7, 8] propusieron una novedosa metodología de *ensemble* llamada *Negative Correlation Learning* (NCL). Esta metodología promueve explícitamente la diversidad, incluyéndola en las funciones objetivo que se usan para optimizar los parámetros de los predictores. En la concepción original de NCL, la función objetivo de cada predictor base del *ensemble* está compuesta por el error cuadrático medio del modelo con respecto a la salida esperada, y un término de penalización asociado a la diversidad del predictor con respecto al *ensemble*.

El entorno de trabajo NCL está inspirado en la investigación de Perrone et al. [38], que describe que el error en la predicción (sesgo) y la diversidad entre los predictores base (varianza) son dos objetivos diferentes que se enfrentan cuando se eligen los parámetros de un predictor base. Además de tener en cuenta el sesgo y la varianza de cada predictor base, la novedad que plantea el entorno de trabajo de NCL es que el error de generalización del *ensemble* también depende de la covarianza entre los predictores [46, 49]. Huanhuan Chen y Xin Yao encontraron que la optimización de NCL es propenso al sobre-ajuste del *ensemble* al ruido en el conjunto de entrenamiento, independientemente del ajuste de los hiperparámetros del modelo. Esto les llevó a proponer la versión regularizada del *ensemble*, *Regularized Negative Correlation Learning* (RNCL) [27]. Además, una versión multi-objetivo fue propuesta [12], que optimizaba simultáneamente tres componentes de la función objetivo: error, diversidad NCL y regularización.

Aunque al principio el entorno de trabajo de NCL fue concebido como una aproximación para problemas de regresión [34, 46], posteriormente esa idea fue adaptada a problemas de clasificación binaria [36], problemas multi-clase [49], problemas de regresión ordinal [18] y semi-supervisados [11]. En su mayoría, estas metodologías de *ensemble* son aplicadas a predictores base que requieren de optimizaciones iteradas (descenso por gradiente) del *ensemble*. Para evitar esto, Wang et al. [49] propuso un término de penalización basado en NCL junto con metodología AdaBoost, haciendo esta propuesta un sistema híbrido entre aproximaciones de muestreo de datos y *Negative Correlation Learning*.

Capítulo 2

Objetivos

El objetivo de esta tesis es establecer una serie de propuestas para métodos de ensemble learning que promueven la diversidad a través de parámetros. Partiendo del entorno de trabajo *Negative Correlation Learning*, la diversidad se introduce en la función objetivo explícitamente. Se introduce en la función objetivo de ELM un término de diversidad, que se controla a través de un parámetro de penalización en la función objetivo.

En la primera propuesta, llamada *Regularized Ensemble for Extreme Learning Machine* (RE-ELM), se promueve en la función objetivo que los pesos de la capa de salida de cada ELM base sean ortogonales a los anteriores pesos. Esto influye en que la predicción de cada ELM sea diferente. Esta diversidad explícita es incluida como un término de diversidad basada en ortogonalidad, siendo esta una medida de la diversidad entre predictores cuantificable [37]. La optimización sucesiva de los problemas de ELM resultantes son simplificadas usando la fórmula de *Sherman-Morrison*, con lo que la optimización del ensemble se reduce al cálculo de una inversa matricial.

En la segunda propuesta, llamada *Negative Correlation for Extreme Learning Machine* (NC-ELM), varios ELMs son entrenados de manera separada en una primera iteración, usando diferentes pesos en la capa oculta obtenidos aleatoriamente. Posteriormente, los ELMs se entrenan en sucesivas iteraciones, en las que se introduce una penalización en la función objetivo de cada predictor base, basada en NCL. La diversidad entre los predictores bases es medida y promovida en esta propuesta a través del ángulo entre las salidas de cada ELM individual y las salidas del ensemble. Las salidas del ensemble son recalculadas después de cada iteración, y los términos ortogonales en las funciones objetivos son actualizados. Por último, el número de inversas matriciales a computar se reduce drásticamente, a través de una implementación de resolución de inversas inspirado en la fórmula de *Sherman-Morrison*. Concretamente, el coste computacional del método propuesto será similar a la resolución de los ELMs base que componen el ensemble.

Por último, la tercera propuesta está orientada a un nuevo entorno de trabajo para ensembles, que se centra en la optimización de la predicción del ensemble de manera global, en lugar de las optimizaciones individuales de los predictores base. De esta manera, la función objetivo propuesta reduce el error de la salida del ensemble, y al mismo tiempo introduce diversidad entre los distintos predictores y la regularización de cada uno de ellos. Esta propuesta se presenta como un entorno de trabajo, que abre la puerta al desarrollo de nuevos métodos de ensemble con diferentes métricas de error (**fit**), diversidad (**div**) y regularización (**reg**). Además, una solución concreta para ELM es desarrolla-

da analíticamente, llamada *Global Negative Correlation for Extreme Learning Machine* (GNC-ELM).

Capítulo 3

Conclusiones

En esta tesis doctoral, se han propuesto aumentar significativamente la eficacia de los métodos de *Extreme Learning Machine* (ELM) a través de metodologías de ensemble. Los avances de esta tesis se pueden presentar como tres propuestas de ensembles, usando ELM como predictores base:

- La primera propuesta, *Regularized Ensemble for Extreme Learning Machine* (RE-ELM), muestra cómo se pueden generar ELMs que sirvan como predictores base de un método ensemble que se construye jerárquicamente. Partiendo de el ELM estándar, se añade la diversidad entre los pesos de la capa de salida como una perturbación, y se añade en la función objetivo a minimizar junto con el error de clasificación. Posteriormente, los predictores son combinados jerárquicamente, pesando progresivamente menos para añadir diversidad sin restar eficacia al algoritmo.

Como consecuencia de la generación jerárquica de predictores y de la propuesta ponderada de combinación de las salidas, este método de ensemble también puede ser entendido como un generador de predictores. Este resultado puede ser interesante para algunos problemas aplicados, donde la sensibilidad ante ciertas variables puede dar lugar a algunas soluciones aplicadas a nuestro problema, ya que cada predictor base del ensemble puede ser entendido como una solución en sí misma.

- La segunda propuesta, *Negative Correlation for Extreme Learning Machine*, (NC-ELM), introduce el entorno de trabajo de *Negative Correlation Learning* (NCL) en la comunidad ELM. En esta propuesta, se generan S ELMs inicialmente, tantos como predictores base vayan a componer el ensemble. Después, a sus funciones objetivos se incorpora un término de penalización inspirado en el entorno de trabajo NCL. El término de penalización que se propone promueve explícitamente la diversidad, mediante el ángulo entre las salidas de cada uno de los predictores base con respecto a las salidas del ensemble.

Adicionalmente, el coste computacional del método propuesto es similar a la resolución de S problemas de optimización de ELM independientes, puesto que los inverses están estimados a través de la fórmula de *Sherman-Morrison*.

- La tercera propuesta presenta un novedoso entorno de trabajo de ensemble learning para regresión y clasificación que se enfoca en la optimización global de la eficacia

del ensemble, en lugar de la minimización de los errores de sus predictores individuales. Esta propuesta se particulariza para Extreme Learning Machine, llamada *Global Negative Correlation for Extreme Learning Machine* (GNC-ELM). Las pruebas experimentales muestran que optimizar de manera global produce ensembles con una eficacia significativamente mejor, comparadas con las de métodos de ensemble con diversidad explícita encontrados en la literatura.

Los experimentos muestran que promover la diversidad explícitamente entre los predictores base generan ensembles con mejores rendimientos significativamente que aquellos que la promueven mediante muestreo de datos. En esta tesis está ordenada de manera que cada artículo científico soluciona las limitaciones de la propuesta anterior. La primera propuesta (RE-ELM) genera modelos perturbando la solución óptima de un ELM. La perturbación añadida a cada ELM que se añadía era mayor, por lo que no se pueden añadir muchos elementos al ensemble y cada añadido contribuía menos a la predicción final.

Para solucionar esto, la segunda propuesta (NC-ELM) entrena todos los predictores al mismo tiempo. Es decir, que no da más importancia a los distintos predictores base, puesto que todos los ELMs difieren inicialmente en el mapeo aleatorio, y posteriormente se diferencian de la predicción final mediante el término de diversidad añadido a su función objetivo. Sin embargo, se optimiza cada predictor por separado, por lo que se necesita además iterar la solución. Cómo se introducía la interacción entre los predictores en las funciones objetivos necesitaba de esta iteración de soluciones. La última propuesta (GNC-ELM) encuentra una solución óptima para todos los ELMs base del ensemble de manera global, de manera que se optimiza directamente, sin iteraciones. Además, se busca la minimización del error global de las predicciones del ensemble sobre los datos de entrenamiento, y no las salidas de cada uno de los predictores base.

Bibliografía

- [1] Eric Bauer and Ron Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning*, 36(1-2):105–139, 1999.
- [2] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.
- [3] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug 1996.
- [4] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [5] Leo Breiman et al. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231, 2001.
- [6] Gavin Brown and Jeremy Wyatt. Negative correlation learning and the ambiguity family of ensemble methods. In *4th International Workshop on Multiple Classifier Systems*, volume 2709, pages 266–275. Springer, 2003.
- [7] Gavin Brown, Jeremy Wyatt, Rachel Harris, and Xin Yao. Diversity creation methods: a survey and categorisation. *Information Fusion*, 6(1):5–20, 2005.
- [8] Gavin Brown, Jeremy L Wyatt, and Peter Tiño. Managing diversity in regression ensembles. *Journal of machine learning research*, 6(Sep):1621–1650, 2005.
- [9] Peter Bühlmann and Bin Yu. Boosting with the l2 loss: regression and classification. *Journal of the American Statistical Association*, 98(462):324–339, 2003.
- [10] Faxian Cao, Zhijing Yang, Jinchang Ren, Weizhao Chen, Guojun Han, and Yuzhen Shen. Local block multilayer sparse extreme learning machine for effective feature extraction and classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5580–5594, 2019.
- [11] Huanhuan Chen, Bingbing Jiang, and Xin Yao. Semisupervised Negative Correlation Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11):5366–5379, 2018.
- [12] Huanhuan Chen and Xin Yao. Multiobjective neural network ensembles based on regularized negative correlation learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(12):1738–1751, 2010.

- [13] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [14] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.
- [15] Shifei Ding, Han Zhao, Yanan Zhang, Xinzheng Xu, and Ru Nie. Extreme learning machine: algorithm, theory and applications. *Artificial Intelligence Review*, 44(1):103–115, 2015.
- [16] Pedro Domingos. Why does bagging work? a bayesian account and its implications. In *3rd International Conference on Knowledge Discovery and Data Mining*, pages 155–158. KDD, 1997.
- [17] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.*, 15(1):3133–3181, 2014.
- [18] Francisco Fernandez-Navarro, Pedro Antonio Gutierrez, Casar Hervas-Martinez, and Xin Yao. Negative correlation ensemble learning for ordinal regression. *IEEE transactions on neural networks and learning systems*, 24(11):1836–1849, 2013.
- [19] Yoav Freund and Robert E Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*. *journal of computer and system sciences*, 55, 1997.
- [20] Mustafa Göçken, Mehmet Özçalıcı, A. Boru, and Ayşe Tuğba Dosdoğru. Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection. *Neural Computing and Applications*, 31(2):577–592, 2019.
- [21] Qing He, Tianfeng Shang, Fuzhen Zhuang, and Zhongzhi Shi. Parallel extreme learning machine for regression based on mapreduce. *Neurocomputing*, 102:52–58, 2013.
- [22] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.
- [23] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.
- [24] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 42(2):513–29, 2012.
- [25] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 985–990. IEEE, 2004.

- [26] Guang Bin Huang, Qin Yu Zhu, and Chee Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501, 2006.
- [27] Huanhuan Chen and Xin Yao. Regularized Negative Correlation Learning for Neural Network Ensembles. *IEEE Transactions on Neural Networks*, 20(12):1962–1979, 2009.
- [28] Muhammad Aminul Islam, Derek T Anderson, John E Ball, and Nicholas H Younan. Fusion of diverse features and kernels using lp-norm based multiple kernel learning in hyperspectral image processing. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2016 8th Workshop on*, pages 1–5. IEEE, 2016.
- [29] Ho Tin Kam. Random decision forest. In *Proc. of the 3rd Int’l Conf. on Document Analysis and Recognition, Montreal, Canada, August*, pages 14–18, 1995.
- [30] Liyanaarachchi Lekamalage Chamara Kasun, Hongming Zhou, Guang-Bin Huang, and Chi Man Vong. Representational learning with elms for big data. 2013.
- [31] A. H.R Ko, R. Sabourin, L. E.S de Oliveira, and A. de Souza Britto Jr. The Implication of Data Diversity for a Classifier-free Ensemble Selection in Random Subspaces. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–5, 2008.
- [32] Nan Liu and Han Wang. Ensemble based extreme learning machine. *IEEE Signal Processing Letters*, 17(8):754–757, 2010.
- [33] Yong Liu and Xin Yao. Ensemble learning via negative correlation. *Neural networks*, 12(10):1399–1404, 1999.
- [34] Yong Liu and Xin Yao. Negatively correlated neural networks for classification. *Artificial Life and Robotics*, 3(4):255–259, Dec 1999.
- [35] Yong Liu and Xin Yao. Simultaneous training of negatively correlated neural networks in an ensemble. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(6):716–725, 1999.
- [36] Yong Liu, Xin Yao, and Tetsuya Higuchi. Evolutionary ensembles with negative correlation learning. *IEEE Transactions on Evolutionary Computation*, 4(4):380–387, 2000.
- [37] Carlos Perales-González, Mariano Carbonero-Ruz, David Becerra-Alonso, and Francisco Fernández-Navarro. A preliminary study of diversity in extreme learning machines ensembles. In *Lecture Notes in Computer Science*, pages 302–314. Springer International Publishing, 2018.
- [38] Michael Perrone and Leon Cooper. When networks disagree: Ensemble methods for hybrid neural networks. Technical report, Brown University Providence, Institute for Brain and Neural Systems, 1992.

- [39] J Platt. Fast training of svms using sequential minimal optimization. *MIT Press: Cambridge*, pages 185–208, 1999.
- [40] John Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.
- [41] Annalisa Riccardi, Francisco Fernández-Navarro, and Sante Carloni. Cost-sensitive AdaBoost algorithm for ordinal regression based on extreme learning machine. *IEEE Transactions on Cybernetics*, 44(10):1898–1909, 2014.
- [42] Stefan Schaal and Christopher G Atkeson. From isolation to cooperation: An alternative view of a system of experts. In *Advances in Neural Information Processing Systems 8*, pages 605–611. NIPS, 1996.
- [43] Mark R Segal. Machine learning benchmarks and random forest regression. 2004.
- [44] Huixin Tian and Bo Meng. A new modeling method based on bagging elm for day-ahead electricity price prediction. In *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, pages 1076–1079. IEEE, 2010.
- [45] Gerhard Tutz and Harald Binder. Boosting ridge regression. *Computational Statistics and Data Analysis*, 51(12):6044–6059, 2007.
- [46] N. Ueda and R. Nakano. Generalization error of ensemble estimators. In *International Conference on Neural Networks*, pages 90–95. IEEE, IEEE, 1996.
- [47] Dianhui Wang and Monther Alhamdoosh. Evolutionary extreme learning machine ensembles with size control. *Neurocomputing*, 102:98–110, 2013.
- [48] Haifeng Wang, Bichen Zheng, Sang Won Yoon, and Hoo Sang Ko. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2):687–699, 2018.
- [49] Shuo Wang, Huanhuan Chen, and Xin Yao. Negative correlation learning for classification ensembles. *Proceedings of the International Joint Conference on Neural Networks*, 2010.
- [50] Michał Woźniak, Manuel Graña, and Emilio Corchado. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16:3–17, 2014.
- [51] Abraham J Wyner, Matthew Olson, Justin Bleich, and David Mease. Explaining the success of adaboost and random forests as interpolating classifiers. *Journal of Machine Learning Research*, 18(1):1558–1590, 2017.
- [52] Xinzhou Xu, Jun Deng, Eduardo Coutinho, Chen Wu, Li Zhao, and Björn W Schuller. Connecting subspace learning and extreme learning machine in speech emotion recognition. *IEEE Transactions on Multimedia*, 21(3):795–808, 2019.
- [53] Jun-hai Zhai, Hong-yu Xu, and Xi-zhao Wang. Dynamic ensemble extreme learning machine based on sample entropy. *Soft Computing*, 16(9):1493–1502, 2012.

- [54] W. Zhang, A. Xu, D. Ping, and M. Gao. An improved kernel-based incremental extreme learning machine with fixed budget for nonstationary time series prediction. *Neural Computing and Applications*, 31(3):637–652, 2019.

UNIVERSIDAD LOYOLA ANDALUCÍA



TESIS DOCTORAL

“Diversidad explícita en modelos de ensembles de Extreme Learning Machine”

Doctorando/a: Carlos Perales González

Directores: David Becerra Alonso
Francisco Fernández Navarro
Mariano Carbonero Ruz

Tutor/a de Doctorado: David Becerra Alonso

Sevilla, 2021