

# Hibridación de Algoritmos de Aprendizaje para Modelos Neurologísticos aplicados a la Clasificación de Cubiertas Vegetales

Pedro A. Gutiérrez<sup>1</sup>, César Hervás<sup>1</sup>, Juan C. Fernández<sup>1</sup>, José M. Peña-Barragán<sup>2</sup>, Montserrat Jurado-Expósito<sup>2</sup>, Francisca López Granados<sup>2</sup>

**Resumen**—Este trabajo aborda la resolución de un problema de clasificación de cubiertas vegetales mediante espectrorradiometría de campo, utilizando una metodología híbrida. Dicha metodología combina regresión logística y modelos evolutivos de redes neuronales de unidades de base radial de tipo Gaussiano, obteniendo lo que se denominan modelos neurologísticos. Para estimar los coeficientes de estos modelos se proponen dos etapas, en la primera se aborda aprendizaje de los radios y los centros de las funciones de base radial, entrenando los modelos de redes neuronales mediante computación evolutiva y, una vez estimados el número de funciones Gaussianas y los estadísticos de centralización y dispersión de estas funciones, se aplica el método de máxima verosimilitud en el espacio de características formado por las covariables iniciales y las nuevas funciones de base. Esta metodología híbrida en el diseño del modelo y en la estimación de los coeficientes se aplica a un problema complejo de clasificación de firmas espectrales para su utilización en teledetección con sensores remotos. Los resultados obtenidos con este modelo y esta metodología híbrida son comparativamente mejores que los obtenidos con una regresión logística estándar, tanto en porcentaje de patrones bien clasificados, como en la precisión obtenida para todas y cada una de las cuatro clases del problema de estudio. El número de coeficientes del modelo y de variables de entrada es muy ajustado, de forma tal que se pueden obviar regiones de la firma espectral, ahorrando tiempo y dinero en el proceso de obtención de los datos.

**Palabras clave**—Regresión logística multiclase, redes neuronales de base radial, algoritmo evolutivo, clasificación, teledetección.

## 1. INTRODUCCIÓN

Existen campos de investigación como medicina, microbiología, epidemiología y muchos otros, donde es muy importante predecir el resultado de una variable de respuesta, o de forma similar obtener la probabilidad de éxito en una variable aleatoria Multinomial, en función de un conjunto de factores explicativos o covariables. De esta forma, si se considera como un problema de aprendizaje supervisado de clasificación multiclase, la meta es aprender como distinguir ejemplos que pertenecen a una de entre  $K$  clases (caracterizadas por los sucesos  $Y = 0, \dots, Y = K - 1$ ) en función de los valores que toman  $k$  variables predictoras o covariables  $X_1, X_2, \dots, X_k$ .

El modelo de regresión logística multiclase ha sido ampliamente empleado como método de regresión en estadística y, recientemente, ha recibido una gran aceptación en la comunidad de "machine learning" debido a su cercana relación con las nuevas teorías asociadas a las máquinas de soporte vectorial SVM [1] y a las técnicas de remuestreo iterativo como puede ser AdaBoost [2]. La regresión logística es un modelo de regresión lineal generalizada que trata de predecir las probabilidades a posteriori de pertenencia de cada uno de los patrones de un conjunto de entrenamiento a uno de los  $K$  valores que toma la variable dependiente, utilizando relaciones lineales con las variables predictoras [3].

Además, en los últimos años se está haciendo hincapié en la comparación de los nuevos algoritmos de clasificación basados en modelos de redes neuronales y árboles de decisión con modelos clásicos estadísticos como el análisis discriminante lineal y cuadrático y los modelos de regresión logística estándar o con restricciones [4],[5].

El método habitual de estimación de los coeficientes de un modelo lineal de regresión logística es el de máxima verosimilitud. Para obtener el óptimo de la función de verosimilitud, se aplica con frecuencia una optimización local basada en un algoritmo iterativo de tipo Newton-Raphson o de mínimos cuadrados con reasignación de pesos (IRLS) [6]. Desde un punto de vista formal, la regresión logística es un procedimiento sencillo y útil, pero no siempre se verifica que las probabilidades de pertenencia a cada clase transformadas mediante una transformación logarítmica presenten una relación lineal, causa-efecto, sobre las covariables.

Una aproximación para evitar estas dificultades es aumentar o reemplazar el vector de entradas  $x$  añadiéndole variables adicionales, o funciones de base, las cuales son transformaciones del vector  $x$ . A continuación, se utilizaría el modelo lineal en este nuevo espacio de características de entrada. La ventaja de esta metodología es que una vez que se determine el número y la estructura de estas funciones de base, el modelo es lineal en estas nuevas variables y los coeficientes son fácilmente estimados. La hibridación de este tipo de modelos

<sup>1</sup> Departamento de Informática y Análisis Numérico, Universidad de Córdoba. E-mail: {chervas,i02gucep}@uco.es, fernandezcaballero@gmail.com

<sup>2</sup> Instituto de Agricultura Sostenible, CSIC, Córdoba. E-mail: pa2pebaj@uco.es, {fgranados,montse.jurado}@ias.csic.es

ha sido justificada teóricamente por Donoho [7], el cual demostró que cualquier función continua se puede descomponer en dos funciones mutuamente excluyentes, tales como una función de tipo "kernel" o radial y una de tipo cresta (basada en funciones de proyección).

Uno de los modelos de redes neuronales más utilizados en una gran variedad de dominios son las funciones de base de tipo ventana "kernel", también llamadas redes de base radial o redes RBF. Así, se han utilizado en aproximación de funciones [8], en clasificación [9], en la predicción de series temporales [10], en minería de datos, en procesamiento de señales, y en el modelado y control de sistemas no lineales [11]. Entre sus propiedades, cabe destacar que son aproximadores universales [12] y que pertenecen a la clase de redes linealmente parametrizadas, siendo la salida de la red una combinación lineal de las funciones de base con coeficientes a determinar. El modelo RBF utiliza, en general, hyper-esferas o hyper-elipsoides para acotar las clases en el espacio de características, lo que las diferencia de las redes de base sigmoide o redes MLP, las cuales determinan sus funciones discriminantes mediante hyper-planos. El modelo RBF se puede considerar como un procedimiento de promediado local que mejora notoriamente la capacidad de aproximación y el diseño de la arquitectura de un modelo de red [13]. Esta simplicidad en la arquitectura permite la aplicación de algoritmos rápidos y eficientes [14]. Sin embargo, solo son aplicadas con éxito en problemas de interpolación para determinados casos [15].

Uno de los problemas más importantes asociados a este tipo de red es su aprendizaje, es decir, la estimación de los coeficientes del modelo. Los coeficientes a determinar que incluyen tanto los vectores de los centros y los radios o desviaciones estándar de las distribuciones Gaussianas que forman las funciones de base radial, como los coeficientes del modelo lineal que conecta las salidas de los nodos de la capa oculta con los nodos de la capa de salida. Otra tarea muy importante en el diseño de un modelo de red RBF es la determinación del número de funciones de base o nodos de la capa oculta, puesto que este parámetro tiene una importancia vital en el rendimiento de la red [16,17]. Ambos problemas han sido tratados mediante una gran variedad de aproximaciones. Una de ellas es fijar "a priori" el número y la posición de las unidades de base radial. Esta determinación se puede hacer mediante algoritmos no supervisados de "clustering" o agrupación de patrones [18] o mediante esquemas de aprendizaje supervisado o procedimientos de crecimiento o poda. La poda se aplica tanto para nodos, como para conexiones [16]. También existen trabajos en los que se utilizan algoritmos evolutivos [17] o híbridos [19].

En este trabajo se propone un modelo neurologístico basado en la hibridación de un modelo lineal utilizando las covariables iniciales y añadiendo funciones de base (no lineales) de tipo radial Gaussiano. Estas funciones expresarán las posibles relaciones locales existentes entre las covariables. Desafortunadamente, en nuestra aproximación, no se puede garantizar la no existencia de óptimos locales en la superficie de log-verosimilitud, como ocurre en los modelos estándar de regresión logística. En efecto, las superficies de error asociadas a las redes RBF son no lineales y pueden presentar numerosos óptimos locales.

El aprendizaje de los vectores de centros y radios y del número de funciones de base se lleva a cabo en dos etapas. En la primera, se utiliza un algoritmo evolutivo para diseñar la estructura y determinar los coeficientes asociados a las funciones Gaussianas. La complejidad de la superficie de error del modelo propuesto justifica la utilización de un algoritmo evolutivo como parte del procedimiento de estimación de los coeficientes.

En una segunda etapa, se aplica un algoritmo de optimización de la función de máxima verosimilitud asociada al modelo ampliado de regresión logística. En concreto, una vez que se dispone del número y de los coeficientes de las funciones de base mediante el algoritmo evolutivo, se puede considerar el modelo como lineal tanto en estas nuevas variables como en las covariables iniciales y así obtener los estimadores de los coeficientes de regresión, mediante el procedimiento estándar de regresión logística. Finalmente, se aplica un método de selección de covariables en función de su capacidad para explicar la variable de respuesta. Controlando de esta forma el número de coeficientes del modelo final, se puede disminuir el riesgo de construir un modelo complejo que sobreaprenda los datos del conjunto de entrenamiento, con la consiguiente pérdida de capacidad de generalización sobre los datos del conjunto de test.

La metodología propuesta se ha aplicado a un problema real de agricultura de precisión. Actualmente, los agricultores olivareros sólo reciben subvenciones de la Unión Europea si no labran y mantienen entre las calles del cultivo cubiertas vegetales. Estas cubiertas aumentan la retención del agua de lluvia y evitan la erosión y la pérdida de suelo fértil. Las inspecciones de campo para controlar la percepción de ayudas se realiza con visitas sólo al 1% del total de fincas. Obviamente este método es inexacto. Para solventar estos problemas, se puede utilizar teledetección. La teledetección consiste en captar información de la superficie terrestre o la atmósfera sin entrar en contacto físico. La captura de la energía reflejada de variables agronómicas se suele realizar en las zonas del visible (azul, verde y rojo, de 400 a 700 nm) e

infrarrojo cercano (701-900 nm) del espectro electromagnético. Cada cubierta terrestre presenta una forma peculiar de reflejar energía que se conoce como *signatura* o *firma espectral*. En las investigaciones de teledetección agronómica es esencial conocer la potencialidad de clasificar la signatura espectral *in situ* de cada una de las cubiertas, utilizando diferentes longitudes de onda. A su vez, se ha comprobado que la firma espectral de cualquier especie vegetal varía según la fase fenológica en la que ha sido tomada [20,21].

Diferentes estudios estadísticos y de minería de datos se han desarrollado para determinar un subconjunto suficientemente estrecho de longitudes de onda sin que se pierda información relevante de la firma espectral. Por ejemplo, se han aplicado redes neuronales artificiales para discriminar diferentes concentraciones de nitrógeno en vegetales [22] o clasificar malas hierbas en diferentes campos de trigo [23]; métodos de análisis multivariante como componentes principales o análisis discriminante para identificar deficiencias nutricionales en olivares [24] o discriminar entre firmas espectrales para detectar las malas hierbas del trigo [25]. Para evaluar el rendimiento de nuestra metodología, se ha aplicado a un problema de clasificación de cubiertas vegetales en olivar ya que, para aplicaciones de teledetección, es necesario que existan diferencias espectrales entre ellas, el olivo y el suelo desnudo.

Por lo anterior, se propone seleccionar las longitudes de onda que permiten clasificar las firmas espectrales de suelo desnudo, olivar y cubiertas vegetales (vivas o inertes). Este primer paso nos permitiría en futuros trabajos estudiar el uso de imágenes remotas de tipo hiperespectral. A partir de estas firmas espectrales las mejores longitudes de ondas del espectro serán seleccionadas, a partir de las conexiones y los coeficientes de los diferentes modelos de clasificación presentados en este trabajo.

El trabajo está organizado de la siguiente forma: en la Sección II se incluye una breve introducción a la regresión logística y se presenta el modelo propuesto en profundidad. En la Sección III, la metodología de estimación de los coeficientes del modelo es descrita, la cual se basa en dos etapas, cada una de ellas con un método diferente de optimización. La Sección IV presenta los experimentos realizados y los resultados obtenidos para el problema de determinación de cubiertas vegetales, y en la sección VI se derivan las principales conclusiones alcanzadas.

## II. REGRESIÓN LOGÍSTICA CON FUNCIONES DE BASE DE TIPO RADIAL

Los métodos de regresión logística son procedimientos estadísticos muy usados para

modelar variables de respuesta discreta (binaria o múltiple y en escala nominal u ordinal). Sea  $Y \in \{0, \dots, K-1\}$ , para  $K > 1$ , una variable de respuesta categórica o nominal (si  $Y = 0$  entonces el valor de la clase es  $G = 1$ , y de forma similar si  $Y = K-1$  entonces la clase es  $G = K$ ), y sea  $\mathbf{X} \in \mathbb{R}^p$  el correspondiente vector de covariables o variables iniciales del problema de clasificación. Se dice que las variables  $(\mathbf{X}, Y)$  verifican un modelo de regresión logística nominal, si para cada  $k = 1, \dots, K-1$ ,  $(\mathbf{X}, Y)$  satisfacen la ecuación:

$$\log \frac{P(G = k | \mathbf{X} = \mathbf{x})}{1 - P(G = K | \mathbf{X} = \mathbf{x})} = f_k(\mathbf{x}, \boldsymbol{\beta}_k) = \boldsymbol{\beta}_k^T \mathbf{x} \quad (1)$$

donde  $\frac{P(G = k | \mathbf{X} = \mathbf{x})}{1 - P(G = K | \mathbf{X} = \mathbf{x})}$  se define como la

posibilidad de que la salida sea  $Y = K$ . Debido a la no linealidad del modelo, la estimación de los coeficientes deberá hacerse mediante un procedimiento iterativo. El modelo de regresión logística [26,27] tiene el objetivo de modelar las probabilidades "a posteriori" de pertenencia del patrón  $\mathbf{x} \in \mathbf{X}$  a las  $K$  clases mediante funciones lineales en  $\mathbf{x}$ , asegurando siempre que la suma de las probabilidades sea la unidad. El modelo se propone como un conjunto de distribuciones de probabilidad condicional, las cuales son obtenidas utilizando el conjunto de entrenamiento  $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}$  para  $i = 1, \dots, N$ , donde la probabilidad de pertenencia a la clase  $k$  viene dada en forma matricial como:

$$p_k = \frac{\exp(\boldsymbol{\beta}_k^T \mathbf{x})}{1 + \sum_{k=1}^{K-1} \exp(\boldsymbol{\beta}_k^T \mathbf{x})}, \quad (2)$$

$$p_K = \frac{1}{1 + \sum_{k=1}^{K-1} \exp(\boldsymbol{\beta}_k^T \mathbf{x})}$$

para  $k = 1, \dots, K-1$  y donde se define  $\boldsymbol{\beta}_k = (\beta_{k,0}, \beta_{k,1}, \dots, \beta_{k,p})$ , ya que se añade el valor  $x_0 = 1$  al vector de características  $\mathbf{x}$ . Los coeficientes de regresión  $\boldsymbol{\beta}_k$  se estiman a partir de los datos y se interpretan como la proporción del logaritmo de las posibilidades asociadas a cada clase  $k$ .

En este trabajo se propone la utilización de modelos de regresión logística, donde la función  $f(\mathbf{x}, \boldsymbol{\beta})$  se construye con una parte lineal y otra parte no lineal obtenida a partir de las covariables iniciales. Utilizaremos funciones Gaussianas de base radial de la forma:

$$B_j = B_j(\mathbf{x}, \mathbf{w}_j) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{2r_j^2}\right) \quad (3)$$

Estos modelos de redes de base radial se pueden representar como una red neuronal. La red tiene  $p$  entradas que representan las covariables del modelo,  $m$  nodos en la capa oculta, asociados a otras tantas funciones de base radial y  $K$  nodos en la capa de salida. La función de activación del  $j$ -ésimo nodo de la capa oculta viene dada por la ecuación (3), siendo  $\mathbf{w}_j = (w_{j0}, w_{j1}, \dots, w_{jp})$ , y  $\mathbf{c}_j = (w_{j1}, \dots, w_{jp})$  y  $r_j = w_{j0}$  los centros (medias) y radio (desviación estándar), respectivamente, de la  $j$ -ésima función Gaussiana. En la Fig. 1 se ha incluido el esquema general del modelo RBF.

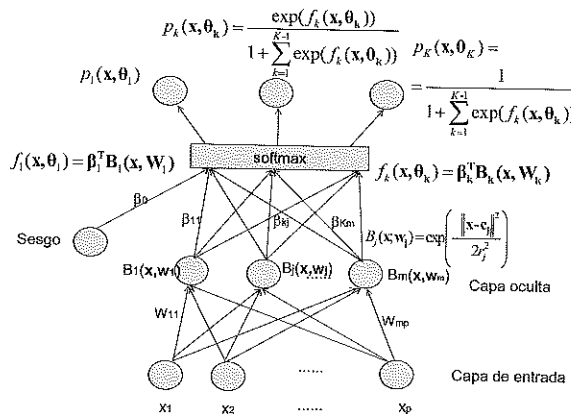


Fig. 1. Grafo del modelo de red RBF para K clases.

Utilizando este modelo, la parte lineal del modelo neurologístico propuesto en este trabajo, formada por las covariables iniciales y las funciones de base radial, NLIRUR, viene dada por:

$$f(\mathbf{x}, \theta) = \alpha^T \mathbf{x} + \beta^T \mathbf{B}(\mathbf{x}, \mathbf{W}) \quad (4)$$

donde  $\mathbf{x} = (1, x_1, \dots, x_p)$ ,  $\mathbf{B}(\mathbf{x}, \mathbf{W}) = (B_1(\mathbf{x}, \mathbf{w}_1), \dots, B_m(\mathbf{x}, \mathbf{w}_m))$  y  $B_j(\mathbf{x}, \mathbf{w}_j)$  viene definida en la ecuación (3).

De este modo, se define el vector de parámetros  $\theta = (\alpha, \beta, \mathbf{W})$ ,  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)$ ,  $\beta = (\beta_1, \dots, \beta_m)$  y  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_m)$ , siendo  $\mathbf{w}_j = (w_{j0}, w_{j1}, \dots, w_{jp})$  con  $w_{ji} \in \mathbb{R}$ . El nuevo conjunto de distribuciones condicionales de probabilidad es ahora:

$$p_k(\mathbf{x}, \theta_k) = \frac{\exp(\alpha_k^T \mathbf{x} + \beta_k^T \mathbf{B}(\mathbf{x}, \mathbf{W}))}{1 + \sum_{j=1}^{K-1} \exp(\alpha_j^T \mathbf{x} + \beta_j^T \mathbf{B}(\mathbf{x}, \mathbf{W}))} \quad (5)$$

$$p_K(\mathbf{x}, \theta_K) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(\alpha_j^T \mathbf{x} + \beta_j^T \mathbf{B}(\mathbf{x}, \mathbf{W}))}$$

para  $k=1, \dots, K-1$ . La función de transformación logit es:

$$\log \frac{P(G=k | \mathbf{X} = \mathbf{x})}{1 - P(G=K | \mathbf{X} = \mathbf{x})} = f_k(\mathbf{x}, \theta_k), \quad (6)$$

para  $k=1, \dots, K-1$ . Así se deriva una estructura de red más compleja, la cual se muestra en la figura 2, donde se ha considerado  $K=2$ , siendo:

$$f_k(\mathbf{x}, \theta_k) = \alpha_k^T \mathbf{x} + \beta_k^T \mathbf{B}(\mathbf{x}, \mathbf{W}) \quad (7)$$

El modelo neurologístico con unidades de base radiales se puede representar de forma equivalente a como se representa una estructura de grafo de tipo perceptrón, con una función de activación logística, como se observa en la Fig. 2 para el caso de clasificación binaria.

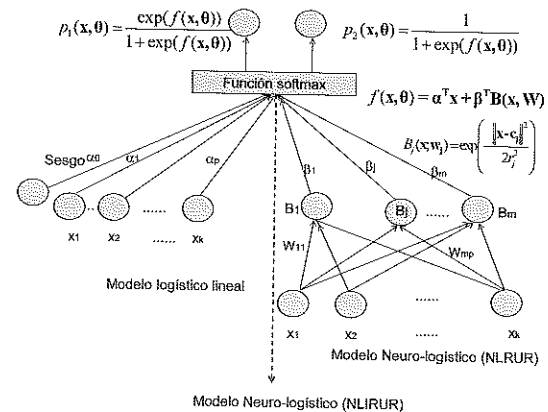


Fig. 2. Grafo del modelo NLIRUR para dos clases.

Si se dispone de un conjunto de datos de entrenamiento,  $D = \{(\mathbf{x}_i, \mathbf{y}_i)\}$  para  $i=1, \dots, N$ , donde  $x_i > 0, \forall i$ , el logaritmo de la función de verosimilitud para esas observaciones es:

$$L(\theta) = \sum_{k=1}^K \sum_{i=1}^{N_k} \left\{ \alpha_k^T \mathbf{x} + \beta_k^T \mathbf{B}(\mathbf{x}, \mathbf{W}) \right\} - \sum_{k=1}^K \sum_{i=1}^{N_k} \left\{ \log \left[ 1 + \sum_{k=1}^{K-1} \exp(\alpha_k^T \mathbf{x} + \beta_k^T \mathbf{B}(\mathbf{x}, \mathbf{W})) \right] \right\} \quad (8)$$

donde se ha considerado que  $f_k(\mathbf{x}, \theta_k) = 0$ , y  $N_k$  es el número de patrones de entrenamiento para la clase  $k$ . Entonces, la regla softmax de decisión para la clasificación es:

$$\text{Si } f_j(\mathbf{x}, \theta_j^*) > 0 \text{ y } f_j(\mathbf{x}, \theta_j^*) = \arg \max_k f_k(\mathbf{x}, \theta_k^*)$$

Entonces  $\mathbf{x} \in G_j$ , para  $j=1, \dots, K-1$

En otro caso  $\mathbf{x} \in G_K$ ,

$$\text{donde } \theta^* = \arg \max_{\theta} L(\theta)$$

La naturaleza y las propiedades de la función  $f_k(\mathbf{x}, \theta_k)$  implican que la matriz Hessiana asociada al procedimiento de maximización de (8) es semidefinida negativa [28], lo que hace que dicha

función sea cóncava sobre los parámetros  $\beta$ . La concavidad, junto con el hecho de que el vector de parámetros  $\beta$  varíe libremente sobre un conjunto convexo, garantiza que no existe un máximo local sobre la superficie del logaritmo de la función de verosimilitud en un modelo de regresión logística. Las condiciones bajo las cuales existe un máximo global, y las condiciones de convergencia de los estimadores obtenidos por el método de máxima verosimilitud se discuten en [28]. La estimación de los parámetros  $\beta$  habitualmente se realiza mediante un procedimiento iterativo como el algoritmo de Newton-Raphson o el algoritmo iterativo de mínimos cuadrados con reasignación de pesos (IRLS) [6]. Por lo general el algoritmo no converge dado que el logaritmo de la función de verosimilitud es cóncavo. En los casos poco probables de que la log-función de verosimilitud decrezca, el tamaño de la etapa garantiza la convergencia. En la siguiente sección se presenta el procedimiento de obtención de los estimadores  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{W})$  de los coeficientes del modelo.

### III. APRENDIZAJE DE LOS PARÁMETROS

La propuesta se basa en la combinación de un algoritmo evolutivo (explorador global) y de un procedimiento de optimización local (explotador local) llevado a cabo mediante un procedimiento de maximización de la función de verosimilitud. En una primera etapa, el proceso evolutivo determina el número  $m$  de funciones de tipo RBF y los vectores  $w_j$ , correspondientes a los centros y radios de cada una de estas  $m$  funciones. Una vez que las funciones base son determinadas mediante el algoritmo evolutivo,

$$B(x, W) = \{B_1(x, w_1), B_2(x, w_2), \dots, B_m(x, w_m)\},$$

se considera un aumento del espacio de entrada añadiendo a las covariables iniciales estas transformaciones no lineales de las covariables. De esta forma, el nuevo modelo es lineal en este nuevo espacio de variables, y los restantes coeficientes del modelo,  $\alpha$  y  $\beta$ , se calculan mediante un procedimiento de máxima verosimilitud condicional asociado a la regresión logística estándar. Por último, un procedimiento por etapas para eliminar variables de forma secuencial del modelo completo se aplica hasta que ninguna nueva eliminación mejore su capacidad de clasificación sobre el conjunto de entrenamiento.

A continuación los diferentes aspectos de esta metodología son mostrados con más detalle.

En la primera etapa se aplica un algoritmo evolutivo para encontrar una estimación de las funciones base de base radial,

$$B(x, \hat{W}) = \{B_1(x, \hat{w}_1), B_2(x, \hat{w}_2), \dots, B_m(x, \hat{w}_m)\},$$

correspondientes a la parte no lineal de la

función  $f(x, \theta)$ . Se debe determinar el número de funciones  $m$  y la matriz de pesos formada por los centros y radios estimados de las funciones Gaussianas  $\hat{W} = (\hat{w}_1, \hat{w}_2, \dots, \hat{w}_m)$ . Entre los diferentes paradigmas de la Computación Evolutiva, el elegido es la Programación Evolutiva, debido al hecho de que se evolucionan modelos cuya representación en grafo está asociada a un modelo de red neuronal artificial con números reales como pesos. El algoritmo evolutivo utilizado para diseñar la arquitectura del modelo y la estimación de los exponentes reales tiene puntos en común con otros algoritmos referenciados en la bibliografía [29],[30], y puede verse con detalle en [31], donde la única diferencia planteada en este trabajo está en la función de aptitud asociada a cada modelo, dado que se aborda un problema de clasificación. De esta forma, consideraremos la expresión de  $L(\theta^*)$  (ecuación (8)) como función de error de un individuo  $f(x, \theta^*)$  de la población. Por tanto, la medida de aptitud de un individuo es definida mediante una transformación estrictamente decreciente de la función de error  $L(\theta^*)$  en la forma

$$A(\theta^*) = \frac{1}{1 + L(\theta^*)}, \text{ donde } 0 < A(\theta^*) \leq 1.$$

La idea básica del algoritmo es la utilización de operadores de selección, replicación y mutación (paramétrica y estructural) en el proceso de evolución. El algoritmo se estructura en los siguientes pasos:

1. Generar la población inicial.
2. Repetir hasta que se cumpla la condición de parada:
  - Calcular la aptitud para cada individuo.
  - Ordenar de mayor a menor según la aptitud.
  - Copiar el mejor en la siguiente generación.
  - Replicar el  $r\%$  mejor de la población y sustituirlo por el  $r\%$  peor.
  - Aplicar mutación paramétrica al  $r\%$  mejor.
  - Aplicar mutación estructural al  $(100-r)\%$  restante.

Una vez que las  $m$  mejores funciones de base asociadas al individuo con una aptitud mejor en la última generación son determinadas, se considera una transformación del espacio de covariables, añadiéndoles las transformaciones no lineales del espacio de entrada, esto es, las funciones de base de dicho individuo:

$$H: R^k \rightarrow R^{k+m}$$

$$(x_1, x_2, \dots, x_k) \rightarrow (x_1, x_2, \dots, x_k, z_1, \dots, z_m)$$

donde  $z_1 = B_1(x, \hat{w}_1), \dots, z_m = B_m(x, \hat{w}_m)$ .

En la segunda etapa, el modelo de regresión logística estándar es aplicado a las variables  $x_1, x_2, \dots, x_k, z_1, \dots, z_m$  en el nuevo espacio de

características de entrada. De esta forma, el máximo de la función de verosimilitud condicional para  $N$  observaciones es obtenido. Para obtener el modelo final, se aplica un método de eliminación de variables no significativas del modelo basado en un procedimiento que elimina una variable en cada paso, empezando con el modelo completo y eliminando las variables secuencialmente hasta que la eliminación de una variable no mejore la capacidad de clasificación del modelo sobre el conjunto de entrenamiento.

#### IV. PARTE EXPERIMENTAL

La metodología híbrida propuesta en las secciones anteriores se ha aplicado para la resolución de un problema real de clasificación de cubiertas vegetales mediante espectrorradiometría de campo.

##### A. Diseño experimental

Los parámetros utilizados en el algoritmo evolutivo de la primera etapa son los siguientes: los centros  $w_{ji}$  se inicializaron en el intervalo  $[-2, 2]$ , los coeficientes  $\beta_j$  en el intervalo  $[-5, 5]$  y los radios en el intervalo  $[0, 1]$ . El máximo número de nodos en la capa oculta de los modelos RBF es  $m = 6$ ; el tamaño de la población es 1.000. El número de nodos que pueden ser añadidos o eliminados mediante una mutación estructural está entre 1 y 2; mientras que el número de conexiones que se pueden añadir o eliminar en una mutación estructural en cada nodo es un número entero del intervalo  $[1, 0.3l]$ , siendo  $l$  el número de conexiones del nodo. Las condiciones propuestas para el criterio de parada son: o bien que en 30 generaciones no se mejore ni el rendimiento medio del 20% de los mejores individuos de la población ni la aptitud del mejor individuo, o bien que el algoritmo alcance 100 generaciones.

El experimento de clasificación de las firmas espectrales en cuatro clases (cubiertas vivas, cubiertas inertes, suelo desnudo y olivar) ha sido realizado en Andalucía, España, en dos fincas llamadas Cortijo del Rey y Matallana a principios de primavera y verano, aunque en esta comunicación sólo se presentan los resultados obtenidos en Cortijo del Rey en primavera. La especie que forma la cubierta viva en esta finca es *Avena sativa*, especie de gramínea que presenta en primavera un color verde y formación de espigas. Las cubiertas inertes están formadas por los restos de hojas y ramas provenientes de la poda en primavera de los olivos. Se han tomado cuarenta firmas espectrales de un conjunto de cubiertas vivas, veinte firmas espectrales de un conjunto de cubiertas vegetales inertes, diez de suelo desnudo y diez de olivar. A partir de estas 80 firmas espectrales, se ha aplicado

un diseño de experimentos de tipo hold out estratificado tomando aproximadamente el 75% de las firmas en cada clase para formar el conjunto de entrenamiento (59 patrones) y el 25% restante para formar el conjunto de generalización (21 patrones). Las medidas se han obtenido utilizando un ASD Handheld FieldSpec Spectroradiometer (Analytical Spectral Devices, Inc., Boulder). El rango hiperespectral se consideró en principio entre 400 y 900 nanómetros, nm (con un ancho de banda de 5 nm). Estudios previos muestran que longitudes de onda no tan cercanas pueden dar similar información [20]. Así, en este estudio se considera un rango entre 575 y 725 nm longitudes de onda con un intervalo de 25 nm, con lo que el número de covariables del modelo es de 7. Este rango se ha obtenido mediante un análisis de componentes principales previo, donde se analizó qué longitudes de onda explicaban más varianza de las componentes principales.

##### B. Resultados

Para validar el rendimiento del modelo se utilizará la proporción de firmas espectrales correctamente clasificadas (CCR) para los conjuntos de entrenamiento y generalización. Hemos elegido los modelos de Análisis discriminante lineal con las covariables iniciales, DA, y con covariables iniciales y unidades de base radial, DAIRU, para así comparar los resultados con aquellos obtenidos utilizando otras metodologías. Para poder seleccionar las variables más significativas del modelo de regresión logística, tanto del modelo LR como del modelo NLIRUR, se aplica un método de eliminación de variables por pasos, utilizando el software SPSS 13.0 [32]. A partir de los resultados incluidos en la Tabla I, se concluye que los modelos NLIRUR producen clasificaciones más precisas (en cuanto a la proporción de patrones bien clasificados para el conjunto de generalización).

TABLA I  
CCR (%) DE LOS MEJORES MODELOS

ENTRENAMIENTO				
DA	DAIRU	RBF	LR	NLIRUR
96,6	88,14	88,14	100	100
GENERALIZACIÓN				
DA	DAIRU	RBF	LR	NLIRUR
90,50	90,50	85,91	95,24	100

Por otra parte, la Tabla II muestra los mejores modelos obtenidos con cada una de las metodologías RBF, LR y NLIRUR, modelos con un número de coeficientes muy ajustado, en los que la longitud de onda 600 nm fue siempre seleccionada; las longitudes de onda 575, 625, 675 y 725 nm se seleccionaron en RBF y NLIRUR; y las longitudes 650 y 700 nm sólo se seleccionó en uno de los modelos aplicados. Lo anterior indica que de 100 longitudes de onda tomadas



en cada cubierta en el espectro de 400 a 900 nm cada 5 nm, tan sólo hay 5 longitudes de onda (575, 600, 625, 675, y 725 nm) que conforman el mejor modelo NLIRUR, por lo que es posible configurar sensores remotos que, al menos en Cortijo 1 y en primavera, recojan sólo estas longitudes de onda. Ello redundará en el precio de las imágenes hiperespectrales que se adquieran. Por ejemplo, el sensor remoto CASI (Compact Airborne Spectral Imager) es capaz de proporcionar imágenes en estas 5 longitudes de onda programadas por el usuario. Por último indicar que de la observación de las matrices de contingencia asociadas a las tres metodologías (RBF, LR y NLIRUR) y presentadas en la Tabla III, se extrae que la mayor dificultad en el proceso de clasificación de las firmas espectrales se centra en la clase olivar (G=4) que a veces se confunde con la clase de cubiertas vivas (G=1).

TABLA III  
MATRICES DE CONTINGENCIA DE LOS MODELOS RBF,  
LR Y NLIRUR

OBS PRE	G=1	G=2	G=3	G=4	FC (%)
G=1	10,10, (10)	0,0,(0)	0,0,(0)	0,0,(0)	100,100, (100)
G=2	0,0, (0)	5,0,(0)	0,0,(0)	0,0,(0)	100,100, (100)
G=3	0,0, (0)	0,0,(0)	3,0,(0)	0,0,(0)	100,100, (100)
G=4	3,1, (0)	0,0,(0)	0,0,(0)	0,2,(3)	0,66,7, (66,7)
CCR					85,7, 95,2, (100)

## V. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo, se ha abordado un problema de clasificación multiclase de formas espectrales, para el que se proponen métodos de regresión logística basados en la combinación de un modelo lineal y un modelo no lineal formado por modelos de redes neuronales de unidades de base radial.

La metodología para el aprendizaje de los coeficientes del modelo se basa en la combinación en dos etapas de un algoritmo evolutivo que nos determine la estructura básica del modelo no lineal y de un procedimiento de optimización local basado en maximizar el logaritmo de la función de verosimilitud, que nos sirve para estimar los coeficientes de regresión del modelo final. Además, un método de selección de características es utilizado para determinar aquellas covariables que mejor expliquen la variable de respuesta. De esta forma, se controla el número de coeficientes del modelo final y se disminuye el riesgo de construir modelos complejos que sobre-aprendan los datos del conjunto de entrenamiento.

El modelo propuesto se ha aplicado a un problema real complejo, en el que se mejoran los

resultados obtenidos con otras metodologías. De esta forma, el modelo híbrido determina un buen balance entre considerar sólo un modelo lineal o sólo un modelo no lineal. Además, los resultados muestran que nuestra metodología es muy prometedora en términos de precisión en la clasificación. Los resultados obtenidos en la finca Cortijo 1 en Primavera no son extrapolables al verano, pero si pueden serlo a Matallana en primavera.

## AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por los proyectos TIN2005-08386-C05-02 (CICYT-FEDER), AGL2008-04670-CO3-03 (CICYT-FEDER) y P08-TIC-3745 (Junta de Andalucía). La investigación de P.A. Gutiérrez y J.C. Fernández ha sido financiada respectivamente por los programas predoctorales FPU (AP-2006-01746) y FPI (BES-2006-12543) del M<sup>o</sup> de Ciencia e Innovación.

## REFERENCIAS

- [1] Vapnik, "The nature of Statistical Learning Theory," Springer, 1999.
- [2] Y. Freund and R. Shapire, "Experiments with a new boosting algorithm," presented at Machine Learning: Proceedings of the Thirteenth International Conference, San Francisco, 1996.
- [3] P. McCullagh and J. A. Nelder, Generalized Linear Models, 2nd edn. London, 1989.
- [4] W. Vach, R. Robner, and M. Schumacher, "Neural Networks and logistic regression: Part II," Computational Statistics & Data Analysis, vol. 21, pp. 683-701, 1996.
- [5] M. Schumacher, R. Robner, and W. Vach, "Neural networks and logistic regression: Part I," Computational Statistics & Data Analysis, vol. 21, pp. 661-682, 1996.
- [6] T. Hastie, R. J. Tibshirani, and J. Friedman, "The Elements of Statistical Learning. Data mining, Inference and Prediction," in Springer, 2001.
- [7] D. Donoho, Projection based in approximation and a duality with kernel methods, Ann. Statist 17 (1989) 58-106.
- [8] L. Yingwei, N. Sundararajan, and P. Saratchandran, A sequential learning scheme for function approximation using minimum radial basis function neural networks, Neural Computation 9(9) (1997) 461-478.
- [9] Y.S. Hwang, and S.Y. Bang, An Efficient Method to Construct Radial Basis Function Neural Network Classifier, Neural Networks 10(8) (1997) 1495-1503.
- [10] B.A. Whitehead, and T.D. Choate, Cooperative-Competitive Genetic Evolution of Radial Basis Function Centers and Widths for Time Series Prediction, IEEE Trans. on Neural Networks 7(4) (1996) 869-880.
- [11] Y.J. Oyang, S.C. Hwang, Y.Y. Ou, C.Y. Chen, and Z.W. Chen, Data classification with radial basis function networks based on a novel kernel density estimation algorithm, IEEE Trans. Neural Networks 16(1) (2005) 225-236.
- [12] J. Park, and I.W. Sandberg, Universal Approximation Using Radial Basis Function Networks, Neural Computation 3(2) (1991) 246-257.
- [13] C.M. Bishop, Improving the generalization properties of radial basis function neural networks, Neural Computation 3(4) (1991) 579-581.
- [14] C. Panchapakesan, M. Palaniswami, D. Ralph, and C. Manzie, Effects of Moving the Centers in an RBF Network, IEEE Trans. On Neural Networks 13(6) (2002) 1299-1307.
- [15] D.S. Broomhead, and D. Lowe, Multivariable Functional Interpolation and Adaptive Networks, Complex Systems 2 (1988) 321-355.

[16] G.B. Huang, P. Saratchandran, and N. Sundararajan, A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation, *IEEE Trans. Neural Networks* 16(1) (2005) 57-67.

[17] F.H.F. Leung, H.K. Lam, S.H. Ling, and P.K.S. Tam, Tuning of the structure and parameters of a neural network using an improved genetic algorithm, *IEEE Trans. Neural Networks* 14(1) (2003) 79-88.

[18] C. Darken, and J. Moody, Fast adaptive K-means clustering: some empirical results, in: *Proceedings IEEE INNS International Joint Conference On Neural Networks*, 1990, pp. 233-238.

[19] Z.Q. Zhao, and D.S. Huang, A mended hybrid learning algorithm for radial basis function neural networks to improve generalization capability, *Applied Mathematical Modelling* 31 (2007) 1271-1281.

[20] J.M. Peña-Barragán, F. López-Granados, M. Jurado-Expósito, and L. García-Torres. Spectral discrimination of *Ridolfia segetum* and sunflower as affected by phenological stage. *Weed Research* 46 (2006) 10-21.

[21] K.S. Schmidt, and A.K. Skidmore. Spectral discrimination of vegetation types in coastal wetland. *Remote Sensing Environment* 85 (2003) 92-108.

[22] P.K. Goel, S.O. Prasher, R.M. Patel, J.A. Landry, R.B. Bonnel, and A.A. Viau. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture* 39 (2003) 67-93.

[23] F. López-Granados, J.M. Peña-Barragán, M. Jurado-Expósito, M. Francisco-Fernández, R. Cao, A. Alonso-Betanzos, O. Fontenla-Romero. Multispectral classification of grass weeds and wheat (*Triticum durum*) crop using linear and nonparametric functional discriminant analysis and neural networks. *Weed Research* 48 (2008) 28-37.

[24] M. Gómez-Casero, F. López-Granados, M. Jurado-Expósito, J.M. Peña-Barragán, L. García-Torres, and R. Fernández-Escobar. Assessing nitrogen and potassium deficiencies in olive orchards through discriminant analysis of hyperspectral data. *Journal of American Society of Horticultural Science* 132 (2007) 611-618.

[25] K.J. Girma, J. Mosali, W.R. Raun, K.W. Freeman, K.L. Martin, J.B. Solie, and M.L. Stone. Identification of optical spectral signatures for detecting cheat and ryegrass in winter wheat. *Crop Science* 45 (2005) 477-485.

[26] D.W. Hosmer, and S. Lemeshow, *Applied logistic regression*. New York: John Wiley & Sons, 1989.

[27] T. P. Ryan, *Modern Regression Methods*. New York: Wiley, 1997.

[28] G. McLachlan, *Discriminant analysis and statistical pattern recognition*. New York: John Wiley & Sons, 1992.

[29] P. J. Angeline, G. M. Saunders, and J. B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 5 (1), pp. 54-65, 1994.

[30] X. Yao and Y. Liu, "A new evolutionary system for evolving artificial neural networks," *IEEE Transactions on Neural Networks*, vol. 8 (3), pp. 694-713, 1997.

[31] A.C. Martínez-Estudillo, C. Hervás-Martínez, F.J. Martínez-Estudillo, and N. García, Hybridation of evolutionary algorithms and local search by means of a clustering method, *IEEE Transaction on Systems, Man and Cybernetics, Part. B: Cybernetics* 36(3) (2006) 534-546.

[32] I. SPSS ©, "SPSS ©9.0 advanced models," Inc, S., Ed. Chicago, IL, 1999.

TABLA II  
FUNCIONES DISCRIMINANTES DE LOS MODELOS RBF, LR Y NLIRUR

MODELO	FUNCIONES DISCRIMINANTES
RBF Nº ENLACES 11	$F_1 = -5.17 * (\text{EXP}(-0.5 * (((LO_{675} - 1.84)^2)^{0.5} / 0.26)^2)) + 1.81$ $F_2 = 5.36 * (\text{EXP}(-0.5 * (((LO_{600} - 1.38)^2 + (LO_{625} - 1.35)^2 + (LO_{675} - 1.69)^2 + (LO_{725} - 1.0)^2)^{0.5} / (0.63)^2)) - 2.36$ $F_3 = 4.54 * (\text{EXP}(-0.5 * (((LO_{675} - 1.84)^2)^{0.5} / 0.265)^2)) - 2.56$
LR Nº ENLACES 15	$F_1 = -203,007 - 259,120 * LO_{575} + 308,418 * LO_{600} - 131,521 * LO_{650} + 79,368 * LO_{700}$ $F_2 = -174,468 - 186,151 * LO_{575} + 188,650 * LO_{600} - 72,274 * LO_{650} + 67,597 * LO_{700}$ $F_3 = -217,393 - 276,693 * LO_{575} + 332,998 * LO_{600} - 134,072 * LO_{650} + 75,874 * LO_{700}$
NLIRUR Nº ENLACES 11	$F_1 = -1369,503 - 2429,896 * LO_{575} + 4551,510 * LO_{625} - 1883,166 * RBF_2$ $F_2 = -1791,112 - 1876,937 * LO_{575} + 3986,671 * LO_{625} - 1123,438 * RBF_2$ $F_3 = -1432,149 - 2607,932 * LO_{575} + 4786,869 * LO_{625} - 1936,293 * RBF_2$ $RBF_2 = \text{EXP}(-0.5 * (((LO_{600} - 1,38)^2 + (LO_{625} - 1,35)^2 + (LO_{675} - 1,69)^2 + (LO_{725} - 1,0)^2)^{0.5} / (0,63)^2))$