Algoritmos de aprendizaje evolutivo y estadístico para la determinación de mapas de malas hierbas utilizando técnicas de teledetección

P.A. Gutiérrez, J.C. Fernández, C. Hervás
Departamento de Informática y Análisis Numérico
Universidad de Córdoba
14071-Córdoba-España.
{zamarck@yahoo.es, chervas@uco.es, i82fecaj@uco.es }

Resumen

Este trabajo aborda la resolución de problemas de clasificación binaria utilizando una metodología híbrida que combina la regresión logística y modelos evolutivos de redes neuronales de unidades producto. Para estimar los coeficientes del modelo lo haremos en dos etapas, en la primera aprendemos los exponentes de las funciones unidades producto, entrenando los modelos de redes neuronales mediante computación evolutiva y una vez estimados el número de funciones potenciales y los exponentes de estas funciones, se aplica el método de máxima verosimilitud al espacio de características formado por las covariables iniciales junto con las nuevas funciones de base obtenidas al entrenar los modelos de unidades producto. Esta metodología híbrida en el diseño del modelo y en la estimación de los coeficientes se aplica a un problema real agronómico de predicción de presencia de la mala hierba *Ridolfia segetum* Moris en campos de cosecha de girasol. Los resultados obtenidos con este modelo mejoran los conseguidos con una regresión logística estándar en cuanto a porcentaje de patrones bien clasificados sobre el conjunto de generalización.

1. Introducción

Existen muchos campos de investigación como medicina, microbiología, epidemiología y muchos otros, donde es muy importante predecir el resultado de una variable de respuesta binaria, o de forma similar obtener la probabilidad de éxito en una variable aleatoria de Bernouilli, en función de la relación que tiene esta variable con un conjunto de variables explicativas o covariables. De esta forma, si lo consideramos como un problema de aprendizaje supervisado de clasificación binaria, la meta es aprender como distinguir ejemplos que pertenecen a una de entre dos clases (caracterizadas por los sucesos Y=1, e Y=0) en función de los valores que toman k variables predictoras o covariables $X_1, X_2, ..., X_k$.

La regresión logística es un modelo de regresión lineal generalizada donde se trata de predecir las probabilidades a posteriori de pertenencia de cada uno de los patrones de un conjunto de entrenamiento a uno de los valores que toma la variable dependiente mediante relaciones lineales con las variables predictoras [10]. Este tipo de modelos de regresión se aplica con mucha frecuencia a problemas donde la variable de respuesta es dicotómica, de forma tal que se le puede asignar los sucesos éxito y fracaso, como se hace por ejemplo en [13]. Por otra parte, un modelo de regresión logística se puede representar de forma equivalente a como se representa una estructura de grafo de tipo perceptrón, con una función de activación logística, representando un esquema de red neuronal lo más sencilla posible, como se observa en la Figura 1.

El procedimiento habitual de estimación de los coeficientes de un modelo lineal de RL es el de máxima verosimilitud, en el que para obtener el óptimo de la función de verosimilitud se utiliza habitualmente un método de optimización local basado en un algoritmo iterativo de tipo Newton-Raphson o de mínimos cuadrados con reasignación de pesos (IRLS) [6]. A la hora de aplicar la regresión logística, 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.

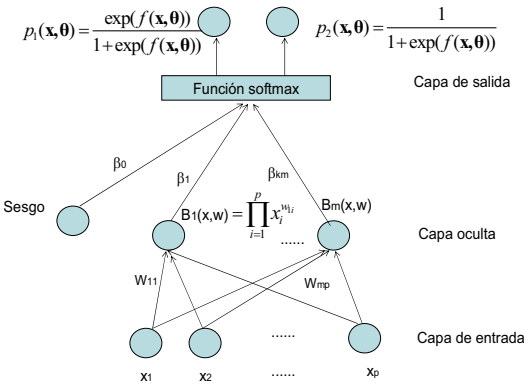

Figura 1. Esquema de red neuronal utilizado.

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** .

De esta forma, para introducir no linealidad en el modelo proponemos en este trabajo un modelo de regresión logística basado en la hibridación del modelo lineal en las covariables iniciales y en conjunto de funciones de base (no lineales) de tipo unidad producto, obtenidas a partir de Redes Neuronales de tipo Unidad Producto (PUNN), que suponen una alternativa de carácter multiplicativo a las funciones de tipo sigmoide. Mediante las unidades producto, pretendemos expresar las posibles interacciones significativas existentes entre las covariables.

Desafortunadamente en nuestra aproximación no podemos 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 PUNN son extremadamente complejas con numerosos óptimos locales y superficies planas.

Por ello, el aprendizaje de los coeficientes y del número de funciones de base se llevará a cabo en varias etapas. En la primera, se utilizará un algoritmo evolutivo (EA) para diseñar la estructura (número de funciones de base y enlaces), y fijar los pesos asociados a los exponentes de las covariables del modelo PUNN. En la segunda etapa, consideramos lineal el modelo tanto en estas nuevas funciones de base como en las covariables iniciales y podemos obtener los estimadores de los coeficientes de regresión mediante un procedimiento de máxima verosimilitud asociado al modelo estándar de regresión logística. Por último, aplicamos un método de selección de covariables en función de su capacidad para explicar la variable de respuesta.

Para evaluar el rendimiento de nuestra metodología utilizamos un problema de clasificación binaria asociado a la determinación de la probabilidad de presencia de la mala hierba *R.segetum* (*Ridolfia segetum* Moris) en función de los valores digitales en la banda Azul, Verde, Roja e Infrarroja Cercana de imágenes aéreas. Los resultados empíricos muestran que tanto la metodología como el modelo propuesto son muy prometedores tanto en su capacidad de predicción en la clasificación, como en su simplicidad, dada por el relativamente pequeño número de coeficientes y de funciones de base utilizadas en los mejores clasificadores.

El trabajo está organizado de forma tal que en la sección 2 se hace una revisión de trabajos relacionados con el tema que nos ocupa. En la sección 3 hacemos una breve introducción a la regresión logística para, a continuación, presentar nuestro modelo en profundidad. En la sección 4 describimos el procedimiento de estimación de los coeficientes del modelo basado en dos etapas con un método diferente de optimización en cada una de ellas. En las secciones 5 y 6 planteamos el problema de predicción de presencia de *R.segetum* en campos de cosecha de girasol y exponemos los resultados obtenidos, y en la sección 7 presentamos las conclusiones.

2. Trabajos relacionados

En esta sección damos una breve reseña acerca de los diferentes métodos que utilizan funciones de base para ir más allá de la linealidad del modelo, entre ellos citaremos trabajos recientes que muestran la proximidad existente entre modelos de regresión logística y métodos de "machine learning".

Los modelos aditivos generalizados [5] comprenden métodos estadísticos automáticos y flexibles que se utilizan para identificar y

caracterizar los efectos de la regresión no lineal. Para un problema de clasificación en dos clases, el modelo de regresión logística es un ejemplo de modelo aditivo generalizado donde se reemplaza cada término lineal por una forma funcional más general que aproxima funciones multidimensionales mediante la suma de funciones unidimensionales.

Uno de los modelos más populares de regresión no lineal es la regresión de tipo "spline" adaptativa multivariante (MARS) [4], donde las funciones base vienen determinadas por el producto de algún número de funciones unidimensionales de tipo "spline". Las funciones base se añaden incrementalmente durante el aprendizaje, utilizando una técnica constructiva de selección secuencial.

3. Regresión logística binaria con funciones de base de tipo producto

Sea $D = \{(\mathbf{x}_n, \mathbf{y}_n); n = 1,2,...,N\}$ una muestra extraída de la población que consideraremos como el conjunto de entrenamiento, donde $\mathbf{x}_n = (x_{1n},...,x_{kn})$ es el vector de medidas que toma valores en $\Omega \subset R^k$ e $\mathbf{y}_n$ una variable aleatoria de Bernouilli asociada a un problema de clasificación en dos clases, de forma tal que si $\mathbf{y}_n = 1$ el n-ésimo patrón pertenece a la primera clase C_1, con una probabilidad p condicionada a los valores de las covariables y si $\mathbf{y}_n = 0$ pertenece a la segunda clase, C_2. El modelo de regresión logística [7], [14], es una técnica habitual en estadística en la cual la probabilidad p de pertenencia a la primera clase se asocia con los valores $\mathbf{x}_n = (x_{1n},...,x_{kn})$ en la forma:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = f(\mathbf{x},\boldsymbol{\beta}) = \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x} = \sum_{i=0}^{k}\beta_i x_i$$

donde $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_k)$ son los coeficientes del modelo. Estos coeficientes $\boldsymbol{\beta}$, se estiman a partir de los datos del conjunto D. El método de estimación de estos coeficientes se basa habitualmente en el método de máxima verosimilitud, construyendo para ello la función:

$$l(\boldsymbol{\beta}) = \sum_{l=1}^{n_T}\{y_l \log p(\mathbf{x}_l;\boldsymbol{\beta}) + (1-y_l)\log(1-p(\mathbf{x}_l;\boldsymbol{\beta})\} = \sum_{l=1}^{n_T}\left\{y_l \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}_l - \log(1+e^{\boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}_l})\right\} \quad (1)$$

La matriz Hessiana asociada al procedimiento de maximización de $l(\boldsymbol{\beta})$ es semidefinida negativa [11], lo que implica que dicha función es cóncava sobre el parámetro $\boldsymbol{\beta}$. La concavidad, junto con el hecho de que el vector de parámetros $\boldsymbol{\beta}$ 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.

En este trabajo proponemos un modelo de regresión logística basado en la hibridación de un modelo estándar de regresión logística y un modelo de red neuronal de unidades producto, PUNN. La expresión general del modelo es la siguiente:

$$f(\mathbf{x},\boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^{k}\alpha_i x_i + \sum_{j=1}^{m}\beta_j \prod_{i=1}^{k} x^{w_{ji}}$$

con funciones de base $B_j(\mathbf{x},\mathbf{w_j}) = \prod_{i=1}^{k} x_i^{w_{ji}}$, y parámetros $\boldsymbol{\theta} = (\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{W})$, $\boldsymbol{\alpha} = (\alpha_0,\alpha_1,...,\alpha_k)$, $\boldsymbol{\beta} = (\beta_0,\beta_1,...,\beta_k)$ y $\mathbf{W} = (\mathbf{w}_1,\mathbf{w}_2,...,\mathbf{w}_m)$, con $\mathbf{w}_j = (w_{j1},w_{j2},...,w_{jk})$, $w_{ij} \in \mathbb{R}$.

De esta manera, la nueva función de probabilidad condicional es:

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{e^{f(\mathbf{x},\boldsymbol{\theta})}}{1+e^{f(\mathbf{x},\boldsymbol{\theta})}}$$

y la transformación logit:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = f(\mathbf{x},\boldsymbol{\theta})$$

En este caso los bordes de decisión son funciones no lineales y están definidas por la hipersuperficie $f(\mathbf{x},\boldsymbol{\theta}) = 0$ en el espacio $\mathbb{R}^k$. Dependiendo del parámetro $\boldsymbol{\theta}$, la hipersuperficie puede incluso ser no conexa. La representación del modelo en grafo se observa en la figura 1. La parte no lineal de $f_l(\mathbf{x},\boldsymbol{\theta}_l)$ se corresponde con redes neuronales *feed-forward* con funciones de transferencia sigmoides o producto, estas últimas introducidas por Durbin y Rumelhart [2].

Por otra parte, y retomando el modelo propuesto en (1) el logaritmo de la función de verosimilitud para n_T observaciones es:

$$l(\boldsymbol{\theta}) = \sum_{l=1}^{n_T} \left\{ y_l f(\mathbf{x}_l, \boldsymbol{\theta}) - \log(1 + e^{f(\mathbf{x}_l, \boldsymbol{\theta})}) \right\} \quad (2)$$

La naturaleza y las propiedades de la función $f(\mathbf{x}, \boldsymbol{\theta})$ implica que la matriz Hessiana asociada a la maximización de (2) es en general indefinida y que puede tener máximos locales, y por tanto, es probable que el algoritmo se quede atrapado en ellos. De esta forma, nuestra aproximación no puede garantizar que se alcance el máximo global en la superficie de verosimilitud. Para salvar este problema utilizamos un algoritmo evolutivo como parte del proceso de estimación de los coeficientes del modelo y también como un método para determinar el número, m, de unidades de base óptimo. En la siguiente sección presentamos el procedimiento de obtención de los estimadores $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{W}})$ de los coeficientes del modelo.

4. Aprendizaje de los parámetros

La propuesta se basa en la combinación de un algoritmo de Programación Evolutiva (explorador global) y de un procedimiento de optimización local (explotador local) llevado a cabo mediante un procedimiento de maximización de la función de verosimilitud asociada a un modelo de regresión logística.

En una primera etapa aplicamos un algoritmo evolutivo para encontrar las funciones base de unidades producto, $B_j(\mathbf{x}, \hat{\mathbf{w}}_j)$, $1 \le j \le m$, que se corresponden a la parte no lineal de la función $f(\mathbf{x}, \boldsymbol{\theta})$. Debemos determinar el número de funciones m y la matriz de pesos formada por los exponentes estimados de las funciones potenciales $\hat{\mathbf{W}} = (\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, ..., \hat{\mathbf{w}}_m)$. Nuestro algoritmo evolutivo tiene puntos en común con otros algoritmos evolutivos referenciados en la bibliografía [1],[16]. Consideraremos la ecuación (2) de $l(\boldsymbol{\theta})$ como la función de error de un individuo f de la población. Por tanto definimos la medida de aptitud de un individuo mediante una función estrictamente decreciente de la función de error $l(\boldsymbol{\theta})$ en la forma $A(f) = \frac{1}{1 + l(\boldsymbol{\theta})}$, donde $0 < A(f) \le 1$.

La idea básica del algoritmo es la utilización de operadores de selección y mutación (paramétrica y estructural) en el proceso de evolución. El algoritmo puede estudiarse detalladamente en [9].

A continuación aplicamos el modelo de regresión logística estándar a las variables $x_1, x_2, ..., x_k, z_1, ..., z_m$ en el nuevo espacio de características de entrada, siendo $z_1 = B_1(\mathbf{x}, \hat{\mathbf{w}}_1), ..., z_m = B_m(\mathbf{x}, \hat{\mathbf{w}}_m)$. De esta forma calculamos el máximo de la función de verosimilitud condicional para n_T observaciones:

$$l(\hat{\boldsymbol{\theta}}^1) = \sum_{i=1}^{n_T} \left\{ y_i f(\mathbf{x}, \hat{\boldsymbol{\theta}}^1) - \log(1 + e^{f(\mathbf{x}, \hat{\boldsymbol{\theta}}^1)}) \right\}$$

donde $\hat{\boldsymbol{\theta}}^1 = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\mathbf{W}})$. Para obtener este procedimiento de estimación utilizamos un método basado en el gradiente, obteniéndose los estimadores $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{W}})$ de los parámetros del modelo:

$$f(\mathbf{x}, \hat{\boldsymbol{\theta}}) = \hat{\alpha}_0 + \sum_{i=1}^{k} \hat{\alpha}_i x_i + \sum_{j=1}^{m} \hat{\beta}_j \prod_{i=1}^{k} x_i^{\hat{w}_{ji}}$$

Por último para obtener el modelo final, utilizaremos un método de eliminación de variables no significativas del modelo basado en un procedimiento de eliminación de 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 generalización. En cada etapa, eliminamos la variable menos significativa.

5. Probabilidad de presencia de la mala hierba *R. segetum*

Se ha comprobado la eficacia de la metodología presentada en un problema agronómico real de predicción de mapas de rodales de mala hierba en cultivos, utilizando para ello datos obtenidos de la teledetección (*remote sensing*).

Los sistemas de teledetección (tecnología de sensores remotos que realizan fotografías aéreas) proporcionan una gran cantidad de información con un coste razonable. El análisis de imágenes

remotas permite modelizar los diversos parámetros agronómicos para su aplicación en agricultura de precisión. Recientes avances en esta tecnología han planteado la necesidad de utilizar modelos más flexibles para estimar diferentes parámetros asociados a la determinación de la productividad y/o la presencia de malas hierbas en fincas. Un aspecto relacionado con la posibilidad de minimizar el impacto de la agricultura en la calidad del medioambiente es el desarrollo de aproximaciones más eficaces para la determinación de la producción y para la aplicación inteligente de herbicidas. El motivo es que, aunque los rodales de malas hierbas se encuentran situados en zonas determinadas del cultivo, los herbicidas se aplican indiscriminadamente en toda la finca, con las consecuentes pérdidas económicas y los problemas medioambientales derivados. En este sentido es esencial el análisis de características agronómicas, que, basándose en la precisión de los mapas de rodales generados, permitan la implantación de modelos capaces de obtener un mayor rendimiento, a partir, por ejemplo, de una disminución de las cantidades de los herbicidas [8]. Los beneficios potenciales económicos y medioambientales de la aplicación específica *in-situ* de herbicidas incluyen un ahorro del consumo de spray, de los costes de los herbicidas y un incremento en el control de los rodales [3] y [12].

Planteamos por tanto un problema de determinación de mapas de rodales de malas hierbas (más en concreto, la mala hierba *Ridolfia segetum* Moris, frecuente en los cultivos de girasol y capaz de ocasionar graves pérdidas en la producción) partiendo de imágenes aéreas. La finca analizada es la finca de Matabueyes, cercana a la ciudad de Córdoba, de la cual disponemos de imágenes obtenidas en los meses de mayo, junio y julio del 2003. En la figura 2 se muestra un ejemplo de dichas imágenes, en concreto la imagen en color de mayo. Para mayo y agosto, se dispone de las imágenes en color e infrarrojo color, mientras que para junio únicamente disponemos de la imagen en color.

En dichas imágenes, mediante un estudio de campo, se estableció la naturaleza de un total de 2400 píxeles considerados como verdad-terreno: 800 píxeles clasificados como *R.segetum*, 800 píxeles clasificados como Suelo Desnudo y 800 píxeles clasificados como Girasol. El objetivo es discernir entre *R.segetum* y todo lo demás (puesto que desde el punto de vista de la aplicación de herbicida no interesa discernir entre suelo y cultivo). Las variables de entradas son los valores digitales de las bandas de cada una de las imágenes disponibles, es decir: Rojo (R), Verde (V) y Azul (A), para la imagen de junio, y R, V, A e infrarrojo cercano (NIR) para las imágenes de mayo y julio. Se utilizó un diseño experimental de tipo *hold-out*, tomando un 70% (1120 píxeles) de los patrones para entrenar y un 30% (480 píxeles) para generalizar, y estratificando los conjuntos de datos para que haya la misma cantidad de píxeles con presencia y ausencia de *R.segetum*.

Figura 2. Imagen en color en mayo de la finca de Matabueyes.

En todos los experimentos el número de generaciones que se ha empleado en el algoritmo evolutivo ha sido de 500 generaciones, el número de ejecuciones ha sido 30 y el máximo de nodos en capa oculta para los modelos PUNN de 3 nodos en junio y 4 nodos en mayo y julio. Para poder seleccionar las variables más significativas del modelo de regresión logística, utilizamos un método de eliminación de variables por pasos, mediante el software SPSS 13.0 [15].

Los modelos comparados en los distintos experimentos son los siguientes: en primer lugar, el mejor modelo de PUNN del algoritmo evolutivo (EPUNN); en segundo lugar, la

regresión logística estándar (LR) sobre las covariables iniciales; y en tercer y cuarto lugar, la regresión logística únicamente sobre las funciones de base del mejor modelo EPUNN (LRPU) y sobre las mismas funciones de base complementadas con las covariables iniciales (LRLPU).

6. Resultados

Para validar el rendimiento de cada modelo utilizamos el porcentaje de patrones de los datos correctamente clasificados para el conjunto de generalización, CCR_G .

En la tabla 1 se presentan las matrices de contingencia de los modelos EPUNN, LR, LRPU y LRLPU, calculado sobre los conjuntos de entrenamiento y generalización. En la Tabla 2 presentamos los mejores modelos obtenidos con redes neuronales evolutivas de unidades producto, EPUNN, y con las diferentes metodologías propuestas de regresión logística para la fecha correspondiente a junio. Observamos que tanto en mayo como en junio el mejor modelo en cuanto a CCR_G es el LRLPU; mientras que en julio el mejor modelo es LPU, siendo en ambos casos muy importantes las diferencias existentes con los resultados obtenidos con el modelo de regresión logística estándar.

A partir de estos modelos y tras la generalización a todos los píxeles de la imagen, se pueden obtener mapas de predicción de rodales de *R.segetum*. En la figura 3 se representan los mapas obtenidos con los cuatro modelos (EPUNN, LR, LRPU y LRLPU) para el mes de junio. En concreto hemos reflejado la probabilidad de presencia de mala hierba predicha por el modelo en cada uno de los píxeles, utilizando una escala entre blanco (mínima probabilidad, cercana a 0) y verde oscuro (máxima probabilidad, cercana a 1). Utilizando estos mapas, el experto puede establecer un umbral de probabilidad a partir de la cuál considerar presencia de mala hierba. Claramente, los modelos LRPU y LRLPU disciernen mejor zonas de elevada densidad de mala hierba de zonas libres de *R.segetum,* siendo más interesantes desde el punto de vista de la aplicación inteligente de herbicidas.

7. Conclusiones

En este trabajo nos hemos centrado en la presentación de una nueva metodología donde se combinan modelos de regresión logística binaria con modelos de redes neuronales con funciones de unidad producto. Para estimar el número de estas funciones y los coeficientes asociados a los exponentes de las mismas, hemos utilizado un algoritmo heurístico basado en el paradigma de la Programación Evolutiva.

Hemos aplicado el modelo propuesto un problema real agronómico de predicción de presencia de la mala hierba *R.segetum* Moris en campos de cosecha de girasol. Los mejores resultados obtenidos de CCR_G para los modelos híbridos mejoran los obtenidos utilizando RL y en algunos casos PUNN. De esta forma, el modelo híbrido determina un buen balance entre considerar sólo un modelo lineal o sólo un modelo no lineal.

Agradecimientos

Este trabajo ha sido financiado en parte por el proyecto TIN2005-08386-C05-02 de la Comisión Interministerial de Ciencia y Tecnología y fondos FEDER.

Referencias

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

[2] Durbin, R., Rumelhart, D.: Products Units: A computationally powerful and biologically plausible extension to back propagation networks. Neural Computation, 1, 133-142, 1989.

[3] Thompson, J.F., Stafford, J.V., Miller P.C.H.: Potential for automatic weed detection and selective herbicide application. Crop Protection, 10, 254–259, 1991.

[4] Friedman, J.: Multivariate adaptive regression splines. Ann. Stat., 19, 1-141, 1991.

[5] Hastie, T.J., Tibshirani, R.J.: Generalized additive models. Chapman & Hall, 1990.

[6] Hastie, T., Tibshirani, R.J., Friedman, J.: The Elements of Statistical Learning. Data mining, Inference and Prediction, Springer, 2001.
[7] Hosmer, D.W. and Lemeshow, S.: Applied logistic regression. John Willey & Sons, New York, 1989.
[8] Karimi, Y., Prasher, S.O., McNairn, H., et al.: Classification accuracy of discriminant analysis, artificial neural networks, and decision trees for weed and nitrogen stress detection in corn. Trans. ASAE 48 (3), 1261-1268, 2005.
[9] Martínez-Estudillo, A.C., Martínez-Estudillo, F. J., Hervás-Martínez, C., et al.: Evolutionary Product Unit based Neural Networks for Regression. Neural Networks, 19 (4), 477-486, 2006.
[10] McCullagh, P., Nelder, J.A.: Generalized Linear Models. 2nd edn., London, 1989.
[11] McLachlan, G.: Discriminant analysis and statistical pattern recognition. John Wiley & Sons, New York, 1992.
[12] Medlin, C.R., Shaw, D.R., Gerard, P.D. et al.: Using remote sensing to detect weed infestations in Glycine max. Weed Science 48 (3) 393-398, 2000.
[13] Prentice, R.L., Pike, R.: Logistic disease incidence models and case-control studies, Biometrika, 66 (3), 403-411, 1979.
[14] Ryan, T.P.: Modern Regression Methods. Wiley, New York, 1997.
[15] SPSS ©: SPSS ©13.0 advanced models. Inc, S., Ed. Chicago, IL, 1999.
[16] Yao, X., Liu, Y.: A new evolutionary system for evolving artificial neural networks. IEEE Trans. on Neu. Net., 8 (3), 694-713, 1997.

Tabla 1. Matrices de contingencia (Y=0, Ausencia de R.segetum; Y=1, Presencia de R.segetum) para todas las fechas utilizando redes neuronales de evolutivas de unidades producto, EPUNN, regresión logística, LR (en itálica), regresión logística solo con PU, LRPU (entre paréntesis) y regresión logística con covariables iniciales y PU, LRLPU (entre corchetes).

Estado Fenológico (Fecha)	Respuesta Real	Aprendizaje: Respuesta Predicha Y=0	Y=1	C	CR (%)	Generalización: Respuesta Predicha Y=0	Y=1	C	CR (%)
Vegetativo	Y=0	384 352	176 208		68.5 62.9	164 148	76 92		68.3 61.7
		(383) [394]	(177) [166]		(68.4) [70.4]	(164) [168]	(76) [72]		(68.3) [70]
(mitad-Mayo)	Y=1	133 171	427 389		76.2 69.5	65 69	175 171		72.9 71.3
		(136) [141]	(424)419]		(75.7) [74.8]	(67) [69]	(173) [171]		(72.1) [71.3]
	CCR (%)		72.4		66.2				70.6 66.5
			(72.		1) [72.6]		(70.		2) [70.6]
Floración	Y=0	547 529	13 31		97,7 94,5	236 226	4 14		98.3 94,2
		(547)[552]	(13) [8]		(97,7)[98,6]	(237)[238]	(3)[2]		(98,8)[99,2]
(mitad-Junio)	Y=1	7 30	553 530		98,8 94,6	2 12	238 228		99,2 95
		(9) [11]	(551)[549]		(98,4) [98]	(2) [2]	(238)[238]		(99,2)[99,2]
	CCR (%)		9		8,2 94,6				98,7 94,6
			(98)[98,3]				(99)		[99,2]
Senescencia	Y=0	443 296	117 264		79.1 52.9	195 138	45 102		81 .2 57.5
		(443) [447]	(117) [113]		(79.1) [79.8]	(425) [189]	(55) [51]		(88.5) [78.8]
(mitad-Julio)	Y=1	105 131	455 429		81.2 76.6	52 60	188 180		78.3 75
		(111) [117]	(449) [443]		(80.2) [79.l]	(53) [50]	(187) [190]		(77.9) [79.2]
	CCR (%)		80.1		64.7				79.8 66.3
			(79.		6) [79.5]		(79.		6) [79]

Tabla 2. Modelos propuestos para determinar la probabilidad de existencia de *R.segetum* para la elaboración de los mapas de malas hierbas en el mes de junio

Método	nº coef	Mejores modelos
EPUNN	8	$P = 1/(1+e(-0,424+75,419\ (V)^{4,633} +0,322\ (R)^{-1,888} +14,990\ (A)^{3,496}\ (V)^{-3,415}))$
LR	4	$P = 1/(1+e(-0,694+8,282\ (A)-63,342\ (V)-11,402\ (R)))$
LRPU	7	$P = 1/(1+e(-17,227+143,012\ (V)^{4,633} +0,636\ (R)^{-1,888} +23,021\ (A)^{3,496}\ (V)^{-3,415}))$
LRLPU	9	$P = 1/(1+e(18,027+130,674\ (A)-133,662\ (V)-29,346(R)+353,147\ (V)^{4,633} -3,396(A)^{3,496}\ (V)^{-3,415})))$

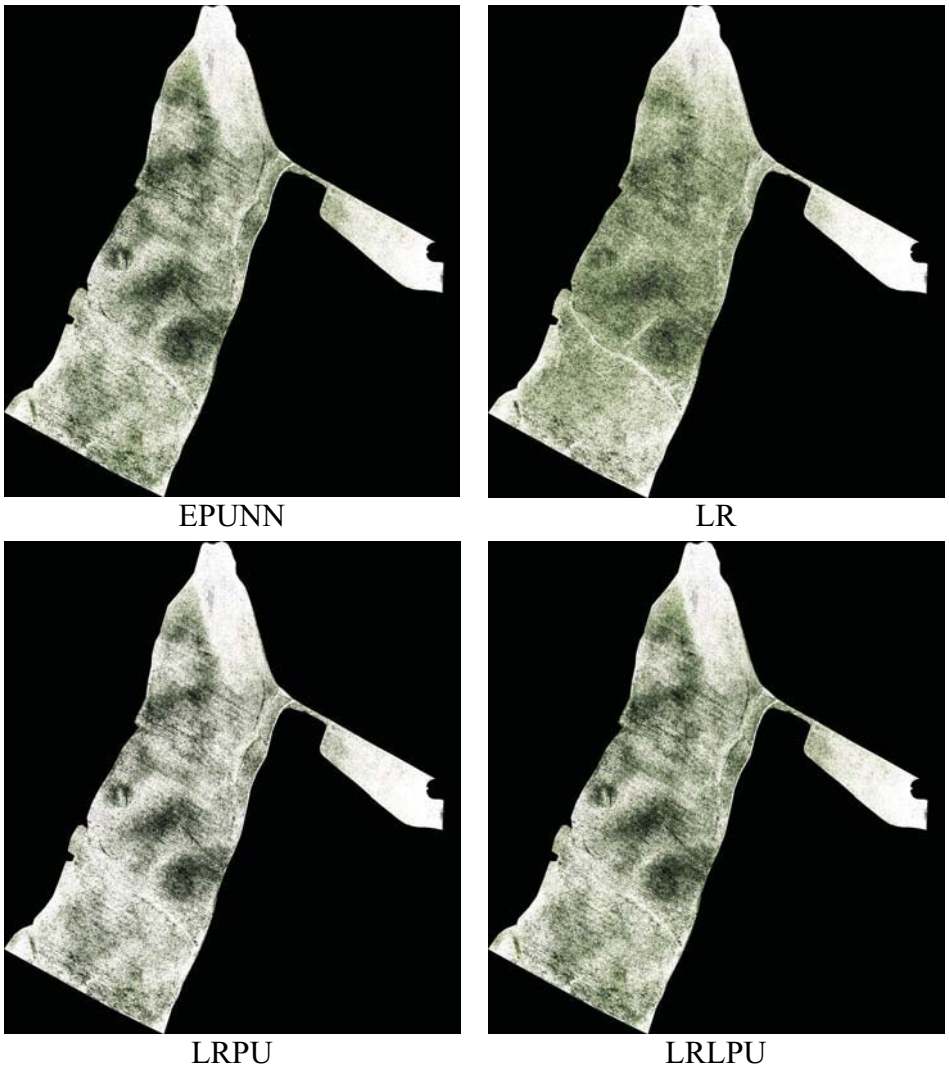

Figura 3. Mapas de predicción de rodales de *R.segetum* en junio obtenidos con cada uno de los modelos estudiados: EPUNN, LR, LRPU y LRLPU.