APRENDIZAJE: TERCER CURSO DEL GRADO DE ING. INFORMÁTICA EN COMPUTACION

Introducción al Aprendizaje Automático

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

2019-2020

Introducción al Aprendizaje Automático

Que es el Aprendizaje Automático?

"Cualquier cambio en un sistema que le permita realizar mejor una tarea la segunda vez que la repita, o haga otra tarea extraída de la misma población de la que se obtuvo la información" (Simon, 1983).

El Aprendizaje Automático estudia cómo construir programas que mejoren automáticamente con la experiencia o los datos.

¿Por que estudiar Aprendizaje Automático?

Dados los recientes avances tanto en teoría como en algorítmica para resolver problemas de clasificación, regresión agrupamiento y optimización

Crecimiento desbordante de datos “en línea” (on line).

Disponibilidad de computadoras suficientemente potentes.

Interés por parte de la industria. Industria 4.0

Introducción al Aprendizaje Automático

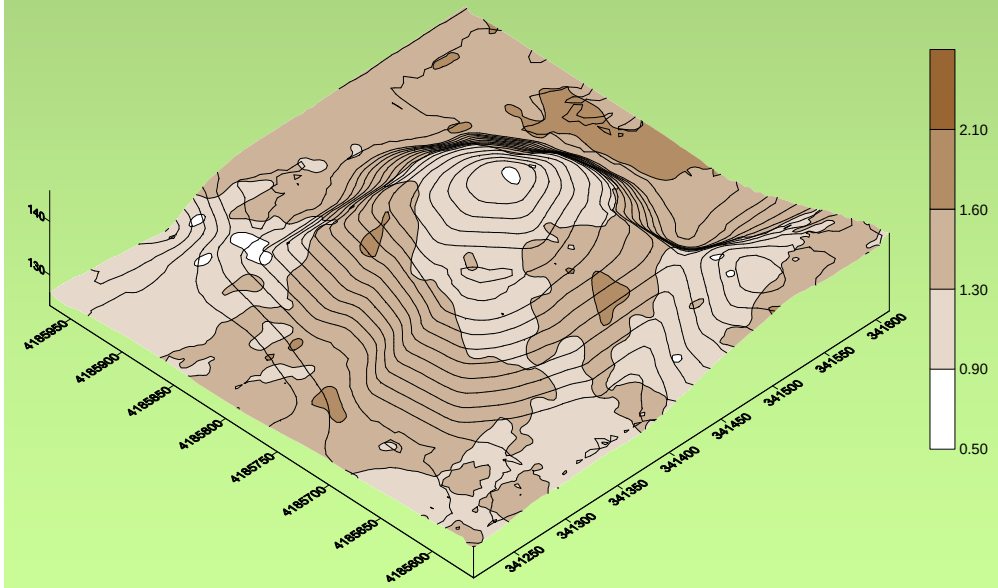
Algunos ejemplos

- **Minería de datos (*Data mining*): Uso de datos históricos para mejora de toma de decisiones. Predicción clasificación y agrupamiento en series temporales**
- **Datos médicos. Decisiones médicas, sistemas de apoyo a la decisión**
- **Conducción autónoma de vehículos: Aprenden a conducirse en autopistas en función de la información visual que reciben.**
- **Reconocimiento del habla, de imágenes de señales de video, etc**
- **Teoría y desarrollo de Juegos**
- **Programas que se modifican a si mismos según las costumbres del usuario**
- **Lector de periódicos que aprende los temas de interés del usuario. Recuperación de documento**
- **Gestor de correo electrónico. Detectores de spam**
- **Domótica. Casas que aprenden a optimizar el gasto de energía en función de las costumbres y hábitos de sus ocupantes.**

APLICACIONES

- **Agroalimentación**

5. Predicción de mapas de cosecha de girasol infestado de Ridolfia segetum mediante Redes Neuronales Evolutivas de Unidades Producto



1. Introducción

Presentación del problema

Problema abordado

Predecir mapas de rendimiento del cultivo de girasol con infestaciones naturales de *Ridolfia segetum*, incorporando la pendiente del campo y los datos espectrales de imágenes aéreas.

Problema de Modelado (o Modelizado) de Sistemas.

- **Modelar sistemas consiste en establecer una relación funcional entre las variables que intervienen en un fenómeno de estudio.**
  - **Modelos de clasificación: Predicen la clase de pertenencia.**
  - **Modelos de regresión: Predicen el valor de una variable del fenómeno.**

1. Introducción

Presentación del problema

Objetivo: Modelo de Regresión que establezca una relación entre:

- **Variable cuyo valor queremos predecir:**
  - ***Rendimiento de girasol*.**
- **Variables de entrada:**
  - ***Valores digitales multiespectrales*.**
  - ***Densidad de mala hierba.***
  - ***Elevación del terreno.***

1. Introducción

Datos de partida

- **Área de estudio**
  - **Finca de Matabueyes, parcela central**
    - **Sembrada de girasol**
    - **Con infestaciones naturales de *R. segetum***
  - **Estudiada en mayo del 2003.**

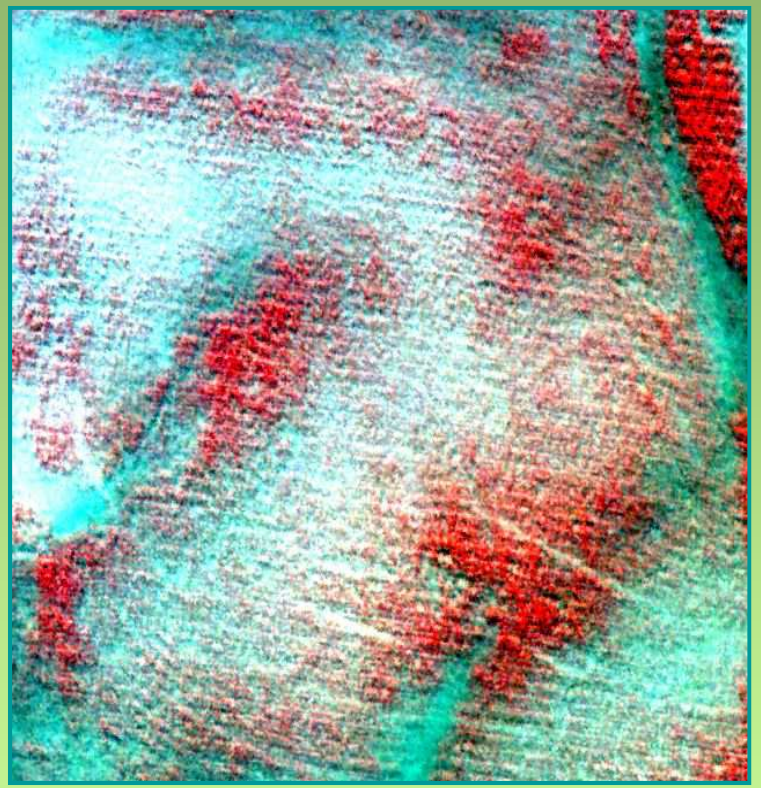
Este trabajo tiene como punto de partida un conjunto de datos resultado de distintos trabajos anteriores en el área de estudio

- **Imágenes aéreas.**
- **Mapa de infestación de malas hierbas.**
- **Modelo Digital de Elevaciones.**
- **Mapa de rendimiento de cosecha.**

1. Introducción

Datos de partida

- **Imágenes aéreas** → (Peña-Barragán et al., 2007)

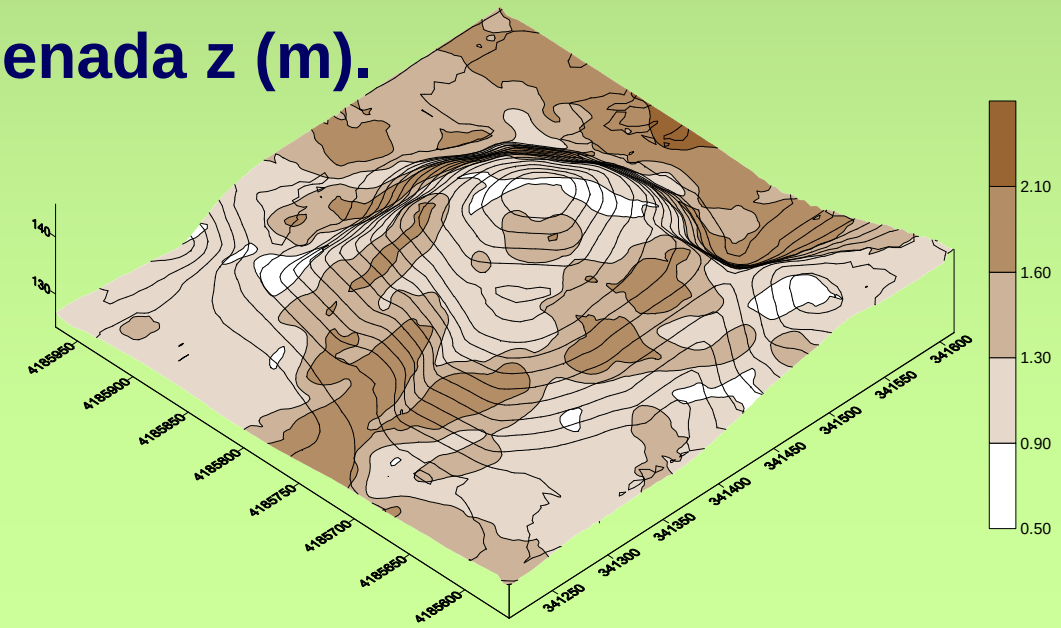


- **Imagen en Color**
  - **Blue (B, 400-500nm) 700-900nm)**
  - **Green (G, 500-600nm)**
  - **Red (R, 600-700nm)**

- **Imagen en Color Infrarrojo**
  - **Near Infrarred (NIR)**

1. Introducción
Datos de partida

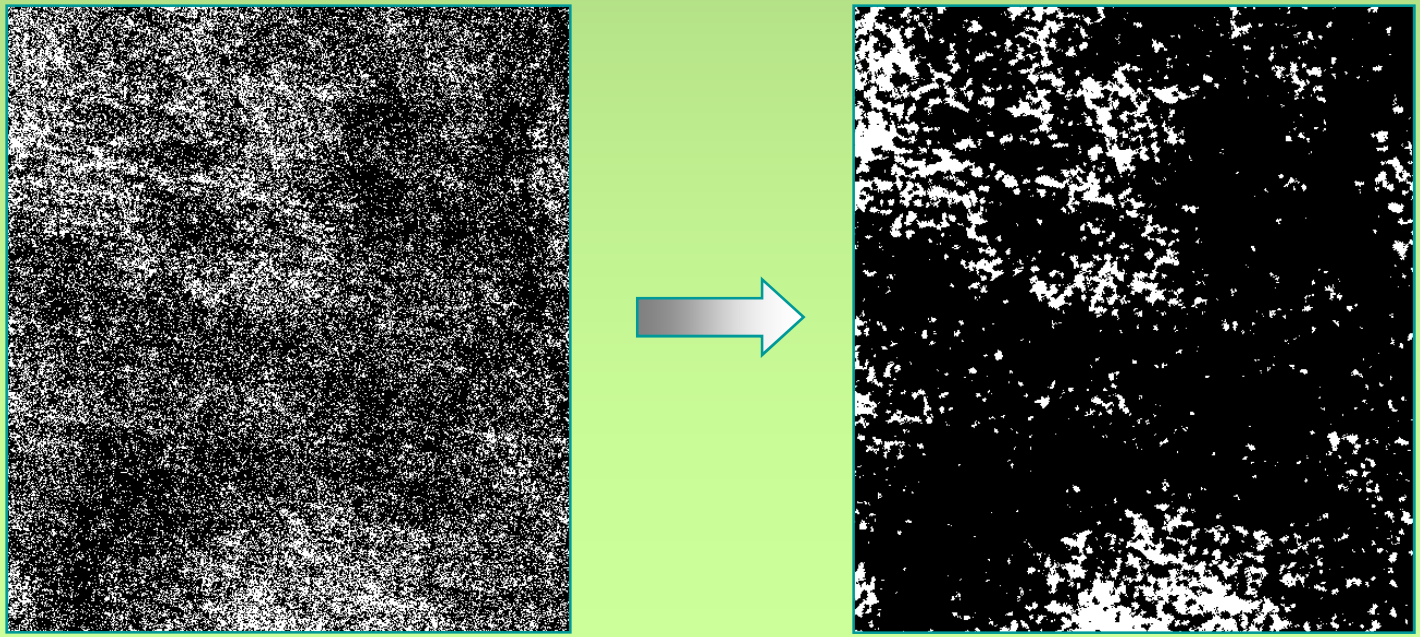
- **Elevación del terreno y Rendimiento de girasol**
  - **Se recolectó el girasol utilizando la cosechadora de precisión con sistema FieldStar de Massey Fergusson en Agosto de 2003.**
  - **También incorporaba un Sistema de Posicionamiento Global (GPS), que sirvió para describir los datos de elevación. Coordenada z (m).**



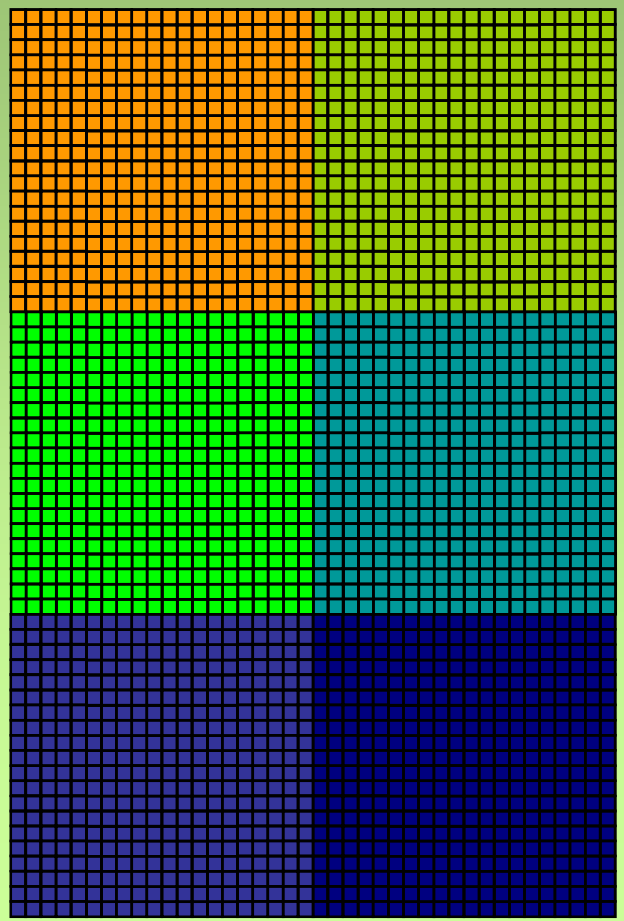
1. Introducción

Datos de partida

- **Infestación de mala hierba**
  - **Se obtuvieron los mapas de rodales de *R. segetum*, tal y como se describe en (Peña-Barragán et al., 2007).**
    - **Utilización del método Mapeador de Ángulo Espectral (SAM)**
    - **Aplicación posterior de un filtro de mediana, utilizando una celda 5x5, para reducir el ruido de Sal**



❑ Reducción del volumen de datos



En cada píxel tenemos información sobre:

- Densidad de mala hierba (pl/m2)
- Valor Digital en la banda Roja
- Valor Digital en la banda Verde
- Valor Digital en la banda Azul
- Valor Digital en la banda Infrarroja
- Coordenada Z (m)
- Rendimiento de cosecha (tn/ha)

Promedio de todos los valores

- Celdas de 20x20 píxels
- Reducción del volumen de datos en un 400%

1. Introducción

Metodologías empleadas

- Regresión Lineal Múltiple por el método Stepwise (SMLR)
- Modelos de Redes Neuronales de Unidad Producto (PUNN). Entrenados con:
  - Programación Evolutiva (EP)
  - Programación Evolutiva Híbrida (HEP)
  - Programación Evolutiva Híbrida con Agrupamiento (HEPC)
  - Programación Evolutiva Híbrida con Agrupamiento Dinámica (HEPCD)
  - Ensembles de HEPCD basados en la Media (E_{MEAN})
  - Ensembles de HEPCD basados en la Mediana (E_{MEDIAN})

Resultados

Resultados estadísticos de las distintas metodologías evolutivas

- **Algoritmos Evolutivos → No deterministas**
  - **Distintas ejecuciones dan lugar a distintos resultados.**
  - **Realización de 30 ejecuciones y obtención de la media del MSE de la mejor ANN de cada metodología a lo largo de las 30 ejecuciones, la desviación típica (SD), el menor MSE y el mayor MSE.**
  - **70% de los datos para entrenamiento y 30% de los datos para generalización.**

Algoritmos	MSE							
	Entrenamiento				Generalización			
	Media	SD	Menor	Mayor	Media	SD	Menor	Mayor
EP	0,0555	0,0022	0,0517	0,0589	0,0520	0,0032	0,0474	0,0579
HEP	0,0508	0,0014	0,0490	0,0552	0,0484	0,0016	0,0457	0,0529
HEPC	0,0506	0,0014	**0,0488**	0,0552	0,0484	0,0016	0,0454	0,0528
HEPCD	**0,0501**	**0,0009**	**0,0488**	**0,0520**	0,0480	0,0012	0,0463	0,0502
E_{MEAN}	0,0503	0,0010	0,0489	0,0530	**0,0475**	**0,0009**	**0,0449**	**0,0490**
E_{MEDIAN}	0,0504	0,0011	0,0489	0,0536	0,0477	**0,0009**	0,0454	0,0493

Resultados
Comparación de las metodologías evolutivas con el análisis SMLR

- **Mejor ANN de todas las metodologías evolutivas (correspondiente a HEPCD) comparado con la realización de una regresión SMLR.**

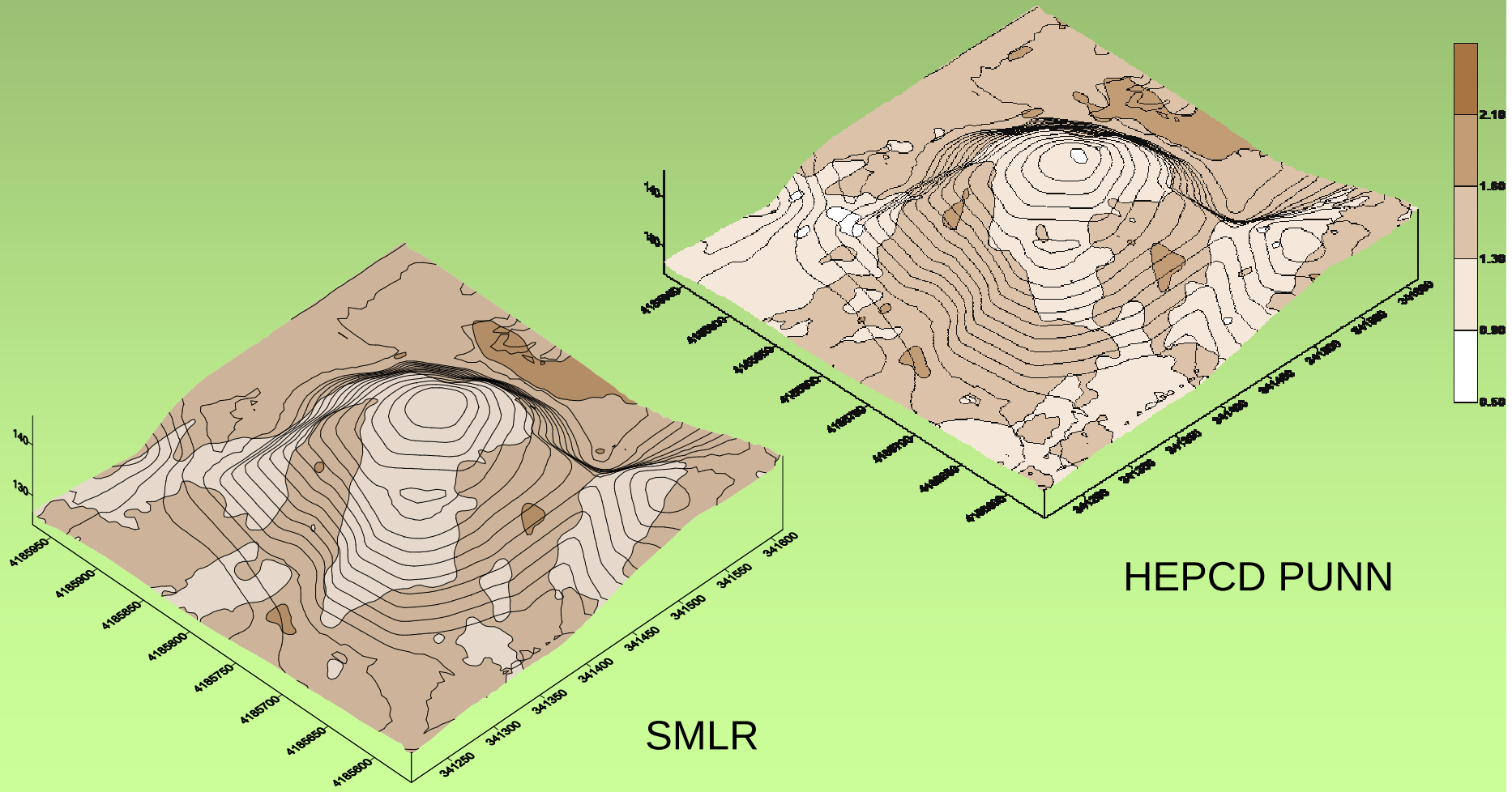
Modelos	$MSE_{Entrenamiento}$	$MSE_{Generalización}$
SMLR	0,0673	0,0642
HEPCD	0,0505	0,0454

Modelos	Ecuaciones de Regresión
SMLR	Rendimiento (tn/ha) = 3,7031- 0,0097*Ridolfia* - 0,0133*Z* - 0,0033*R* + 0,0003*NIR*
HEPCD	Rendimiento (tn/ha) = 0,7396 $(G^{6,1364}\ Z^{-1,1831}\ Ridolfia^{0,4289})$
	-0,0709 $(NIR^{-0,7263}\ R^{6,2797}\ B^{-2,3923}\ Z^{-2,1374}\ Ridolfia^{0,5795})$
	-13,1973 $(NIR^{-0,0902}\ R^{2,8042}\ G^{-0,3956}\ B^{1,6049}\ Z^{0,9406})$
	+33,8768 $(R^{1,5069}\ B^{2,9757}\ Z^{0,5926}\ Ridolfia^{0,0552})$
	-22,3513 $(NIR^{0,0758}\ G^{0,9941}\ B^{3,7274}\ Z^{0,3144}\ Ridolfia^{0,1106})$
	+1,6973

Z: pendiente en metros. G, B, R y NIR: valores digitales espectrales de las bandas del verde (G), azul (B), rojo (R) e infrarrojo cercano (NIR); *Ridolfia*: nº pls. m^{-2}.

Resultados
Comparación de las metodologías evolutivas con el análisis SMLR

- **Mapas de cosecha obtenidos por ambos modelos.**



HEPCD PUNN

SMLR

APLICACIONES

- **Biomedicina : Detección de la enfermedad de Parkinson**

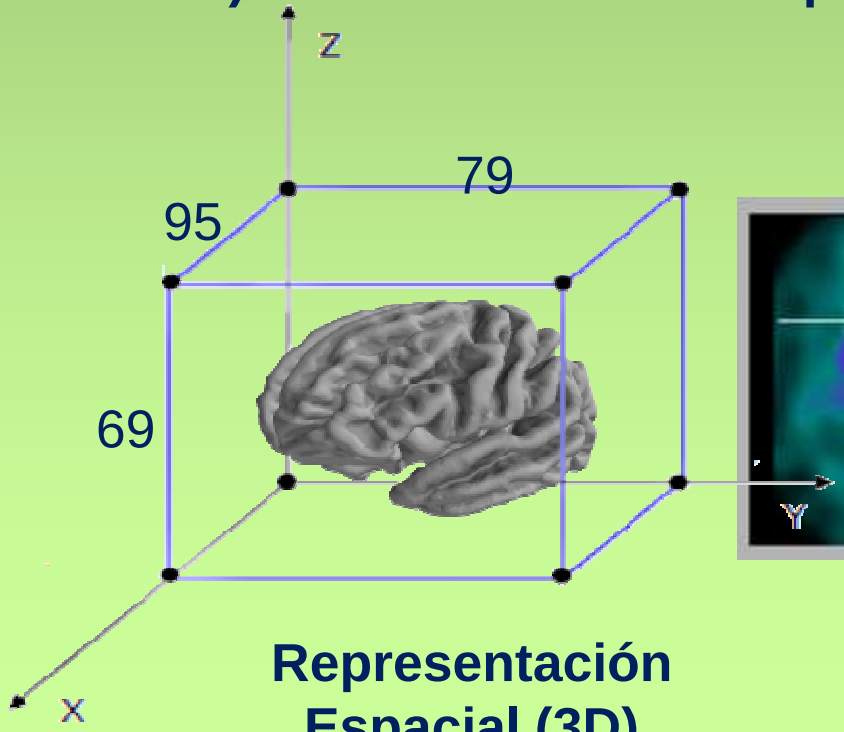
Enfermedad de Parkinson: Modelo de clasificación binaria

Número de pacientes: 165

65 padecen la enfermedad de Parkinson (patológico) y 100 no la padecen (normal),

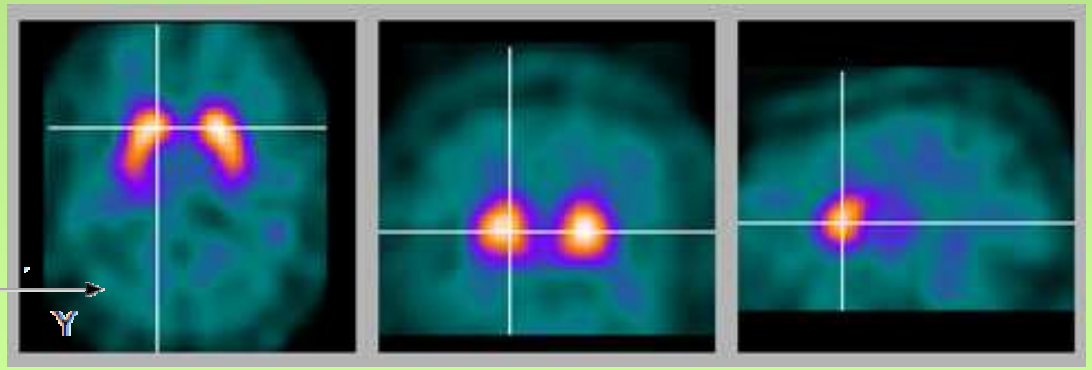
Número de clases 2: Sanos, clase 0, y Enfermos clase 1

La media de edad de los pacientes es de 70,46 años (35-89) con una desviación típica de 11,85 años



Representación Espacial (3D)

Planos Cerebrales (2D)



Axial **Coronal** **Sagital**

Metodología
Regresión logística regularizada, RLR

Función de verosimilitud de la RL

$$L(\mathbf{a},b)=\sum_{i=1}^{M} y_i \ln(p(y_i \mid \mathbf{x}_i))-\sum_{i=1}^{M}(1-y_i)\ln(1-p(y_i \mid \mathbf{x}_i))$$

siendo M el número de patrones

Función de verosimilitud regularizada de la RLR

$$L_{L1}(\mathbf{a},b)=L(\mathbf{a},b)+C\sum_{i=1}^{V}|a_i|$$

$$L_{L2}(\mathbf{a},b)=L(\mathbf{a},b)+C\sum_{i=1}^{V}(a_i^2)$$

siendo V el número de voxeles

RESULTADOS

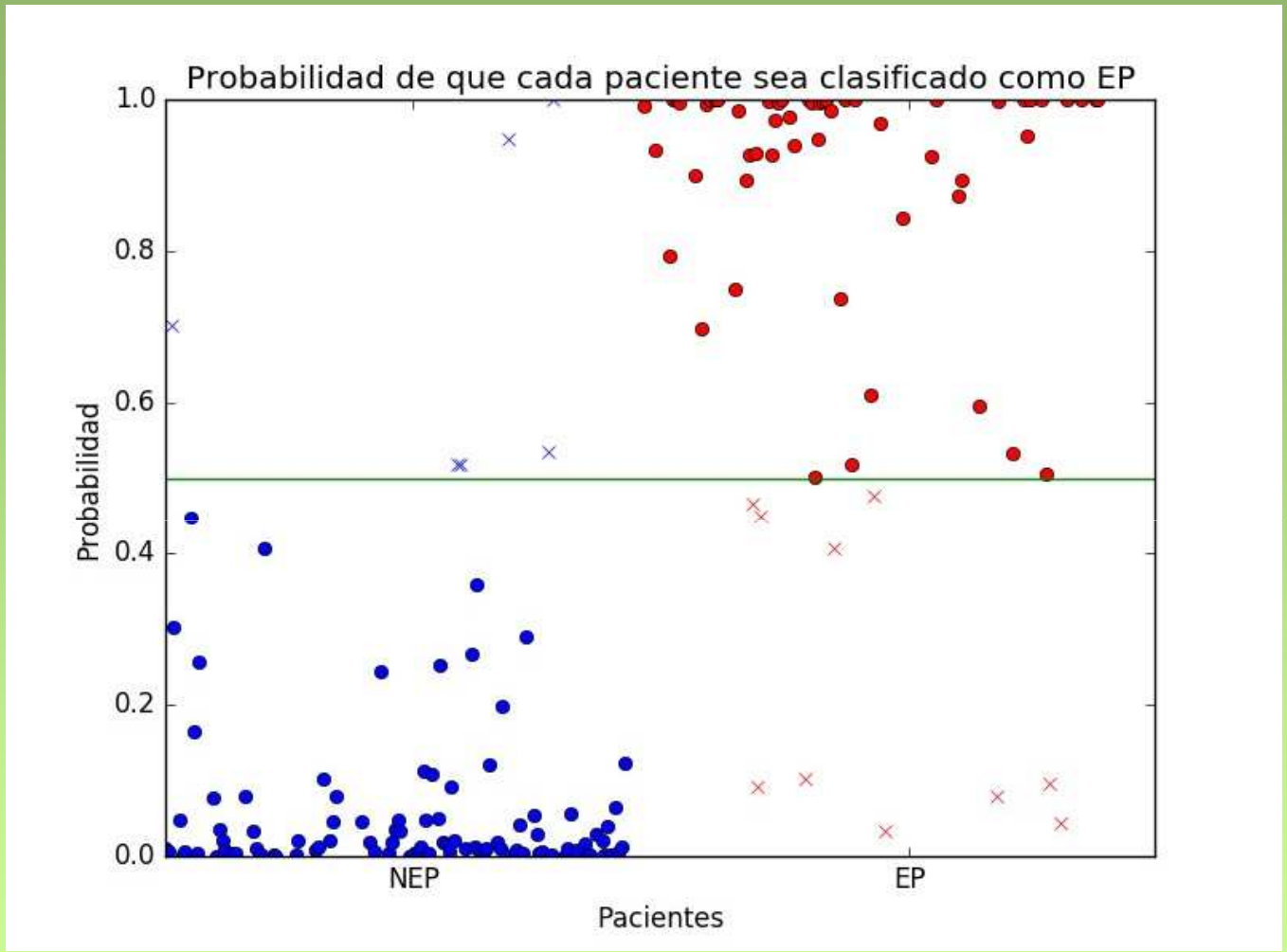


Figura. Probabilidad de cada paciente de pertenecer a la clase patológico o enfermo de Parkinson.

RESULTADOS

Subconjunto	CCR	Sensibilidad clase enfermo SEN	Sensibilidad clase sano SPC	AUC
1er *fold*	0,8788	0,9231	0,8500	0,9298
2º *fold*	0,9091	0,8462	0,9500	0,9615
3er *fold*	0,9091	0,8462	0,9500	0,9563
4º *fold*	0,8788	0,7692	0,9500	0,9428
5º *fold*	0,8485	0,7692	0,9000	0,9663
Media	**0,8848**	**0,8308**	**0,9200**	**0,9574**

Cuadro. Resultados de buena clasificación obtenidos en el diseño de un 5-fold validación cruzada

RESULTADOS

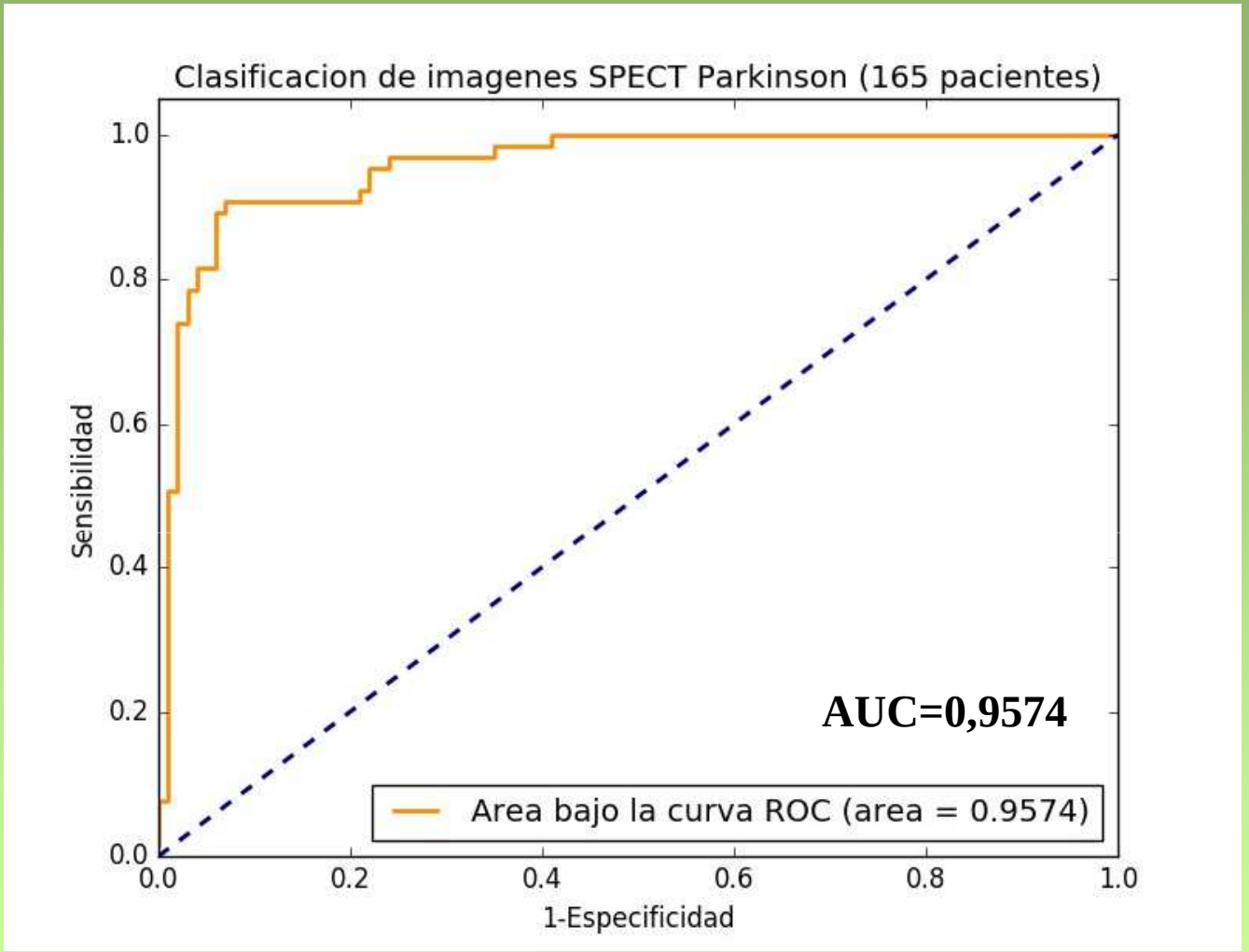


Figura. Área bajo la curva ROC del mejor modelo de RLR obtenido.

APLICACIONES

- **Biomedicina : Detección de los diferentes estadíos de Melanoma**

Biomedicina: Diagnóstico de Melanoma mediante clasificación multiclase y ordinal

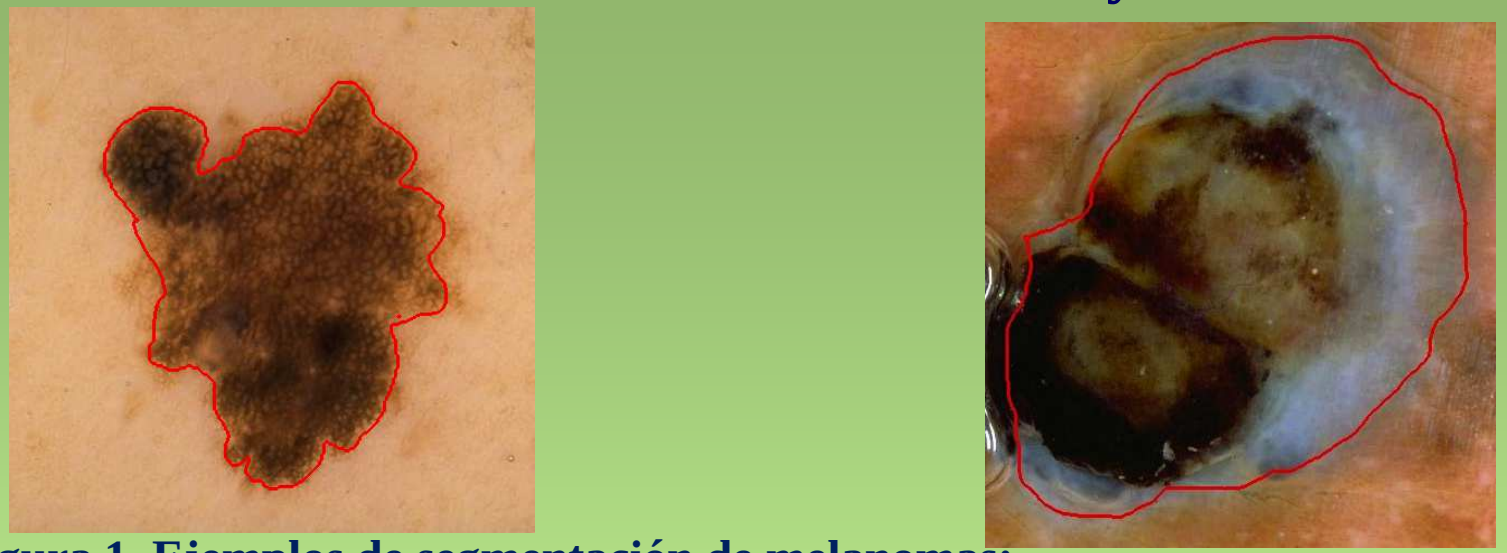


Figura 1. Ejemplos de segmentación de melanomas: (a) melanoma <0.76 mm, (b) melanoma >0.76 mm

Figura 2. (a) Imagen Original (b) Detección mediante una red pigmentada

Enfoques de descomposición de etiquetas en orden parcial para el diagnóstico de melanoma

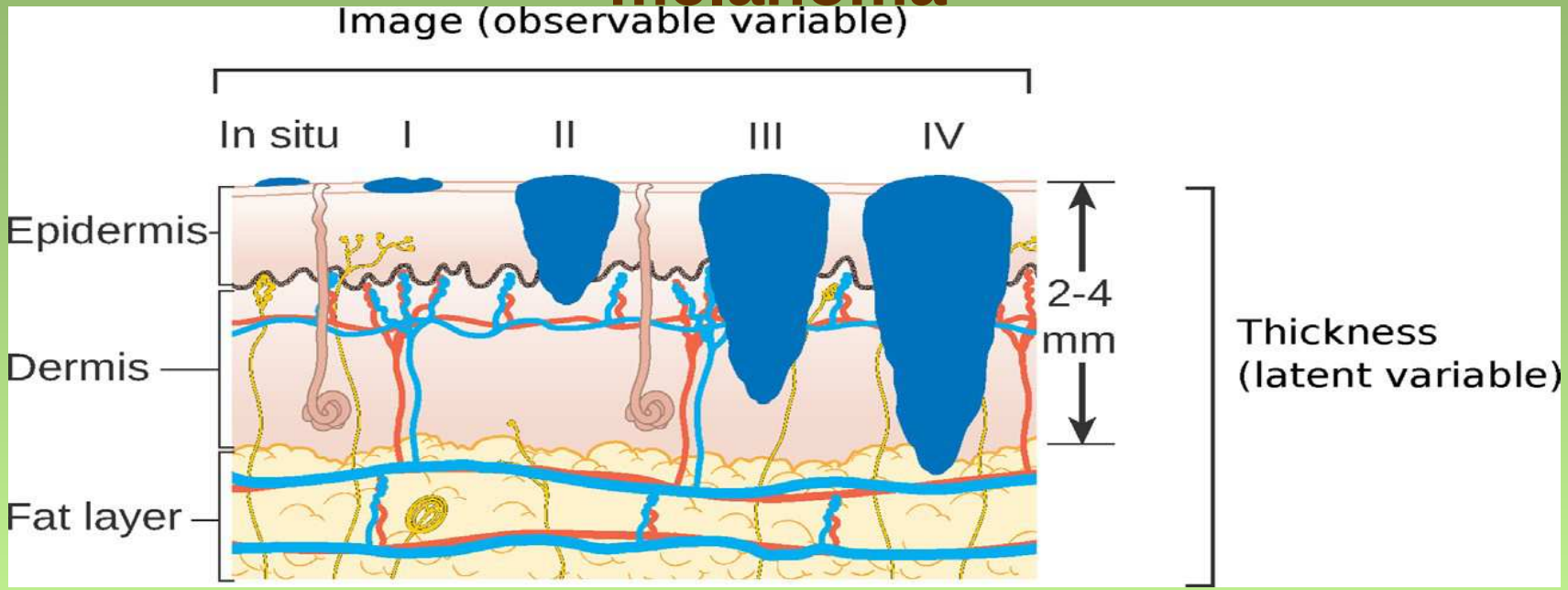


Fig. 1. Representación gráfica de las diferentes etapas del melanoma, donde se pueden analizar tanto los datos observables (imagen dermatoscópica) como la variable no observable o latente (grosor del tumor).

Crédito de la imagen: Cancer Research UK / Wikimedia Commons.

Clasificación parcialmente ordenada

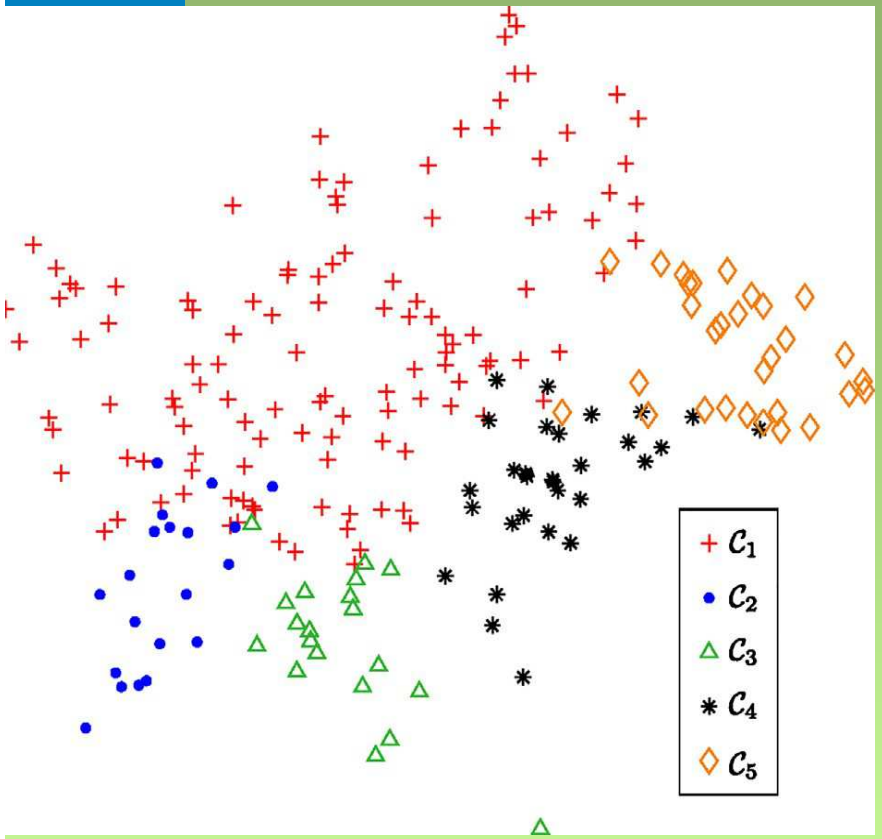


Fig. 2. Ejemplo de una base de datos parcialmente ordenada

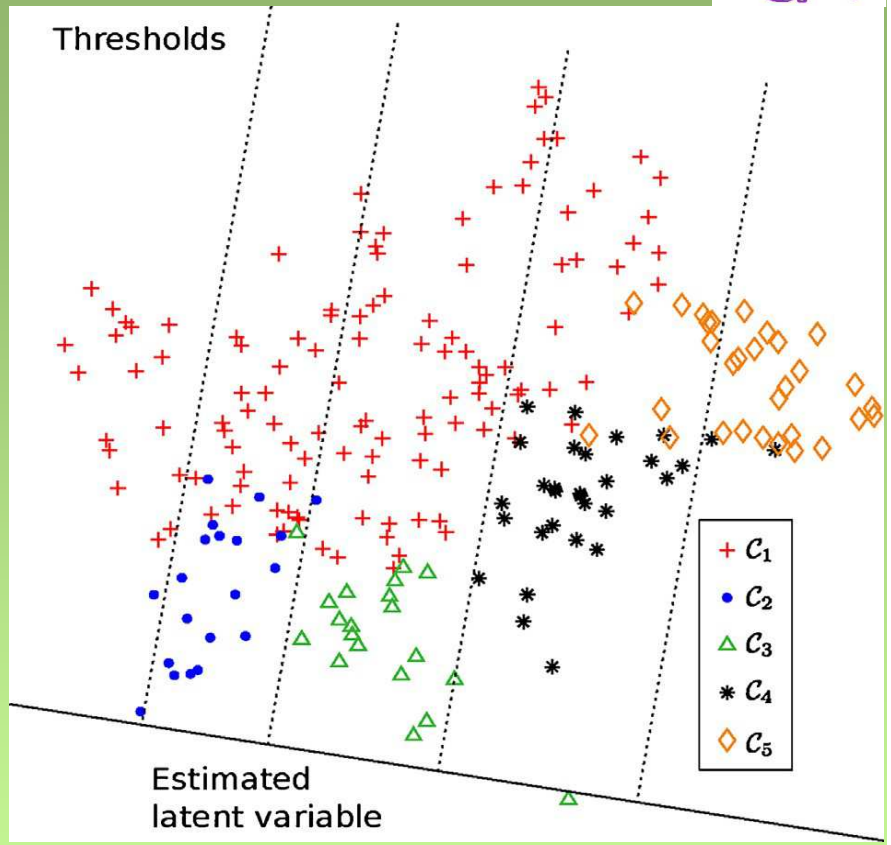
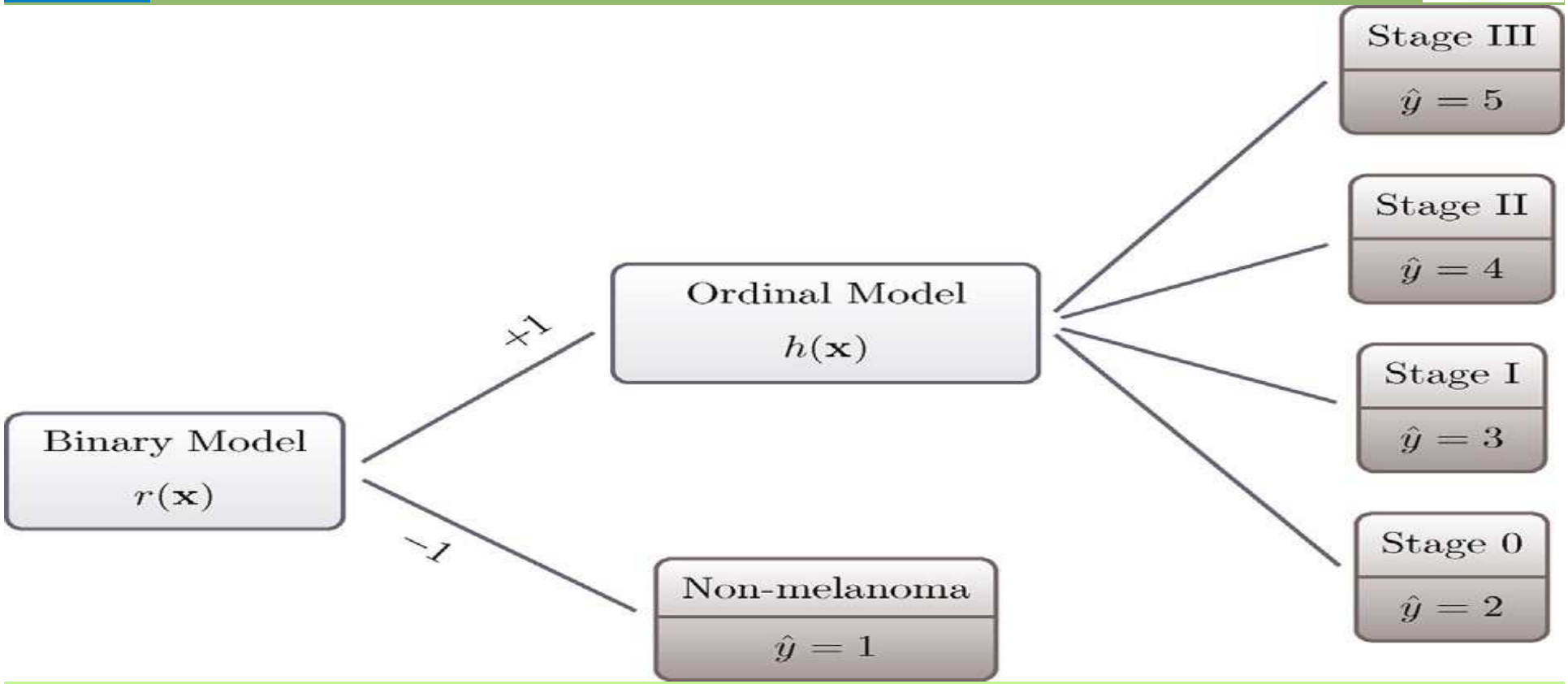


Fig. 3. Ejemplo de un modelo de umbral ordinal que se ha entrenado considerando los datos de la figura 2. Se puede ver que debido a la estructura de los datos, el modelo ordinal no es capaz de obtener una solución adecuada.

Diagnóstico de Melanoma

Proceso de predicción del clasificador jerárquico

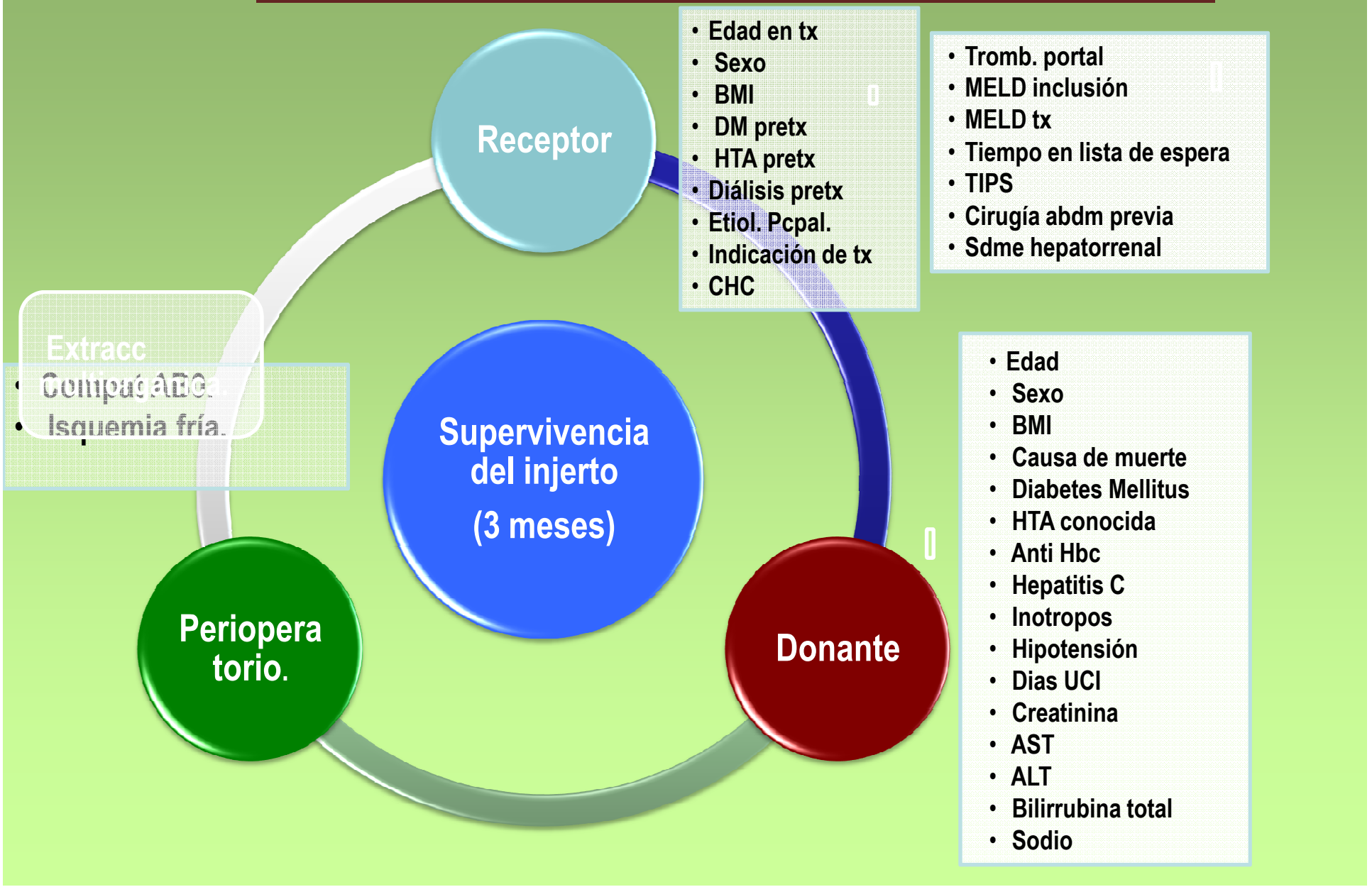


El modelo de la estructura jerárquica está basado 1) En la Regresión Logística (LR) para el modelo binario y 2) En el Modelo de probabilidades proporcionales para el clasificador ordinal, POM (el POM adapta la LR estándar al caso ordinal).

APLICACIONES

- **Biomedicina : Modelos de asignación Donante-Receptor en trasplante hepático**

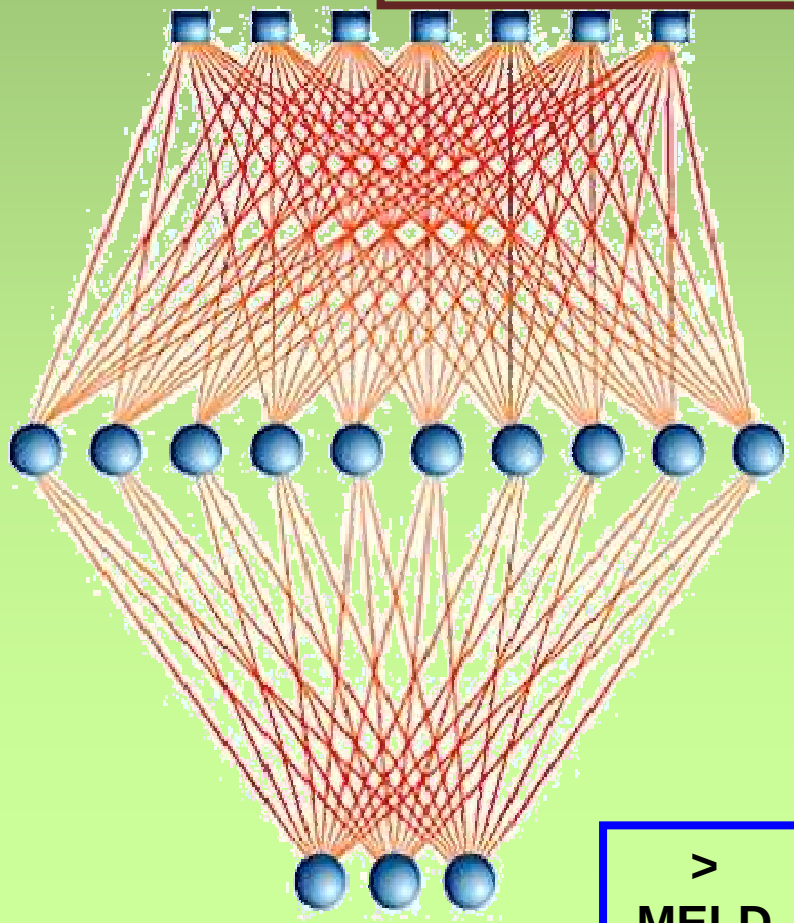
Utilidad de las redes neuronales artificiales en la asignación donante-receptor en trasplante hepático

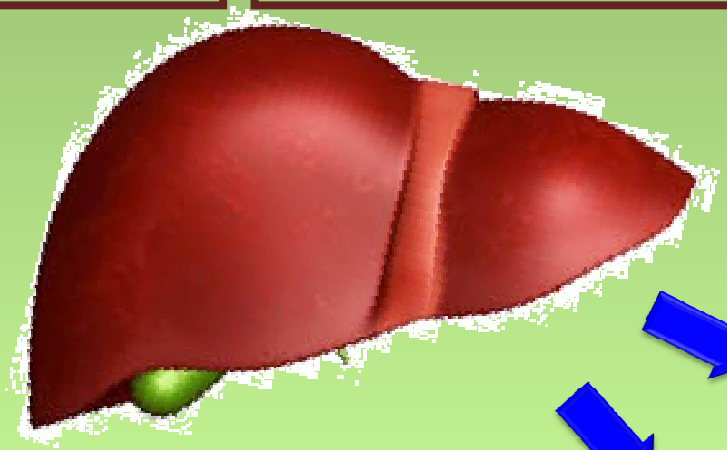


Utilidad de las redes neuronales artificiales en la asignación donante-receptor en trasplante hepático

“Supervivencia del injerto a 3 meses”
Probabilidad de aceptación (modelo CCR)

“No Supervivencia del injerto a 3 meses”
Probabilidad de rechazo (modelo MS)





R1

R2

R3

R4

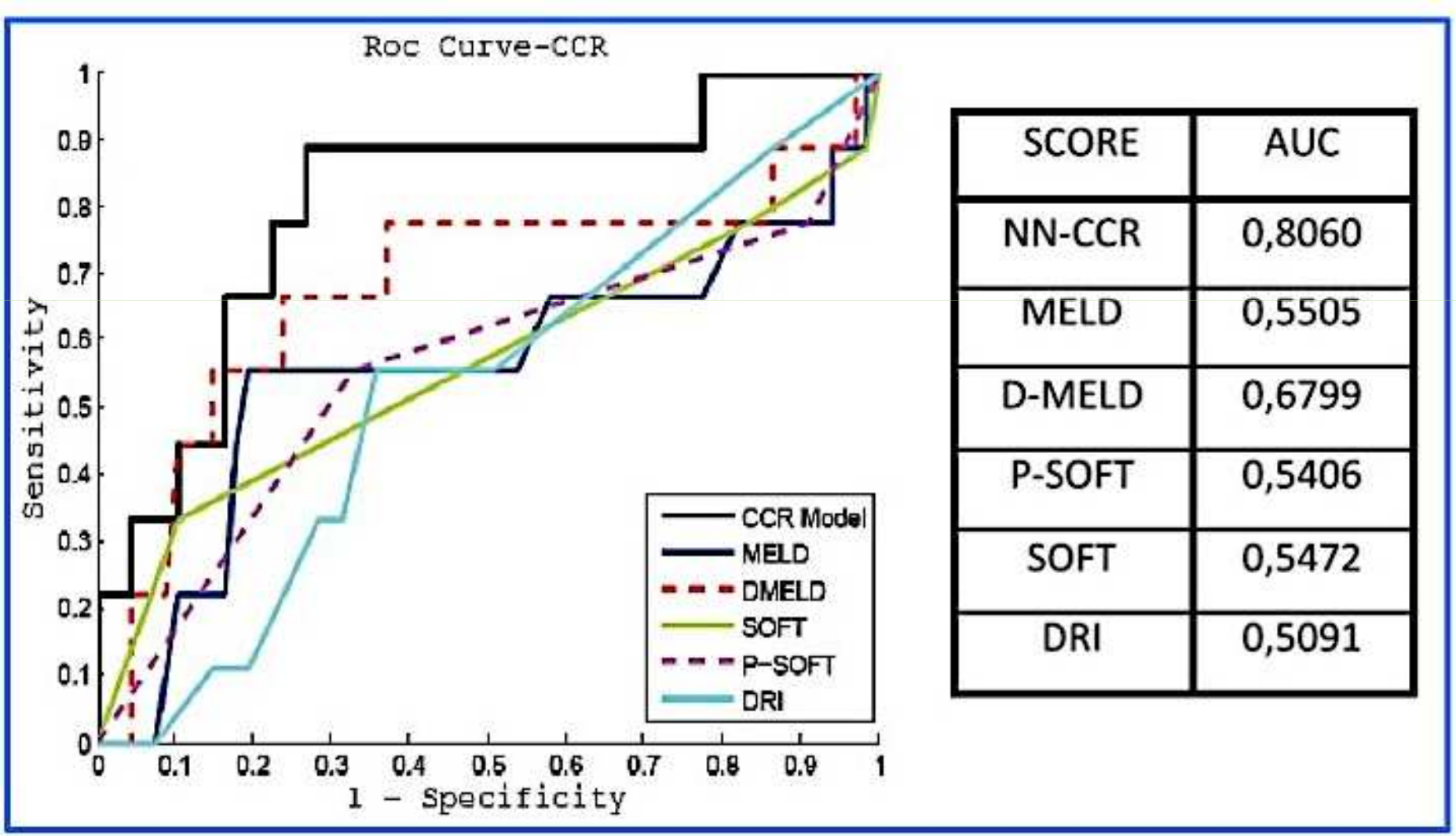
R5

R6

> MELD

< MELD

Utilidad de las redes neuronales artificiales en la asignación donante-receptor en trasplante hepático



SCORE	AUC
NN-CCR	0,8060
MELD	0,5505
D-MELD	0,6799
P-SOFT	0,5406
SOFT	0,5472
DRI	0,5091

Biomedicina / Biométrica

- **Medicina:**
  - Diagnóstico a través de imágenes
  - Diagnóstico y pronóstico
  - Detección de drogas
- **Seguridad:**
  - Reconocimiento de imágenes
  - Firmas / Huellas dactilares / Verificación del iris
  - DNA fingerprinting

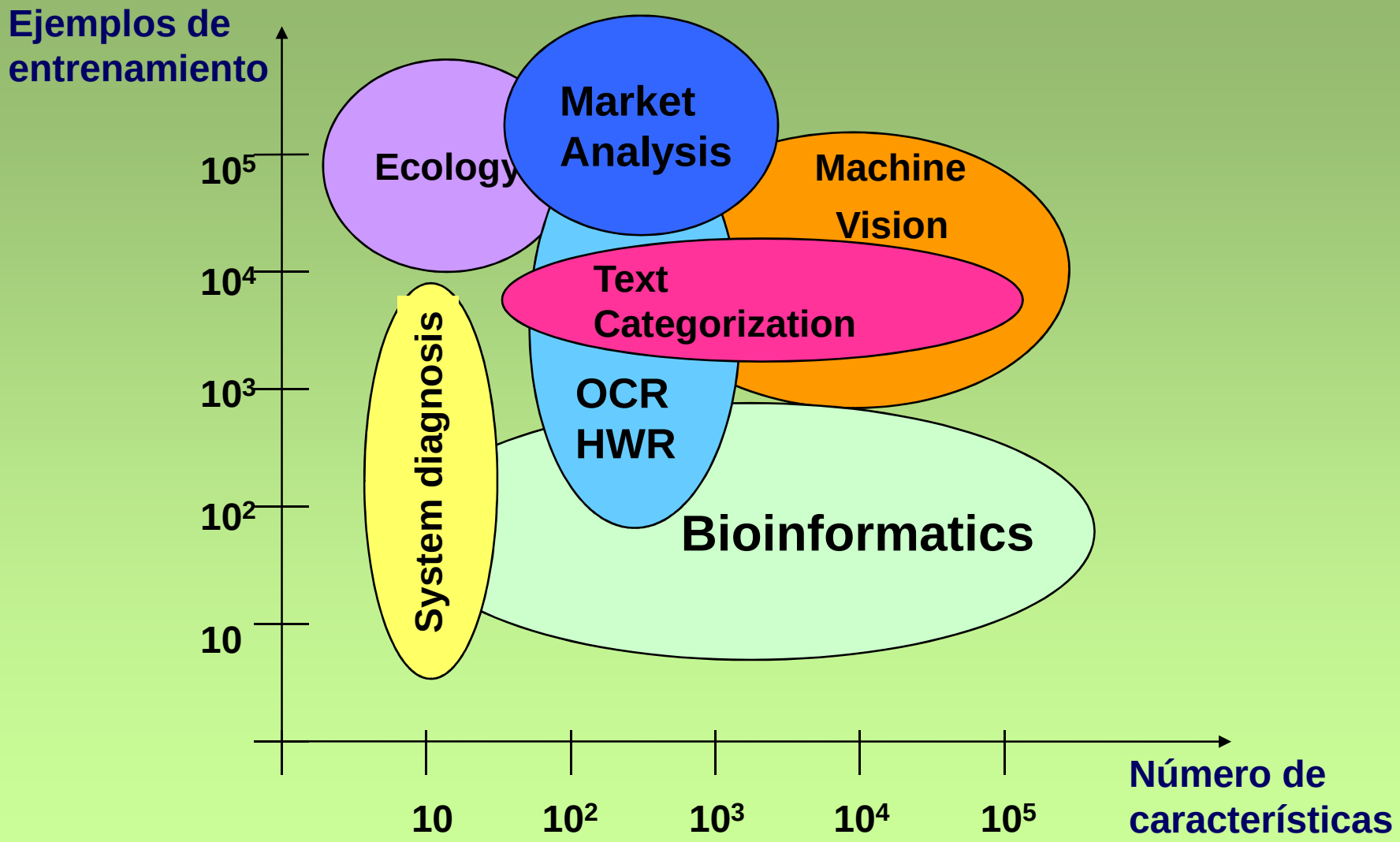
APLICACIONES

- **Banca / Telecomunicaciones / Ventas**

- **Identificar:**
  - **Futuros clientes**
  - **Clientes insatisfechos**
  - **Buenos clientes**
  - **Malos pagadores**
- **Obtener:**
  - **Mejoras en la efectividad de la publicidad**
  - **Menor riesgo de crédito**
  - **Reducir el fraude**

- **Interfaces de ordenador**
  - Detección de número de asistentes
  - Reconocimiento de escritura y de habla
  - Ondas cerebrales
- **Internet**
  - Resultados de clasificación
  - Detección de correo no deseado “spam”
  - Categorización de textos
  - Traducción de textos
  - Sistemas de recomendación

Aplicaciones



Hacia donde vamos

Hoy: La punta del iceberg

Redes neuronales superficiales y profundas

Árboles de decisión

Programación lógica inductiva

Regresión No lineal

Clasificación multietiqueta

Clasificación ordinal

Agrupamiento

. . .

- **Aplicados a bases de datos bien estructuradas**
- **Interés por parte de la industria. Industria 4.0**

Hacia donde vamos

Oportunidades para el futuro:

- **Aprendizaje a partir de datos de varias fuentes simultaneas: Bases de datos internas, web,. . .**
- **Aprendizaje por experimentación activa.**
- **Aprendizaje de decisiones en lugar de predicciones.**
- **Aprendizaje acumulativo de larga duración.**
- **¿Lenguajes de programación con aprendizaje incorporado?**

Disciplinas Relacionadas

- Análisis Matemático
- Inteligencia Artificial
- Métodos Bayesianos
- Teoría de la Complejidad
- Teoría de Control
- Teoría de la Información
- Filosofía
- Psicología y Neurobiología
- Estadística
- Investigación Operativa
- ….

Problema de aprendizaje

Aprendizaje: Mejora de alguna tarea mediante la experiencia

Tarea (*T*): Lo que se debe aprender

Experiencia (*E*): La que se tiene en relación a lo que se debe aprender

Rendimiento (*R*): Medida de la calidad de lo aprendido

Se dice que un sistema aprende de la experiencia, si el rendimiento *R* de la tarea *T* aprendida crece al crecer *E*.

Problema de aprendizaje

¿Con que experiencia hay que contar?

¿Que debemos aprender?

¿Como representamos el conocimiento?

¿Que algoritmo usaremos para aprenderlo?

¿Como se mide la mejora de lo aprendido?

Ejemplos

Una forma de aprender a jugar al ajedrez (de mejorar nuestro juego), es jugar contra nosotros mismos. Normalmente la forma de saber si hemos aprendido es jugar contra otros.

Caracterización de este problema de aprendizaje

***T*: Jugar al ajedrez**

***E*: Conjunto de partidas jugadas contra uno mismo**

***R*: Porcentaje de partidas ganadas contra otro jugador**

Ejemplos

Una forma de que un sistema aprenda a reconocer palabras en un texto manuscrito, puede ser a partir de una base de datos con imágenes de palabras manuscritas y sus correspondientes transcripciones.

La forma de saber si el sistema ha aprendido a reconocer palabras, sería darle un texto manuscrito y ver cuantas transcripciones correctas hace.

Caracterización de este problema de aprendizaje

- ***T*: Reconocer palabras manuscritas**
- ***E*:Base de datos de palabras con sus transcripciones**
- ***R*: Porcentaje de palabras reconocidas**

Ejemplos

Se puede enseñar a un vehículo a conducir automáticamente, guiándose por lo que ve (usando sensores de visión), y suministrándole una base de datos en la que se hayan registrado las imágenes tomadas mientras un conductor humano conducía el vehículo, junto con las correspondientes acciones que hizo.

Caracterización de este problema de aprendizaje

- ***T*: Conducir un vehículo**
- ***E*: Base de datos de imágenes, y las acciones correspondientes, registradas durante la conducción por parte de un conductor humano**
- ***R*: Distancia recorrida sin cometer ningún error**

¿Qué estudiar en A. A.?

- ¿Que algoritmos pueden aproximar funciones correctamente?
- ¿Como influye el numero de ejemplos en la exactitud?
- ¿Como influye la complejidad de la representación de las hipótesis?
- ¿Como influye el ruido?
- ¿Cuales son los límites teóricos del aprendizaje?
- ¿Como puede ayudar el conocimiento a priori?
- ¿Que esquemas del aprendizaje biológico podemos adoptar?
- ¿Como pueden los sistemas alterar su propia representación?

Bibliografía

Libros de texto

- C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.

- Tom M. Mitchell, Machine Learning. McGraw-Hill, 1997.

Libros de referencia

- R. O. Duda, P. E. Hart, D. G. Stork, Pattern classification, 2nd edition, John Wiley & Sons, 2001.

- E. Alpaydin, Introduction to Machine Learning, The MIT Press, 2004.

- J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

- Bernhard Schölkopf, Alexander J. Smola, Learning with Kernels: Support Vector Machine, Regularization, Optimization, and Beyond, The MIT Press, 2002.

- Vojislav Kecman, Learning and Soft computing, The MIT Press, 2001.

APRENDIZAJE: TERCER CURSO DEL GRADO DE ING. INFORMÁTICA EN COMPUTACION

Introducción al Aprendizaje Automático

GRACIAS POR SU ATENCIÓN

César Hervás-Martínez
Grupo de Investigación AYRNA

Departamento de Informática y Análisis Numérico
Universidad de Córdoba
Campus de Rabanales. Edificio Einstein.
Email: chervas@uco.es

Curso 2019-2020