

# PRÁCTICA II

---

## Aprendizaje Automático I

**Carmen Lozano | Borja Souto | Nina López | Carlos Hermida**  
**[Grupo 11C]**

—  
**2º GCID – Curso 2021/2022**



**UNIVERSIDADE DA CORUÑA**

## 1. INTRODUCCIÓN

Actualmente es muy común el uso de redes de neuronas artificiales para la detección de objetos en imágenes. En este trabajo, nos dedicaremos en concreto a la detección de ojos.

Para los humanos, una de las maneras más sencillas de reconocer a una persona, es mediante el rostro. Es por esto por lo que este problema puede ayudar a reconocer a personas en una imagen, ya que nuestro ojo es muy característico de nuestra especie, y no muy parecido al de otras. Además, no se parece demasiado a ningún otro objeto cotidiano.

En concreto en esta práctica intentaremos que un método de aprendizaje supervisado consiga reconocer ojos en imágenes de primeros planos de personas, basándose en características aportadas por nosotros.

## 2. DESCRIPCIÓN DEL PROBLEMA

Como ya introdujimos en el apartado anterior en esta práctica lo que buscamos es encontrar características en una imagen que encuadre a un ojo, de manera que pasándole estas a una red de neuronas artificiales, o una máquina de soporte vectorial o cualquier método del estilo, sea capaz de reconocer el ojo en una imagen de una persona.

Para entrenar a la red empleamos 30 ventanas (es decir, recortes de imágenes más grandes de aproximadamente 50 píxeles de ancho y 25 de alto) en las que aparecen ojos encuadrados, a las que llamaremos ventanas positivas; y otras 30 del mismo tamaño aproximadamente, en las que aparecen partes de imágenes que claramente no son ojos.

De las ventanas positivas, tratamos de sacar características que pensamos que podrían ayudar a distinguir un ojo de entre otros objetos. Sobre todo, utilizamos datos relacionados con el color de las imágenes, así como de la distribución de estos en la ventana o la simetría.

Para comprobar finalmente el correcto funcionamiento de nuestro trabajo emplearemos imágenes rectangulares que son primeros planos de personas, y todas del mismo tamaño y orientación (horizontales), con una iluminación parecida entre todas ellas y con un fondo sin mucho detalle (una pared, el fondo de una habitación...), y en las que la persona destaca claramente.

Estas imágenes se recorrerán por ventanas rectangulares de los tamaños de las que usamos para entrenar y cuando se encuentre un ojo, la ventana se resalta en rojo

## 3. ANÁLISIS BIBLIOGRÁFICO

La detección de objetos en imágenes basándonos en características es algo que se ha usado mucho en tiempos recientes. Se puede aplicar a diferentes tareas y diversos ámbitos. Por ejemplo, en el estudio que realizó (Symeonidis et al., 2019) usa redes de neuronas artificiales para detectar semáforos y en el estudio realizado por (Shi et al., 2017) las usa para detectar coches.

El reconocimiento facial es un problema complicado dentro del reconocimiento de objetos ya que el rostro se caracteriza por tener diferencias individuales que pueden ser la base para la verificación de la identidad, (Fernando et al., 2019) y (Shi et al., 2020) realizaron estudios en los que se detectan caras con redes de neuronas artificiales y máquinas de soporte vectorial respectivamente. Otros autores como (Pérez-Gaspar et al., 2015) utilizaron redes de neuronas artificiales para detectar las emociones de las personas.

## 4. DESARROLLO

El problema que se nos plantea es: dada una imagen localizar con distintos tamaños de ventanas los ojos en esa imagen. Para resolver este problema cogimos distintas características de las ventanas de entrenamiento y las combinamos de distintas maneras para entrenar la red.

Antes de pasarle las características a la red las normalizamos entre el máximo y el mínimo ya que todos los valores se encuentran en el mismo rango porque están acotados

Para evaluar la eficiencia de todas las características de cada aproximación usamos el método de validación cruzada con  $k=10$  para evaluar la eficiencia, es decir dividimos los datos de la muestra en 10 subconjuntos, uno de estos lo usamos como prueba y el resto los usamos como datos de entrenamiento.

La red devuelve la probabilidad de ser un ojo, con un valor entre  $[0,1]$ . Consideramos que se trata de un ojo cuando este valor es superior a 0.9.

### 4.1. Aproximación 1

Para comenzar creamos una función a la que llamamos *featureExtraction1* para la que utilizamos las características que se nos propusieron en el enunciado de la práctica, es decir, la media y desviación típica de cada uno de los colores de la imagen: rojo, verde, y azul. Y, además añadimos lo mismo pero para el caso de una imagen en blanco y negro.

En ambos casos, para que la iluminación de la imagen no influya en las características obtenidas, lo que hacemos es dividir toda la matriz por el promedio de sus valores antes de obtener las métricas que vamos a usar.

Lo primero es pasar la imagen a una matriz tridimensional, donde cada capa de la matriz representa un color y guarda, para cada píxel, la cantidad del color correspondiente. Lo que hacemos es quedarnos con cada capa y hacer la media y desviación típica de todos sus elementos y añadir estas métricas a una matriz fila, de manera que ya tendremos 6 características.

A continuación, transformamos la misma imagen a una matriz bidimensional, en la que para cada píxel guarda la cantidad de blanco (valores cercanos a 0 son casi negros, mientras que los valores más altos con cercanos a blanco). De nuevo, obtenemos media y desviación típica de esta matriz, lo que nos aporta 2 características más que añadimos a la matriz creada anteriormente.

Una vez tenemos todas las características evaluamos su eficiencia aplicando validación cruzada en una red de neuronas artificial con topología [4] que significa que solo tiene una capa con cuatro neuronas y obtuvimos los siguientes resultados *Figura 1*.

FOLD	ACCURACY (EN %)	F1 SCORE (EN %)
1/10	66.67	80.0
2/10	99.33	99.60
3/10	100	100
4/10	62	75.80
5/10	100	100
6/10	63.33	77
7/10	88.77	93.20
8/10	52.67	66.0
9/10	100	100
10/10	92.67	95.6

MEDIA ACCURACY	MEDIA F1 SCORE	DESVIACIÓN TÍPICA
82.53	-	19.07
-	88.72	12.75

*Figura 1. Tablas de los resultados de validación cruzada con las características de la primera aproximación*

Como la media de la precisión y del F1-score, tras emplear el método de validación cruzada, superan el 80%, utilizamos estas características para entrenar una red con la misma topología, y comprobamos en las imágenes de prueba el funcionamiento. En la *Figura 2* se muestran los resultados que obtuvo. Este gráfico muestra la cantidad de ojos que la red reconoce como tales de los 24 que hay en las 12 imágenes de prueba.



*Figura 2. Resultados obtenidos por la RNA con la primera aproximación.*

Según este gráfico, podría parecer que lo hace bien, pero como podremos ver en algunas de los ejemplos siguientes, tiene demasiados falsos positivos como para considerar esta una buena aproximación. Llama la atención que reconoce todo aquello que es oscuro.

Algunas de las imágenes obtenidas son las siguientes:



*Figura 3. Ejemplos de los resultados obtenidos con la primera aproximación*

#### 4.2. Aproximación 2

Para esta segunda aproximación hemos cogido las características de la aproximación 1 y les hemos incluido dos nuevas características, la media y la desviación típica de la simetría tanto vertical y horizontal ya que un ojo es prácticamente simétrico.

Para tener en cuenta la simetría creamos una función a la que le pasamos una imagen en forma de matriz y la giramos según el eje vertical u horizontal según el caso. Una vez tenemos las matrices giradas calculamos las diferencias respecto a la matriz original y calculamos la media y la desviación típica de las diferencias.

Una vez tenemos todas las características, para evaluar su eficiencia usamos el método de validación cruzada con una red de neuronas artificial y topología [2, 2] (2 capas ocultas con 2 neuronas cada una) y obtuvimos los siguientes resultados *Figura 4*.

FOLD	ACCURACY (EN %)	F1 SCORE (EN %)
1/10	100	100
2/10	100	100
3/10	100	100
4/10	66.66	80
5/10	70.66	77.8
6/10	86.66	92
7/10	100	100
8/10	70.0	82.0
9/10	100	100
10/10	100	100

MEDIA ACCURACY	MEDIA F1 SCORE	DESVIACIÓN TÍPICA
89.40	-	14.62
-	93.18	9.52

*Figura 4. Tablas de los resultados de validación cruzada con las características de la primera aproximación*

Como la media y la desviación típica de la precisión y de F1-score son aproximadamente del 90% hemos usado estas características para entrenar una red y comprobamos en las imágenes de test su funcionamiento.



*Figura 5. Resultados obtenidos por la RNA con la segunda aproximación.*

Como podemos ver en este gráfico reconoce tan solo la mitad de los ojos de las imágenes a pesar de que la media de la precisión y del F1-score son mejores que en la primera aproximación, esto se puede deber a que estamos entrenando con demasiados atributos o atributos que no son relevantes.

En las imágenes resultantes, es cierto que se conseguían menos falsos positivos que en la anterior aproximación, pero el número de ojos que reconocía es también mucho menor, tal vez porque algunos de ellos son muy simétricos en un eje, pero no tanto en otros. Que haya menos falsos positivo se puede deber a que las ventanas tienen que cumplir la simetría por colores, pero también en blanco y negro. Es por esto que decidimos hacer una ligera modificación que mostraremos en el apartado de la tercera aproximación.

#### 4.3. Aproximación 3

En esta aproximación usamos las mismas características que en la segunda aproximación, pero le quitamos la simetría vertical dado que consideramos que es más relevante la simetría horizontal.

Dado que esta aproximación es muy similar a la anterior, entrenamos directamente la red con estas características y topología [2, 2], comprobamos los resultados en las imágenes de test y obtuvimos los siguientes resultados *Figura 6*.



*Figura 6. Resultados obtenidos por la RNA con la tercera aproximación.*

Como se puede ver en las siguientes imágenes de la Figura 7. Sigue cogiendo muchas partes de la imagen que no contienen ojos, pero en comparación con la primera aproximación, son muchas menos, y la cantidad de ojos que reconoce sigue siendo muy buena.

Si nos fijamos podemos ver que coge como falsos positivos, sobre todo bordes y las partes más oscuras de las imágenes

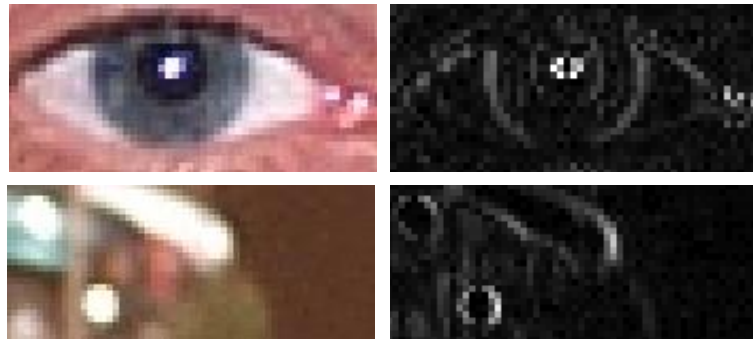


*Figura 7. Ejemplos de los resultados obtenidos con la tercera aproximación*

#### 4.4. Aproximación 4

Para esta aproximación las características que cogemos son la media y desviación típica de los colores, la simetría según ambos ejes solo para los colores, así como la media de las diferencias entre un píxel y el de al lado.

Para sacar esta nueva característica, implementamos una función que para cada fila mira la diferencia entre los valores de un píxel y el de al lado de una imagen en blanco y negro, y las métricas que saca son la media de esas diferencias de las tres filas superiores, las tres filas inferiores y las tres del medio



*Figura 8. Ejemplos del uso de la función de diferencias*

Una vez tenemos todas las características se las pasamos a la red para probar su eficiencia mediante el método de validación cruzada usando una red de neuronas artificial con topología [2, 2] y obtuvimos los siguientes resultados *Figura 8*.

FOLD	ACCURACY (EN %)	F1 SCORE (EN %)
1/10	100	100
2/10	100	100
3/10	100	100
4/10	100	100
5/10	100	100
6/10	100	100
7/10	100	100
8/10	100	100
9/10	100	100
10/10	100	100

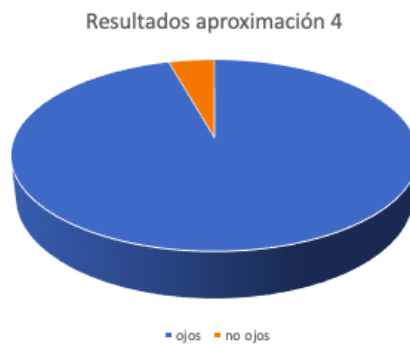
  

MEDIA ACCURACY	MEDIA F1 SCORE	DESVIACIÓN TÍPICA
100	-	0
-	100	0

*Figura 9. Tablas de los resultados de validación cruzada con las características de la cuarta aproximación*

Dado que los resultados parecen ser muy buenos hemos usado estas características para entrenar una red y comprobamos en las imágenes de prueba su funcionamiento.





*Figura 10. Resultados obtenidos por la RNA con la cuarta aproximación.*

Como podemos ver en el gráfico de la *Figura 9*, localiza prácticamente todos los ojos de las imágenes a pesar de que sigue cogiendo ventanas erróneas, en la *Figura 10*, podemos ver algunos de los resultados obtenidos.



*Figura 11. Ejemplos de los resultados obtenidos con la cuarta aproximación*

#### 4.5. Aproximación 5

Para esta aproximación nos quedamos con las características de la media y la desviación típica de los colores y añadimos la media y la desviación típica de las subventanas de las imágenes en blanco y negro.

Para esta nueva característica creamos una función que divide las ventanas en seis partes y calculamos la media y la desviación típica de la cantidad de blanco en todas ellas.

Una vez tenemos las nuevas características se las pasamos a la red para probar su eficiencia mediante el método de validación cruzada usando una red de neuronas artificial con topología [2, 2] y obtuvimos los siguientes resultados *Figura 11*.

FOLD	ACCURACY (EN %)	F1 SCORE (EN %)
1/10	100	100
2/10	100	100
3/10	66.67	80.0
4/10	100	100
5/10	100	100
6/10	99.33	99.6
7/10	100	100
8/10	100	100
9/10	100	100
10/10	100	100

MEDIA ACCURACY	MEDIA F1 SCORE	DESVIACIÓN TÍPICA
96.6	-	10.52
-	97.96	6.31

*Figura 12. Tablas de los resultados de validación cruzada con las características de la quinta aproximación*

A pesar de que los resultados parece que van a ser muy buenos, apenas selecciona ventanas, en alguna de las que selecciona coincide que hay un ojo, pero en otras muchas no. Esto se puede deber a que le estamos pasado datos poco relevantes o demasiado específicos.



*Figura 13. Ejemplos de los resultados obtenidos con la quinta aproximación*

#### 4.6. Aproximación 6

Para esta aproximación usamos características similares a la aproximación cinco, nos quedamos con la media y desviación típica de los colores, calculamos la media y desviación típica de la segunda y la quinta subventana en blanco y negro ya que son las que deberían de tener mayor cantidad de blanco (y las que son más parecidas en todos los ojos que miran de frente) y añadimos la simetría vertical y horizontal de los colores rojo, verde y azul.

Le pasamos estas nuevas características a la red para probar su eficiencia mediante el método de validación cruzada usando una red de neuronas artificial con topología [2, 2] como en aproximaciones anteriores y obtuvimos los siguientes resultados *Figura 13*.

FOLD	ACCURACY (EN %)	F1 SCORE (EN %)
1/10	100	100
2/10	67.33	80.4
3/10	67.33	80.4
4/10	96.28	96.28
5/10	97	96.74
6/10	80.88	78
7/10	86.66	84.35
8/10	95.11	93.71
9/10	87	89.6
10/10	79.33	89.6

MEDIA ACCURACY	MEDIA F1 SCORE	DESVIACIÓN TÍPICA
85.69	-	11.86
-	88.71	7.80

*Figura 14. Tablas de los resultados de validación cruzada con las características de la sexta aproximación*

Probamos a entrenar una red con estas características y a pesar de quitar características y añadir la simetría nos da unos resultados peores a los obtenidos anteriormente.

No adjuntamos imágenes ya que los resultados son peores a los obtenidos anteriormente. A mayores también probamos a introducir solo la simetría horizontal y nos seleccionaba más verdaderos positivos (ventanas con ojos) pero también nos seleccionaba otros muchos falsos positivos (ventanas sin ojos).

## 5. CONCLUSIONES

Tras observar todos los resultados anteriores, concluimos que el mejor conjunto de características es el de la aproximación 4.

En esta usábamos la media y desviación típica de los colores (rojo, verde y azul), así como la media y desviación típica de la simetría de estas capas. Para la imagen en blanco y negro obteníamos la media de las diferencias entre un píxel y el de al lado de las tres filas superiores, las tres inferiores y las tres centrales.

Consideramos la cantidad de color de cada capa y no la cantidad de blanco, ya que los colores en una ventana positiva son característicos, por el color de los ojos, pero también por el de la piel y la parte blanca del ojo.

Además, al mirar las diferencias entre cada píxel nos quedamos con la información de esas franjas ya que por ejemplo la del medio más o menos coincidirá en todas porque es por donde está la pupila y también el borde del iris y el lagrimal, y en las de arriba y abajo es más probable que no haya tantos “bordes” porque hay piel, o que haya los “bordes” correspondientes a las pestañas.

A continuación, mostramos cual fue el resultado del entrenamiento de esta red con estas características en el conjunto de prueba:



Vistos nuestros resultados, pensamos que una red de neuronas artificiales es un buen método para resolver el problema que se nos ha planteado, ya que obtenemos resultados bastante buenos (con un estudio más profesional podrían haber sido muy buenos o prácticamente perfectos).

Lo más difícil, creemos que es encontrar un conjunto de características propias de un ojo y que sirvan para identificarlo de cualquier otra cosa (es decir, que sea poco probable que las presenten todas en conjunto otras ventanas/objetos), y que además no sean demasiado específicas como para que solamente reconozcan cierto tipo de ojos (de un color, de cierta forma ...)

## 6. TRABAJO FUTURO

Teniendo en cuenta el objetivo de esta práctica, que era encontrar características de un ojo que permitieran a una RNA reconocerlos en imágenes de personas, estamos bastante contentos con los resultados, ya que son todas características obtenidas “manualmente” por nosotros.

Pensamos que es posible que la red pudiera reconocer los ojos con más precisión, o detectar menos falsos positivos con características más detalladas obtenidas mediante métodos más elaborados (o con librerías pensadas para procesamiento de imágenes), pero esto, de momento, no entra dentro de los conocimientos que tenemos de la materia.

Además, pensamos que las muestras de entrenamiento no son del todo adecuadas ya que, en el caso de las ventanas positivas, no cumple que la muestra sea representativa, ya que todos los ojos son del mismo color, que es el azul, y luego en el conjunto de test, muchas de las personas de las imágenes tienen ojos oscuros o no tan claros como los de entrenamiento. Además, las formas de los ojos del conjunto de test son muy diferentes entre ellas, más de lo que lo son las de entrenamiento.

Por lo tanto, pensamos que añadiendo más variedad de ejemplos en el conjunto de entrenamiento podría salir mejor.

Ya que en este trabajo nos ceñimos al uso de redes de neuronas artificiales, para trabajos futuros podríamos probar a centrarnos más en otros métodos como máquinas de soporte vectorial, árboles de decisión, o el método de los k vecinos más cercanos.

Este tipo de modelo que resuelva correctamente este problema, creemos que podría ser útil en la vida real, para el reconocimiento facial de los móviles o para ayudar en el CGI (Computer Generated Imagery) que se usa por ejemplo en el cine para efectos especiales.

## 7. REFERENCIAS

- Fernando, E., Andriyan, D., Fitria Murad, D., Touriano, D., & Irsan, M. (2019). Face recognition system Using Deep Neural Network with Convolutional Neural Networks. *Journal of Physics: Conference Series*, 1235(1). <https://doi.org/10.1088/1742-6596/1235/1/012004>
- Pérez-Gaspar, L.-A., Caballero-Morales, S.-O., & Trujillo-Romero, F. (2015). Factores en el reconocimiento facial de emociones y la integración de optimización evolutiva. *Research in Computing Science*, 91(1). <https://doi.org/10.13053/racs-91-1-4>
- Shi, L., Wang, X., & Shen, Y. (2020). Research on 3D face recognition method based on LBP and SVM. *Optik*, 220. <https://doi.org/10.1016/j.ijleo.2020.165157>
- Shi, L., Wang, Y., Cao, Y., & Wei, L. (2017). Vehicle type recognition based on deep convolution neural network. *Communications in Computer and Information Science*, 728. [https://doi.org/10.1007/978-981-10-6388-6\\_42](https://doi.org/10.1007/978-981-10-6388-6_42)
- Symeonidis, G., Groumpos, P. P., & Dermatas, E. (2019). Traffic Light Detection and Recognition Using Image Processing and Convolution Neural Networks. *Communications in Computer and Information Science*, 1084, 181–190. [https://doi.org/10.1007/978-3-030-29750-3\\_14](https://doi.org/10.1007/978-3-030-29750-3_14)