

EJERCICIO 1

Sobre una cinta transportadora circulan naranjas y melones. Se busca obtener un clasificador de frutas que facilite su almacenamiento. Para cada fruta, se conoce su diámetro en centímetros y la intensidad de color naranja medida (entre 0% y 100%). Se recabaron las siguientes muestras como ejemplos de cada tipo de fruta:

Diámetro	10	8	7	15	20	26	24	19
Intensidad	75 %	60 %	67 %	95 %	11 %	13 %	12.5 %	12 %
Tipo	Naranja	Naranja	Naranja	Naranja	Melón	Melón	Melón	Melón

Entrene un perceptrón que permita resolver el problema y utilícelo para clasificar los siguientes ejemplos, verificando que los resultados obtenidos sean los correctos:

$X = [[12, 80], [23, 15], [27, 10], [11, 80]]$

$T = ['Naranja', 'Melón', 'Melón', 'Naranja']$

EJERCICIO 2

Se ha realizado un análisis químico a tres tipos distintos de vinos producidos en una misma región de Italia. Se analizaron 59 muestras del Tipo 1, 71 muestras del Tipo 2 y 48 muestras del Tipo 3. El archivo 'vinos.npy' permite observar los resultados de este análisis. Cada fila representa una muestra distinta y está formada, en primer lugar, por el número del tipo al cual pertenece el vino analizado seguido por los 13 atributos que lo caracterizan. Por ejemplo, la siguiente fila (`vinos[80]`):

```
array([ 2. , 12. , 0.92, 2. , 19. , 86. , 2.42,
        2.26, 0.3 , 1.43, 2.5 , 1.38, 3.12, 278. ])
```

es el resultado del análisis de un vino del Tipo 2 (1er. valor de la fila) seguido por 13 valores separados por comas que indican los niveles de las mediciones realizadas a dicho vino.

Entrene un perceptrón para clasificar los vinos del Tipo 1 utilizando un porcentaje de las muestras existentes. Los patrones que no se utilicen en el entrenamiento serán empleados para probarlo. Realice el mismo procedimiento para los vinos del Tipo 2 y del Tipo 3.

Detalle la cantidad de pasos que fueron necesarios para lograr la mejor clasificación. Pruebe variar el tamaño del conjunto de patrones de entrenamiento observando para cada caso la proporción de patrones correctamente clasificados por el perceptrón.

EJERCICIO 3

Retomando el problema de la clasificación de vinos a partir de su análisis químico, obtenga un clasificador empleando una red neuronal artificial entregada mediante el algoritmo de backpropagation.

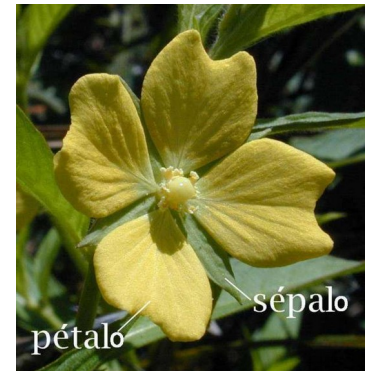
El archivo 'vinos.npy' contiene los datos del análisis químico de 178 vinos. Cada fila representa una muestra distinta y está formada, en primer lugar, por el número del tipo al cual pertenece el vino analizado seguido por los 13 atributos que lo caracterizan.

El archivo 'vinos.entrenamiento.npy' contiene un subconjunto de las muestras, formado por el 90% de las muestras de cada tipo. Pueden usarse estos datos para el entrenamiento y el conjunto completo de las muestras para medir el desempeño del clasificador.

Detalle la estructura de la red neuronal empleada (cantidad de entradas, de salidas y de neuronas en la capa oculta) e informe la velocidad de aprendizaje usada, la cantidad de pasos necesarios para lograr la mejor clasificación y el error cometido al final del entrenamiento. Construya la matriz de confusión producida usando todos los patrones (no solo los de entrenamiento).

EJERCICIO 4

El conjunto de datos "Iris" ha sido usado como caso de prueba para una gran cantidad de clasificadores y es, quizás, el conjunto de datos más conocido de la literatura específica. Iris es una variedad de planta que se la desea clasificar de acuerdo a su tipo. Se reconocen tres tipos distintos: '*Iris setosa*', '*Iris versicolor*' e '*Iris virgínica*'. El objetivo es lograr clasificar una planta de la variedad Iris a partir del largo y del ancho del pétalo y del largo y del ancho del sépalo.



El conjunto de datos *Iris* está formado en total por 150 muestras, siendo 50 de cada uno de los tres tipos de plantas. Cada muestra está compuesta por el tipo de planta, la longitud y ancho del pétalo y la longitud y ancho del sépalo. Todos son atributos numéricos continuos.

a) Entrene perceptrones para que cada uno aprenda a reconocer uno de los distintos tipos de plantas Iris. Informe los parámetros usados para el entrenamiento y el desempeño obtenido. Emplee todos los patrones para el entrenamiento. Muestre la matriz de confusión para la mejor clasificación obtenida luego del entrenamiento, informando los patrones clasificados correcta e incorrectamente.

b) Entrene una red neuronal artificial usando backpropagation como algoritmo de aprendizaje con el fin de lograr la clasificación pedida. Emplee todos los patrones para el entrenamiento. Detalle los parámetros usados para el entrenamiento así como la arquitectura de la red neuronal. Repita más de una vez el procedimiento para confirmar los resultados obtenidos e informe la matriz de confusión para la mejor clasificación obtenida.

X	Setosa	Versicolor	Virgínica	Inválidas
Setosa	50	0	0	0
Versicolor	0	50	0	0
Virgínica	0	0	50	0

Matriz de confusión ideal: Sin errores en la clasificación.

EJERCICIO 5

El hormigón es el material más importante en lo que a la ingeniería civil concierne. La resistencia a compresión del hormigón resulta de fundamental interés ya que permite asegurar cuanta carga podrá soportar una mezcla de hormigón particular. La resistencia a compresión depende de la edad del hormigón y de la mezcla de materiales usados para producirlo. La resistencia a compresión del hormigón se obtiene en ensayos de rotura a compresión de probetas cilíndricas normalizadas realizados a los 28 días de edad y fabricadas con las mismas amasadas puestas en obra.

Para tratar de modelar la resistencia a compresión del hormigón a partir de la mezcla que lo constituye y la edad del mismo, se realizaron sucesivas pruebas en laboratorio con distintas proporciones y en diferentes estadías.

En total se usaron 1030 muestras, sabiendo su edad (en días) y analizando que cantidad de siete componentes diferentes (cemento, agua, cenizas volante, etc.) las constituye, siendo todos estos atributos valores reales continuos. Para cada una de las muestras se obtuvo su resistencia a compresión.

Entrene una red neuronal artificial para tratar de predecir cual será la resistencia a compresión de una muestra de hormigón a partir de su composición, conociendo su edad y que cantidad de cada uno de los siete materiales analizados la constituye. Utilice todas las muestras para el entrenamiento. Detalle la arquitectura empleada y los valores de los parámetros usados. Documente todos los intentos realizados. Informe los mejores resultados obtenidos.

EJERCICIO 6

Con el fin de predecir la puntuación que un catador profesional le otorgará a un determinado vino, se tomaron resultados de análisis físico-químicos de rutina realizados durante la producción de los vinos para tratar de descubrir de que manera estos factores influyen en la puntuación obtenida.

Construir un modelo de este tipo resulta beneficioso no solo para anticipar la puntuación que un enólogo otorgará al vino sino también para poder analizar como la variación de determinadas características afectará, tanto positiva como negativamente, la calidad percibida del vino antes de que éste sea producido.

La bodega portuguesa "Vinho Verde" ha facilitado los resultados de analizar 1599 muestras de varios tipos de vino tinto producidos por esa bodega, junto con las calificaciones que cada uno de ellos obtuvo, las cuales van de 0 a 10, siendo 0 la peor puntuación posible y 10, la mejor. De cada vino se analizaron 11 atributos diferentes (pH, densidad y grado de alcohol, para nombrar sólo algunos), siendo todos estos atributos valores reales continuos.

Entrene una red neuronal artificial mediante backpropagation para que sea capaz de predecir cual será la puntuación que un vino obtendrá a partir de los 11 atributos analizados. Detalle la arquitectura usada y los parámetros empleados en su entrenamiento. Reserve un subconjunto razonable de los patrones disponible para evaluar el desempeño de la red luego de ser entrenada, informando el error obtenido.

EJERCICIO 7

A partir de análisis clínicos y de la edad y el sexo de pacientes de una clínica ubicada en el noreste de Andhra Pradesh, India, se intentará obtener un clasificador automático que sirva para diagnosticar a pacientes con problemas de hígado.

Para esto, se recabaron muestras de ocho análisis distintos realizados a 579 pacientes que, junto con su edad y sexo, se dividieron en dos grupos: 414 de ellos diagnosticados con problemas de hígado por expertos en el área mientras que los 165 restantes fueron señalados como exentos de ese problema.

Los 11 atributos que constituyen una muestra son los indicados en la tabla de la derecha. Todos son atributos son valores numéricos continuos a excepción del atributo "Sexo", en donde el valor 1 representa "HOMBRE" y el valor 2 representa "MUJER", y del atributo "Diagnóstico", donde el valor 1 representa "CON PROBLEMA DE HÍGADO" mientras que el valor 2 representa "SIN PROBLEMA DE HÍGADO".

1	Edad
2	Sexo
3	Bilirrubina Total
4	Bilirrubina Directa
5	Fosfatasa Alcalina
6	Alanina Aminotransferasa
7	Aspartato Aminotransferasa
8	Proteínas Total
9	Albúmina
10	Relación Albúmina/Globulina
11	Diagnóstico (<i>valor a predecir</i>)

Utilice perceptrones o una red neuronal artificial (según resulte más conveniente). Informe el motivo por el que se eligió el tipo de clasificador. Detalle la arquitectura y los parámetros usados en su entrenamiento (según corresponda). Documente todos los intentos realizados.

Para el entrenamiento emplee sólo el 90% de las muestras disponibles de cada tipo. Informe la matriz de confusión que produce el mejor clasificador obtenido al evaluarlo con las muestras de entrenamiento e indique la matriz que ese clasificador produce al usarlo sobre el resto de las muestras reservadas para prueba.

EJERCICIO 8

El trigo es uno de los tres granos más ampliamente producidos globalmente, junto al maíz y el arroz, y el más ampliamente consumido por el hombre en la civilización occidental desde la antigüedad. El grano de trigo es utilizado para hacer harina, harina integral, sémola, cerveza y una gran variedad de productos alimenticios.

Se requiere clasificar semillas de trigo pertenecientes a las variedades Kama, Rosa y Canadiense.

Se cuenta con 70 muestras de cada una de las variedades, a cuyas semillas se le realizaron mediciones de diferentes propiedades geométricas: Área, perímetro, compacidad, largo, ancho, coeficiente de asimetría, largo del carpelo (todos valores reales continuos).

Utilice perceptrones o una red neuronal artificial (según resulte más conveniente) para lograr producir un clasificador de los tres tipos de semillas de trigo a partir de las muestras obtenidas. Informe el criterio empleado para decidir el tipo de clasificador entrenado y la arquitectura y los parámetros usados en su entrenamiento (según corresponda).

Utilice para el entrenamiento sólo el 90% de las muestras disponibles de cada variedad. Informe la matriz de confusión que produce el mejor clasificador obtenido al evaluarlo con las muestras de entrenamiento e indique la matriz que ese clasificador produce al usarlo sobre el resto de las muestras reservadas para prueba.

EJERCICIO 9

La segmentación de imágenes digitales en regiones con apariencia de piel ha sido utilizado como un paso de pre-procesamiento en varias aplicaciones de detección y seguimiento de caras humanas y personas.

Con el fin de generar un detector de piel/no-piel, se realizó un muestreo de fotografías digitales de caras de personas de diversas edades, géneros y razas al mismo tiempo que se hizo otro muestreo de imágenes que no incluyeran personas. El muestreo consistió en obtener mediciones del color de pixels de cada una de las clases requeridas (piel/no-piel). La medición del color se realizó en el espacio de colores RGB, obteniendo un total de 51444 muestras de las cuales 14654 son muestras de piel mientras que las restantes 36790 son muestras de otras texturas. Cada muestra está representada por el nivel de azul (B), verde (G) y rojo (R), siendo el nivel un valor entero entre 0 (mínimo nivel posible) y 255 (máximo nivel posible).

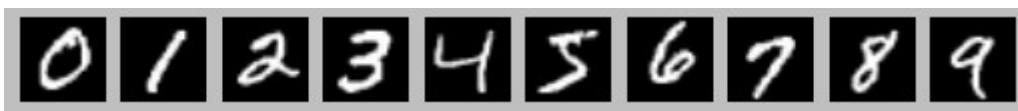
Entrene una red neuronal artificial para tratar de predecir a partir del color de un pixel dado si se trata o no de piel humana. Utilice todas las muestras para el entrenamiento.

Detalle la arquitectura empleada y los valores de los parámetros usados. Documente todos los intentos realizados. Informe los mejores resultados obtenidos. Utilice la imagen suministrada junto con el enunciado para validar la efectividad del clasificador.

EJERCICIO 10

La base de datos de dígitos manuscritos MNIST es un buen ejemplo de un conjunto de datos del mundo real que no requiere que uno se preocupe por el preprocesamiento de los mismos para comenzar a trabajar en su clasificación.

Esta base de datos cuenta con un conjunto de dígitos manuscritos producidos por una gran cantidad de personas diferentes. Estos dígitos han sido escaneados a partir de documentos impresos, y los dígitos han sido centrados y escalados a un tamaño de 28x28 pixeles usando 256 tonos de grises, de forma de hacerlos comparables entre ellos. Para este ejercicio, se cuenta con un subconjunto de la base de datos original, compuesto por 5000 patrones de entrenamiento y 1000 patrones independientes para validar la clasificación.



Ejemplos de dígitos a reconocer

Entrene una red neuronal artificial mediante backpropagation para que sea capaz de identificar cada uno de los dígitos manuscritos. Detalle la arquitectura usada y los parámetros empleados en su entrenamiento. Informe la matriz de confusión una vez realizado el entrenamiento y muestre ejemplos de imágenes a las que no se pudo reconocer y ejemplos donde la identificación fue incorrecta, ya sea porque se identificó el dígito equivocado (era un 4 y se reconoció un 7) o la red respondió reconociendo más de un único dígito (por ejemplo, era un 3 y se reconoció tanto al un 3 como a un 8).