|  |
| --- |
| Asignatura: |
| Inteligencia Artificial II |

|  |
| --- |
| Título del documento: |
| **Laboratorio 1:**  **Implementación de un SOM** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | Fecha |
| Grupo: |  | 2 |  | 12/02/18 |
|  |  |  |  |  |
| Miembros: | 1- | Gonzalo de las Heras |  |  |
|  |  |  |  |  |
|  | 2- | Jorge de la Fuente |  |  |
|  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nombre de fichero: |  | Fecha: |  |  |  |  |
| LAB01-GRUPO02-MEMORIA.docx **(.pdf)** |  | **13/02/18** |  |  |  |  |

Índice

[1 Objetivo del Laboratorio 1 6](#_Toc506322980)

[2 Resultados Práctica 1 (Colores) 7](#_Toc506322981)

[3 Resultados Práctica 2 (Números escritos a mano) 10](#_Toc506322982)

[4 Discusión general 13](#_Toc506322983)

[4.1 Proceso de búsqueda del mapa óptimo 13](#_Toc506322984)

[4.1.1 Procedimiento 13](#_Toc506322985)

[4.2 Mapa Práctica 1 (Colores) 14](#_Toc506322986)

[4.2.1 Conjunto de entrenamiento 14](#_Toc506322987)

[4.2.2 Lado del mapa (L) óptimo 14](#_Toc506322988)

[4.2.3 Periodo (P) óptimo 14](#_Toc506322989)

[4.2.4 Resultados 14](#_Toc506322990)

[4.2.5 Prueba de clasificación 16](#_Toc506322991)

[4.3 Mapa Práctica 2 (Números escritos a mano) 18](#_Toc506322992)

[4.3.1 Conjunto de entrenamiento 18](#_Toc506322993)

[4.3.2 Lado del mapa (L) óptimo 18](#_Toc506322994)

[4.3.3 Periodo (P) óptimo 18](#_Toc506322995)

[4.3.4 Resultados 18](#_Toc506322996)

[4.3.5 Prueba de clasificación 19](#_Toc506322997)

[5 Reflexiones finales 21](#_Toc506322998)

[6 Bibliografía 22](#_Toc506322999)

[7 ANEXO A. Documentos adjuntados 23](#_Toc506323000)

[8 ANEXO B. Números escritos a mano, conjunto completo 24](#_Toc506323001)

[8.1 Conjunto de entrenamiento 24](#_Toc506323002)

[8.2 Resultados 24](#_Toc506323003)

Índice de ilustraciones

[Ilustración 2‑1: Variación de distancia media con respecto al lado del mapa de Kohonen (colores). 7](#_Toc506322962)

[Ilustración 2‑2: Variación de la distancia media con respecto al periodo (colores). 8](#_Toc506322963)

[Ilustración 3‑1: Variación de distancia media con respecto al lado del mapa de Kohonen (números). 10](#_Toc506322964)

[Ilustración 3‑2: Variación de la distancia media con respecto al periodo (números). 11](#_Toc506322965)

[Ilustración 4‑1: Mapa (colores) inicial sin entrenar. 14](#_Toc506322966)

[Ilustración 4‑2: Mapa (colores) entrenado. 15](#_Toc506322967)

[Ilustración 4‑3: Mapa (colores) de clasificación del conjunto de entrenamiento. 15](#_Toc506322968)

[Ilustración 4‑4: Mapa (colores) de activaciones del conjunto de entrenamiento. 16](#_Toc506322969)

[Ilustración 4‑5: Mapa (colores) de activaciones (dibujadas las clases del mapa). 16](#_Toc506322970)

[Ilustración 4‑6: Mapa (colores) de activaciones (dibujados los colores proporcionados). 17](#_Toc506322971)

[Ilustración 4‑7: Mapa (colores) de activaciones (Pruebas). 17](#_Toc506322972)

[Ilustración 4‑8: Mapa (500 números) de clasificación del conjunto de entrenamiento. 18](#_Toc506322973)

[Ilustración 4‑9: Mapa (números) de activaciones del conjunto de entrenamiento. 19](#_Toc506322974)

[Ilustración 4‑10: Conjunto de pruebas (números). 19](#_Toc506322975)

[Ilustración 4‑11: Mapa (500 números) de activaciones (números sustituidos). 20](#_Toc506322976)

[Ilustración 4‑12: Mapa (500 números) de activaciones (solo los buscados). 20](#_Toc506322977)

[Ilustración 8‑1: Mapa (1797 números) de activaciones (números sustituidos). 24](#_Toc506322978)

[Ilustración 8‑2: Mapa (1797 números) de activaciones (solo los buscados). 25](#_Toc506322979)

Índice de tablas

[Tabla 2‑1: Valores óptimos del clasificador (colores). 9](#_Toc506285537)

[Tabla 3‑1: Valores óptimos del clasificador (números). 12](#_Toc506285538)

Listado de abreviaturas

|  |  |
| --- | --- |
| SOM | *Self-Organizing Map* (Mapa autoorganizado) |
|  |  |
|  |  |
|  |  |
|  |  |

Listado de definiciones

***Dataset***. Conjunto de datos.

# Objetivo del Laboratorio 1

Desarrollar de forma autónoma una herramienta de Generación de Mapas Autoasociativos que permita crear y entrenar redes SOM a partir de un dataset de entrada y unos parámetros introducidos en tiempo de ejecución. Usar la herramienta para analizar y resolver dos casos y responder al cuestionario que se plantea para cada uno de ellos.

Casos:

* Colores expresados en formato de tripletes de enteros con valores desde 0 a 255.
* Números escritos a mano representados mediante una matriz de 8x8 píxeles con valores desde 0 a 16.

# Resultados Práctica 1 (Colores)

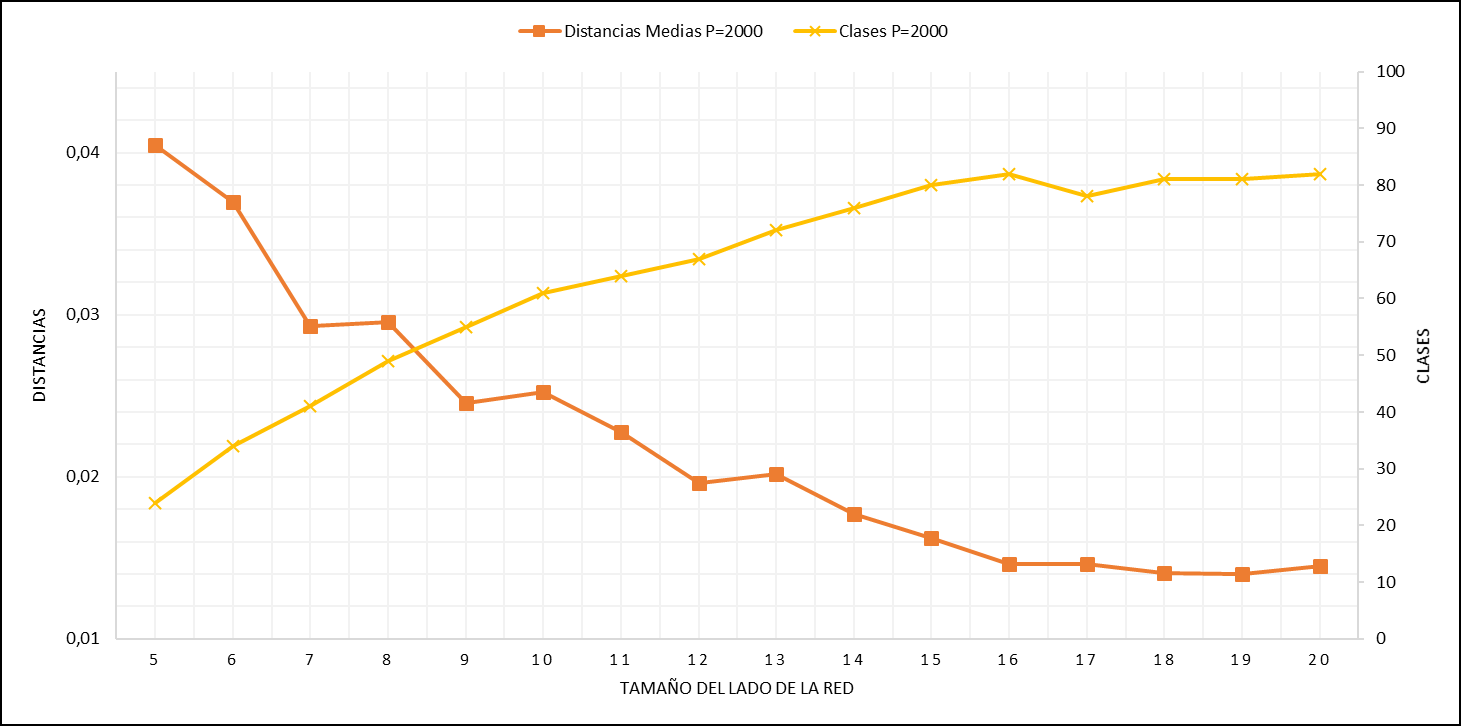


Ilustración 2‑1: Variación de distancia media con respecto al lado del mapa de Kohonen (colores).

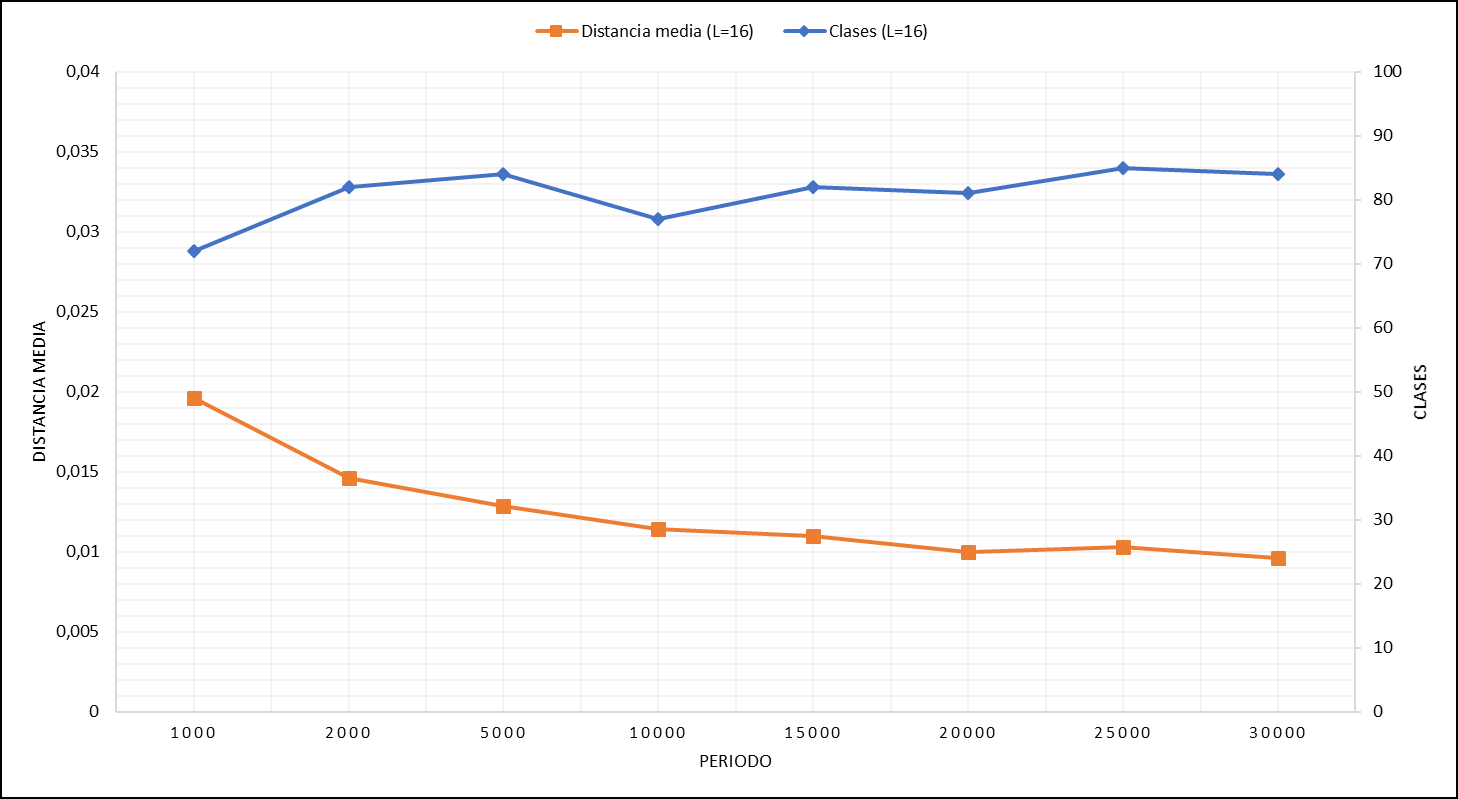


Ilustración 2‑2: Variación de la distancia media con respecto al periodo (colores).

|  |  |
| --- | --- |
| **Variable** | **Valor** |
| Matriz de pesos inicial | Aleatorio |
| Cálculo de distancias | Euclídea |
| Lado del mapa (L) | 16 |
| Mapa de Kohonen (K) | 16 x 16 |
| Periodo | 10.000 |
| Función de decrecimiento de η | Lineal |
| Coeficiente de aprendizaje (η) | 0.1 |
| Vecindario | L / 2 |

Tabla 2‑1: Valores óptimos del clasificador (colores).

# Resultados Práctica 2 (Números escritos a mano)

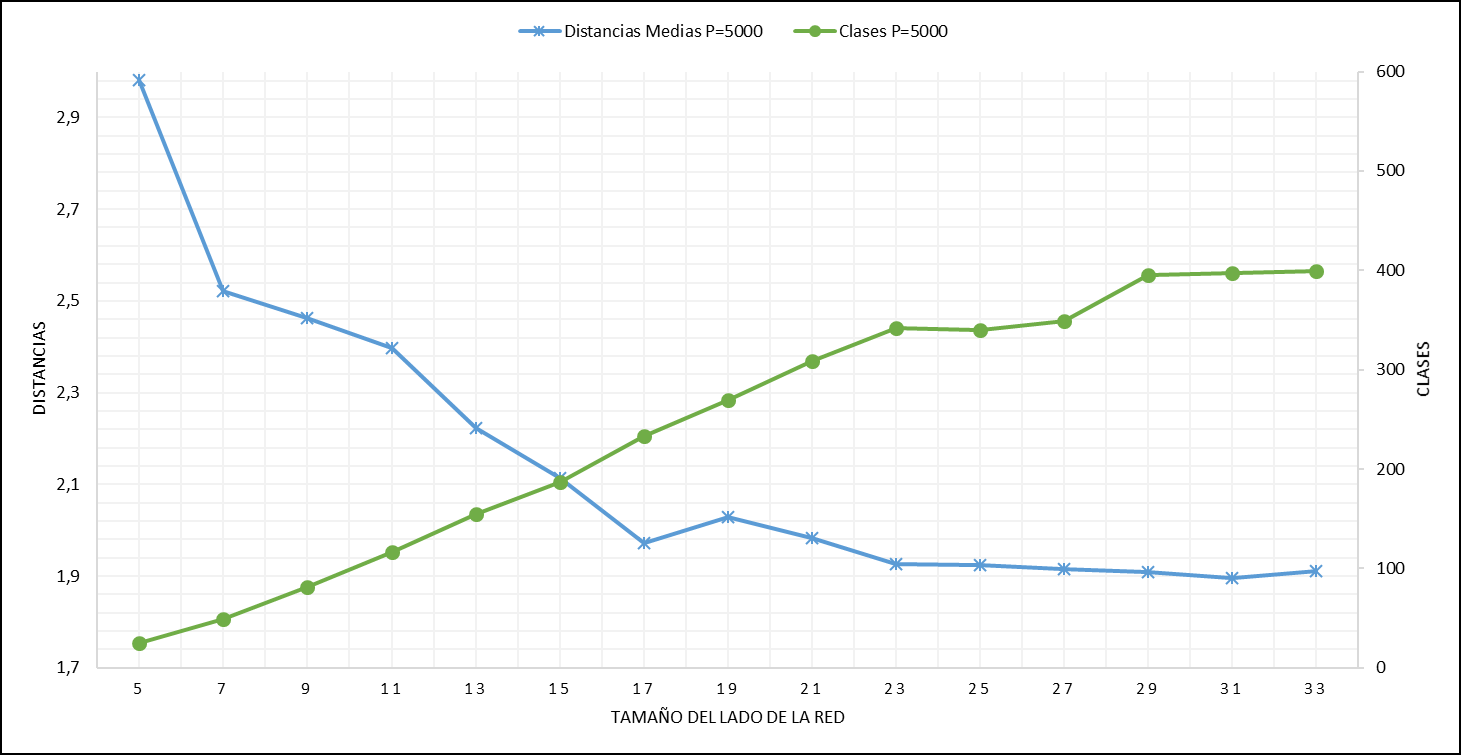


Ilustración 3‑1: Variación de distancia media con respecto al lado del mapa de Kohonen (números).

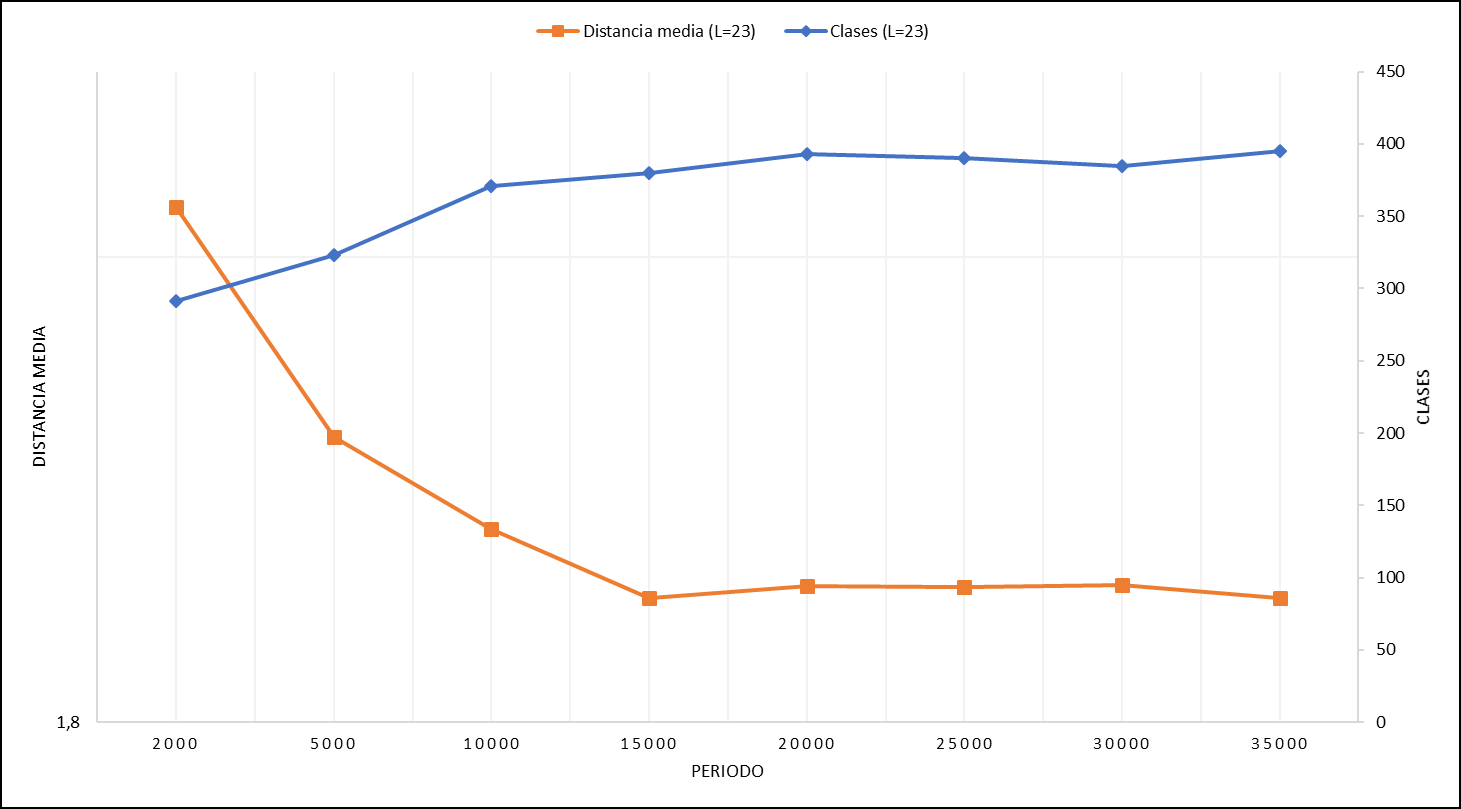


Ilustración 3‑2: Variación de la distancia media con respecto al periodo (números).

|  |  |
| --- | --- |
| **Variable** | **Valor** |
| Matriz de pesos inicial | Aleatorio |
| Cálculo de distancias | Euclídea |
| Lado del mapa (L) | 23 |
| Mapa de Kohonen (K) | 23 x 23 |
| Periodo | 15.000 |
| Función de decrecimiento de η | Lineal |
| Coeficiente de aprendizaje (η) | 0.1 |
| Vecindario | L / 2 |

Tabla 3‑1: Valores óptimos del clasificador (números).

# Discusión general

## Proceso de búsqueda del mapa óptimo

El proceso de búsqueda de la mejor configuración del clasificador con respecto a las variables: lado del mapa de Kohonen (L), periodo (P) y coeficiente de aprendizaje (η) se basa en la experimentación. Es un procedimiento por el cual se realizan diversas pruebas con distintos valores, para posteriormente proceder a un análisis de los resultados atendiendo a las variaciones de las distancias medias de las clases reconocidas.

Realizando una serie de gráficas, observaremos que llegado cierto lado del mapa (L), las distancias disminuyen ínfimamente, instante en el cual el mapa comienza a dividir en exceso las clases reconocidas, y por tanto es el punto donde el mapa comienza a bajar su rendimiento como clasificador.

### Procedimiento

El procedimiento establecido para este experimento es:

1. Buscar el lado del mapa (L) óptimo.

Para ello, al calcularse la matriz de pesos de forma aleatoria, se realizará una media entre 3 resultados con el mismo lado y posteriormente se analizarán los resultados referentes a la distancia media de cada mapa.

1. Buscar el periodo (P) óptimo.

Una vez se tenga el valor de L óptimo, se realizará otro banco de pruebas para analizar la distancia media de los mapas, con L constante y variando P.

Se ha establecido coeficiente de aprendizaje (η) constante en todas las pruebas con un valor de 0,1.

## Mapa Práctica 1 (Colores)

### Conjunto de entrenamiento

Para la búsqueda del mejor clasificador se ha empleado un conjunto de entrenamiento de 100 colores seleccionados de manera aleatoria, lo que no asegura que la distribución de los colores sea uniforme, es decir, puede ocurrir que haya desequilibrios entre la cantidad de instancias de un color de manera que haya muchos colores de un tipo en concreto y pocos o ninguno de otros.

### Lado del mapa (L) óptimo

Atendiendo a la ilustración 2-1 se ha establecido como valor de L óptimo 16 (como los mapas de 256 colores de los archivos .gif), debido a que, a partir de ese valor, la variación de las distancias medias es mínimo.

### Periodo (P) óptimo

Atendiendo a la ilustración 2-2 se ha establecido como valor de L óptimo 10.000, debido a que, a partir de ese valor, la variación de las distancias medias es despreciable.

### Resultados

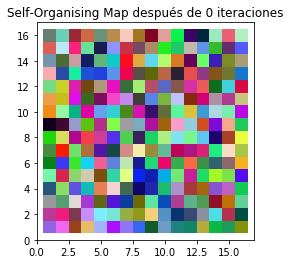


Ilustración 4‑1: Mapa (colores) inicial sin entrenar.

La ilustración 4-1 muestra la matriz de pesos generada de forma aleatoria previa al entrenamiento.

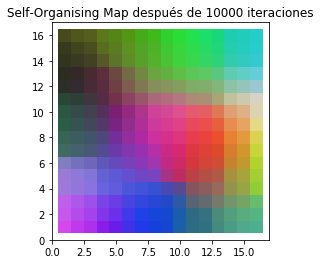


Ilustración 4‑2: Mapa (colores) entrenado.

La ilustración 4-2 muestra la matriz de pesos generada de forma aleatoria después del entrenamiento.

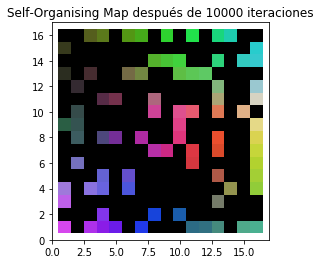


Ilustración 4‑3: Mapa (colores) de clasificación del conjunto de entrenamiento.

La ilustración 4-3 muestra el mapa de colores clasificados, usado como datos de entrada el propio conjunto de entrenamiento.

|  |
| --- |
| [[2 0 1 2 0 0 0 0 2 0 0 0 1 0 1 0]  [0 0 0 0 0 1 0 1 1 1 0 1 0 0 0 0]  [1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1]  [1 1 0 1 1 0 0 1 0 0 1 0 0 0 0 1]  [1 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0]  [0 0 0 1 2 0 0 0 0 0 0 0 1 0 0 1]  [1 0 0 0 0 0 0 2 0 0 0 0 2 0 0 1]  [0 3 0 0 0 0 1 0 0 1 1 0 0 1 0 0]  [0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 2]  [0 1 0 0 0 0 0 1 1 2 0 0 1 1 0 0]  [1 0 0 0 0 1 1 0 0 1 0 0 1 0 0 3]  [1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0]  [1 0 1 0 2 0 1 3 0 1 1 1 0 1 0 1]  [0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 2]  [1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0]  [3 0 0 1 1 1 1 1 1 0 1 1 0 1 2 0]] |

Ilustración 4‑4: Mapa (colores) de activaciones del conjunto de entrenamiento.

La ilustración 4-4 muestra la matriz de activaciones o el número de patrones reconocidos por cada neurona.

Distancia media del mapa: 0.0121223652813.

Los valores de la matriz de pesos entrenada y el mapa de distancias medias se pueden consultar en el fichero L1P1-SOM\_colores-resultado.ipynb.

### Prueba de clasificación

Conjunto de clasificación: [255, 255, 255] [255, 0, 0] [0, 255, 0] [0, 0, 255] [255, 255, 0] [0, 255, 255]

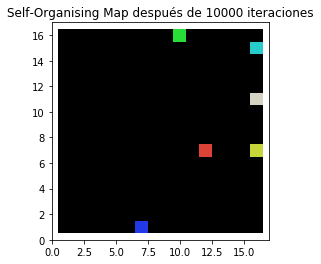


Ilustración 4‑5: Mapa (colores) de activaciones (dibujadas las clases del mapa).

La ilustración 4-5 muestra el mapa de los colores reconocidos. Se dibujan las clases de la matriz de pesos.

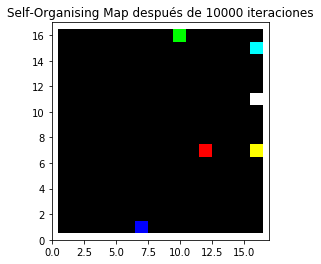


Ilustración 4‑6: Mapa (colores) de activaciones (dibujados los colores proporcionados).

La ilustración 4-6 muestra el mapa de los colores reconocidos. Se dibujan los colores del conjunto de pruebas.

|  |
| --- |
| [[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  [0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0]] |

Ilustración 4‑7: Mapa (colores) de activaciones (Pruebas).

La ilustración 4-7 muestra la matriz de activaciones o el número de patrones reconocidos por cada neurona.

Distancia media del mapa: 0.492107023265.

El mapa de distancias medias se puede consultar en el fichero L1P1-SOM\_colores-resultado.ipynb

## Mapa Práctica 2 (Números escritos a mano)

### Conjunto de entrenamiento

Para la búsqueda del mejor clasificador se ha empleado un conjunto de entrenamiento de 500 números escritos a mano de los 1797 números provistos en el paquete sklearn (debido al tiempo de cálculo que se requeriría con todos). En este paquete de datos, hay una distribución equilibrada de cantidad de ocurrencias de un mismo número, es decir, hay más o menos la misma cantidad de representantes de los distintos números.

### Lado del mapa (L) óptimo

Atendiendo a la ilustración 3-1 se ha establecido como valor de L óptimo 23, debido a que, a partir de ese valor, la variación de las distancias medias es despreciable.

### Periodo (P) óptimo

Atendiendo a la ilustración 3-2 se ha establecido como valor de L óptimo 15.000, debido a que, a partir de ese valor, la variación de las distancias medias es despreciable.

### Resultados

|  |
| --- |
|  |

Ilustración 4‑8: Mapa (500 números) de clasificación del conjunto de entrenamiento.

La ilustración 4-8 muestra el mapa de números clasificados, usado como datos de entrada el propio conjunto de entrenamiento.

|  |
| --- |
| [[17 8 3 1 8 15 24 22 7 3 0 2 38 15 10 11 14 5 9 10 3 2 17]  [ 2 4 0 0 9 6 2 2 0 6 3 4 10 9 2 4 9 4 1 0 0 1 6]  [ 9 3 1 0 6 16 9 2 1 6 0 0 0 1 1 11 3 0 1 0 2 5 12]  [11 1 3 2 1 6 1 6 2 7 0 0 1 0 3 0 0 0 1 1 7 0 5]  [ 9 1 7 2 0 0 1 1 2 1 0 0 5 8 2 8 3 0 2 9 1 0 0]  [ 3 0 11 1 0 2 2 0 1 5 1 5 5 0 3 3 6 3 4 1 0 0 0]  [ 0 1 0 4 4 11 7 11 1 0 2 27 4 1 2 3 11 3 2 1 6 9 17]  [ 0 0 1 0 2 1 10 11 16 1 1 6 13 6 5 3 2 1 3 4 2 2 1]  [11 1 0 0 1 0 3 2 8 3 1 2 5 6 15 0 0 0 1 6 4 0 7]  [ 0 1 1 7 1 0 3 3 2 11 1 0 1 7 3 0 0 0 15 2 0 6 20]  [ 1 1 0 4 4 3 5 1 1 5 2 1 0 0 0 0 3 4 0 4 1 2 33]  [ 2 2 0 1 0 0 5 8 1 0 2 9 2 1 0 0 4 5 1 5 12 8 5]  [15 3 2 1 0 1 3 6 2 0 0 0 2 4 2 2 4 0 0 3 12 7 9]  [ 0 4 1 3 2 1 1 3 0 3 1 0 1 3 8 4 2 0 4 3 3 0 13]  [ 3 1 2 6 1 1 2 0 1 3 0 1 0 0 6 4 4 2 7 1 0 1 7]  [ 2 0 3 2 0 2 0 0 0 0 4 0 3 3 1 3 1 1 6 1 1 1 7]  [ 1 1 0 3 0 2 10 1 0 0 0 3 3 2 2 21 0 1 1 5 1 5 2]  [ 2 0 3 5 2 0 0 0 0 2 1 2 3 6 2 9 0 0 2 20 1 0 0]  [ 2 5 0 6 0 0 2 2 0 5 5 3 24 9 3 3 0 1 1 1 0 0 7]  [ 5 1 4 2 3 2 2 1 0 11 0 1 1 0 1 1 0 2 0 0 1 2 16]  [ 1 9 5 4 1 0 0 0 0 0 1 0 1 0 0 6 0 8 2 2 1 5 11]  [ 0 1 1 2 1 0 1 0 1 0 1 1 0 0 3 0 3 0 1 0 7 7 0]  [ 0 0 3 2 7 0 3 3 3 0 5 6 9 2 9 5 1 0 7 3 10 12 10]] |

Ilustración 4‑9: Mapa (números) de activaciones del conjunto de entrenamiento.

La ilustración 4-9 muestra la matriz de activaciones o el número de patrones reconocidos por cada neurona.

Distancia media del mapa: 1.98612924251.

El mapa de distancias medias se puede consultar en el fichero L1P2-SOM\_numeros-resultado.ipynb

### Prueba de clasificación

Conjunto de clasificación: {3, Matriz en blanco, 0, 3, 5}



Ilustración 4‑10: Conjunto de pruebas (números).

Distancia media del mapa: 0.968525264882.

El mapa de distancias medias se puede consultar en el fichero L1P2-SOM\_numeros-resultado.ipynb.

|  |
| --- |
|  |

Ilustración 4‑11: Mapa (500 números) de activaciones (números sustituidos).

|  |
| --- |
|  |

Ilustración 4‑12: Mapa (500 números) de activaciones (solo los buscados).

# Reflexiones finales

* Es importante disponer de un conjunto de entrenamiento que represente correctamente el espacio muestral que posteriormente se va a clasificar. Si tenemos en el entrenamiento, muchos casos de una clase y pocos de otras, el clasificador puede bajar su rendimiento para algunos casos.
* Llegado cierto número de presentaciones (periodo) la red deja de aprender, por lo que no por realizar más presentaciones la red mejorará continuamente, se deberá explorar otras soluciones como la del refuerzo.
* Ciertos elementos de prueba no son correctamente clasificados. Como respuesta, se debería planificar un escenario de refuerzo para mejorar el rendimiento del mapa de Kohonen.
* Debido a que la matriz de pesos desde la que se parte en el entrenamiento se genera de manera aleatoria, para varias ejecuciones con los valores de P, L y η iguales, puede existir cierta variación en la distancia media de los mapas.

# Bibliografía

**Cáceres-Alonso, Paloma y García-Tejedor, Álvaro.** *Non-Supervised Neural Categorisation of NIR Applied to Pure Compounds NirJournal95.* Madrid, España : s.n.

**García-Tejedor, Álvaro.** Moodle UFV. Recursos de la asignatura. [En línea] http://moodleufv.ufv.es/moodle/.

# ANEXO A. Documentos adjuntados

Documentos entregados referentes al Laboratorio 1:

* /Práctica 1 – Colores
  + Evolucion.mp4
  + L1P1-SOM\_colores-resultado.ipynb
* /Práctica 2 - Números escritos a mano
  + L1P2-SOM\_numeros-resultado.ipynb
  + L1P2-SOM\_numeros-resultado-1797.ipynb
* LAB01-GRUPO02-MEMORIA.docx
* Datos.xlsx

# ANEXO B. Números escritos a mano, conjunto completo

## Conjunto de entrenamiento

En este caso se han empleado los 1797 números provistos en el paquete sklearn para construir el clasificador con el lado del mapa (L) igual a 43 y el periodo (P) igual a 30.000.

## Resultados

|  |
| --- |
|  |

Ilustración 8‑1: Mapa (1797 números) de activaciones (números sustituidos).

|  |
| --- |
|  |

Ilustración 8‑2: Mapa (1797 números) de activaciones (solo los buscados).

El mapa de distancias medias y el resto de información se puede consultar en el fichero L1P2-SOM\_numeros-resultado-1797.ipynb.