UNIVERSIDAD FRANCISCO DE VITORIA

ESCUELA POLITÉCNICA SUPERIOR



Práctica I

Inteligencia artificial II

Grupo B02

Realizada por:

Javier Garzón

Diego Guadalupe

Alejandro Monteliú

Carlos Pujols

# Control de Versiones

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ed.** | **Fecha** | **Cambio**  **(Incluya el capítulo / subcapítulo y una corta descripción)** | **Preparado**  **por** | **Nota de**  **Cambio** |
| 0.1 | 11/02/23 | Creación de memoria. | Carlos David Pujols | N/A |
| 0.2 | 12/02/23 | Implementación de practica 1. | Alejandro Monteliu | N/A |
| 0.3 | 19/02/23 | Implementación de practica 2. | Diego Guadalupe | N/A |
| 0.4 | 19/02/23 | Implementación bibliografía. | Javier Garzón |  |

\* N/A = No aplicable

ÍNDICE

[Control de Versiones](#_Toc127665735)

[1. Resultados de la práctica 1](#_Toc127665736)

[1.1 ¿Cuáles son los valores de Lado Mapa, Periodo y Eta más adecuados?](#_Toc127665737)

[1.2 Información generada en la ejecución y explicación detallada de cada una.](#_Toc127665738)

[1.2.1 Gráfico RGB de los pesos iniciales 6](#_Toc127665739)

[1.2.2 Gráfico RGB de los pesos entrenados 7](#_Toc127665740)

[1.2.3 Número de clases, mapa de clasificación, mapa de activaciones, mapa de distancias y errores. 7](#_Toc127665741)

[1.3 Clasificar el data set de prueba](#_Toc127665742)

[2. Resultados de la práctica 2](#_Toc127665743)

[2.1 ¿Cuáles son los valores de Lado\_Mapa, Periodo y Eta más adecuados?](#_Toc127665744)

[2.2 Número de clases, mapa de clasificación, mapa de activaciones, mapa de distancias y errores.](#_Toc127665745)

[2.3 Observa el mapa de clasificación y contesta](#_Toc127665746)

[2.3.1 ¿Qué grupos se pueden establecer? 26](#_Toc127665747)

[2.3.2 Analiza los vectores de pesos de las neuronas de Venezuela, Noruega y Etiopía 26](#_Toc127665748)

[2.3.3 Colorea el mapa de clasificación 29](#_Toc127665749)

[2.3.4 ¿Cómo se encuentran los países normalmente considerados subdesarrollados o en vías de desarrollo en tu mapa? 30](#_Toc127665750)

[2.4 Crea un mapamundi y coloréalo con el GDP per cápita](#_Toc127665751)

[3. Bibliografía](#_Toc127665752)

ÍNDICE DE FIGURAS

[Ilustración 1 Gráfica del codo del error 5](#_Toc127643449)

[Ilustración 2 Gráfico RGB de pesos iniciales 6](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643450)

[Ilustración 3 Gráfico RGB de pesos entrenados 7](#_Toc127643451)

[Ilustración 4 Matriz de clasificación 10](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643452)

[Ilustración 5 Mapa de activaciones 11](#_Toc127643453)

[Ilustración 6 Mapa de distancias 12](#_Toc127643454)

[Ilustración 7 Matriz de clasificación de prueba 16](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643455)

[Ilustración 8 Mapa de activaciones de prueba 17](#_Toc127643456)

[Ilustración 9 Mapa de distancias de prueba 17](#_Toc127643457)

[Ilustración 10 Mapa de prueba de los pesos 19](#_Toc127643458)

[Ilustración 11 Gráfica del codo práctica 2 20](#_Toc127643459)

[Ilustración 12 Mapa de clasificación práctica 2 21](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643460)

[Ilustración 13 Mapa en escala de grises 22](#_Toc127643461)

[Ilustración 14 Mapa de activaciones práctica 2 23](#_Toc127643462)

[Ilustración 15 Mapa de distancias de la práctica 2 24](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643463)

[Ilustración 16 Vector de pesos de los países 26](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643464)

[Ilustración 17 PCA de los 3 países 27](#_Toc127643465)

[Ilustración 18 PCA de todos los países 28](#_Toc127643466)

[Ilustración 19 Mapas de clasificación 29](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643467)

[Ilustración 20 Mapamundi por GDP 30](file:///C:\Users\Usuario\Documents\IA-II\Memoria_PRC1.docx#_Toc127643468)

# Resultados de la práctica 1

## 1.1 ¿Cuáles son los valores de Lado Mapa, Periodo y Eta más adecuados?

Para poder comprobar cuáles son los mejores valores para entrenar un SOM hemos realizado numerosas pruebas con la matriz de pesos y su set de entrenamiento de 100 colores para ver dónde disminuye el error.

De acuerdo con la siguiente gráfica creada en el Jupyter notebook de la práctica:

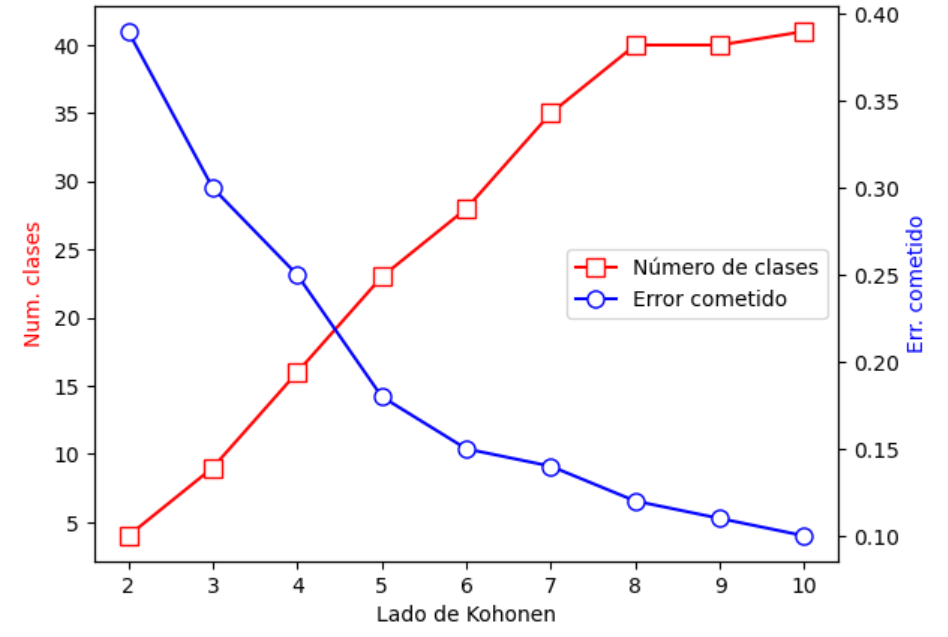


Ilustración 1 Gráfica del codo del error

Podemos comprobar que el error llega a un valor de 0.1 cuando el lado de mapa de Kohonen es igual a 10. Asimismo, vemos que el número de clases sube de manera lineal hasta el valor 8 de lado de mapa de Kohonen, dónde se estabiliza.

Además, tras las numerosas pruebas realizadas, podemos concluir que el periodo que menor error nos ha dado ha sido 10.000, con un Eta de 0.1.

Esto quiere decir que, a partir de ese número, la disminución del error era tan pequeña que no es rentable subir de ese 10.000 de periodo ya que los cambios no serán perceptibles, pero si requerirá una gran capacidad de cómputo.

## 1.2 Información generada en la ejecución y explicación detallada de cada una.

### 1.2.1 Gráfico RGB de los pesos iniciales

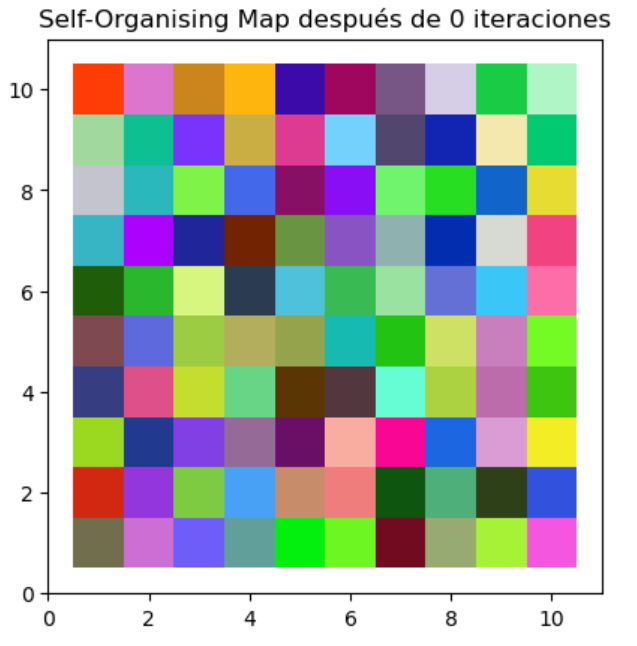


Ilustración 2 Gráfico RGB de pesos iniciales

En la foto anteriormente adjuntada, podemos observar el mapa inicial de la matriz de pesos inicial.

### 1.2.2 Gráfico RGB de los pesos entrenados

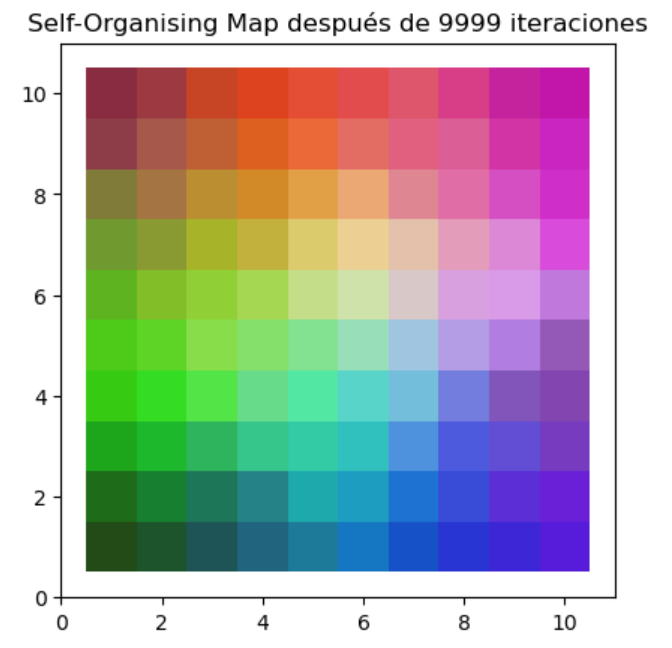


Ilustración 3 Gráfico RGB de pesos entrenados

En la foto adjuntada, podemos ver la matriz de pesos entrenada, por medio de pintar el mapa de pesos, ya que es una matriz con dimensión 3 y es posible pintarla con el método RGB.

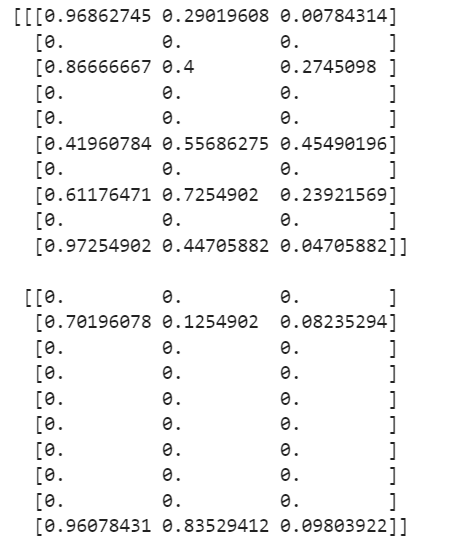
### 1.2.3 Número de clases, mapa de clasificación, mapa de activaciones, mapa de distancias y errores.

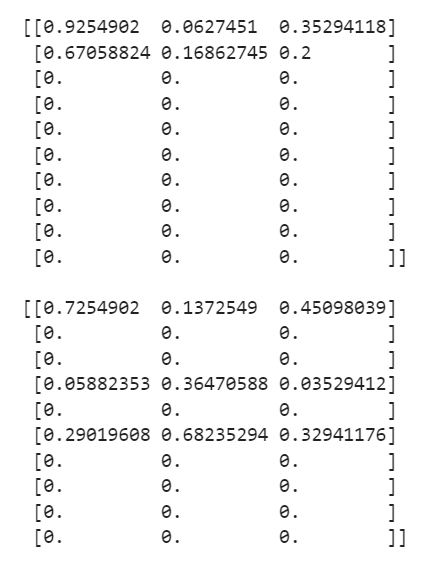
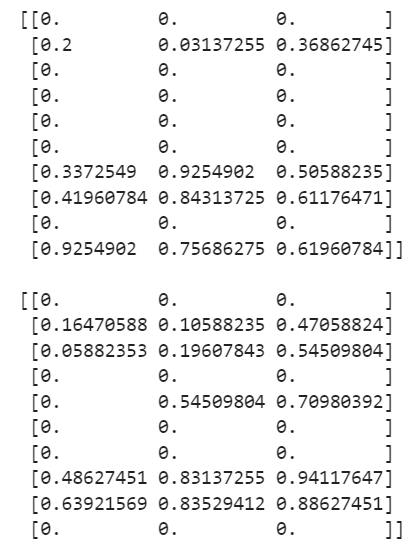
El número de clases, como hemos visto en la primera gráfica (gráfica del codo), son alrededor de 40.

Esto varía a la hora de entrenar la red, ya que haciendo la clasificación de otro dataset de colores pueden salirnos entre 38 y 44 clases.

La media de las pruebas que hemos realizado fueron 41, y por ello el resultado de la gráfica.

-Mapa de clasificación:





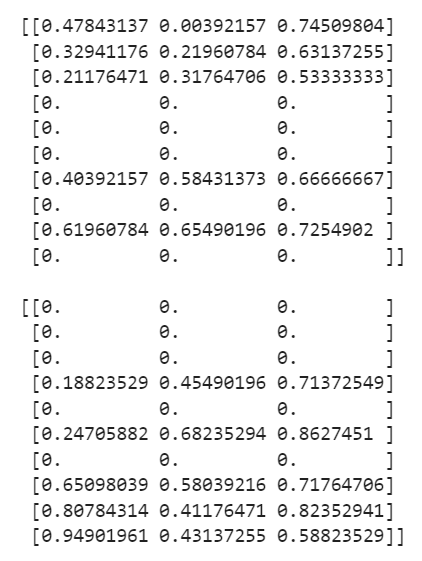
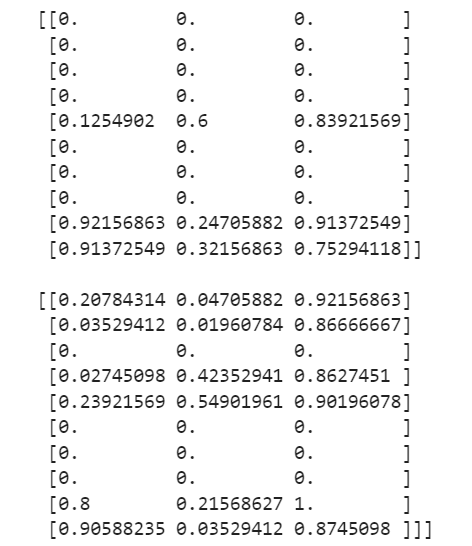


Ilustración 4 Matriz de clasificación

Cómo podemos ver en las distintas capturas, la matriz de clasificación es una matriz de dimensiones iguales a las de la matriz de pesos, en este caso 10x10x3.

El mapa de clasificación guarda en cada neurona del SOM el último patrón del set de clasificación. Es decir, en la matriz de clasificación se guarda el patrón en la misma posición de la BMU de la matriz de pesos al presentarle el patrón.

Gracias a esta matriz, podemos ver que no todas las neuronas han sido ganadoras en algún momento de la clasificación y también podemos observar el último patrón reconocido por cada neurona.

-Mapa de activaciones:

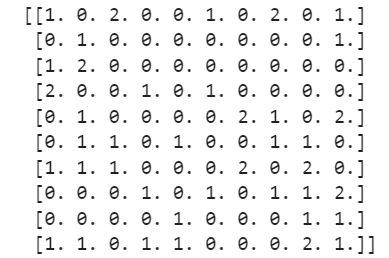


Ilustración 5 Mapa de activaciones

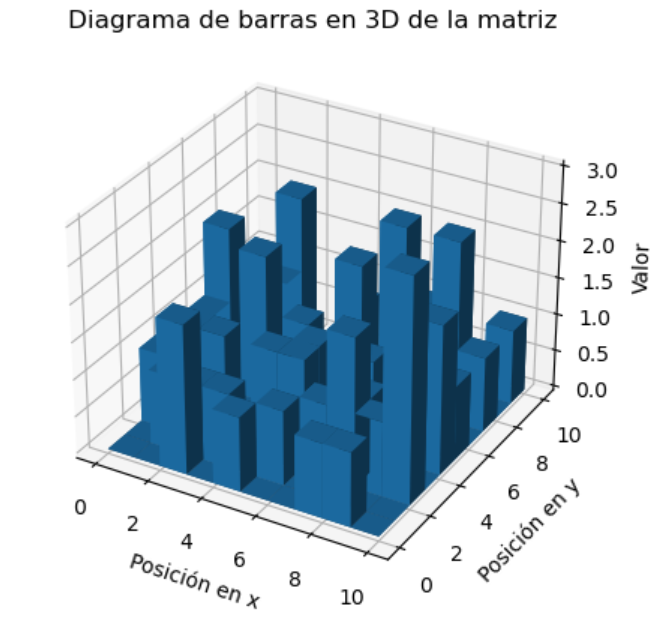


Ilustración Histograma 3D del mapa de activación de la clasificación

El mapa de activaciones nos muestra las veces que se ha activado cada neurona. Como ya sabíamos gracias al mapa de clasificación, no todas las neuronas se han activado a la hora de reconocer patrones en la clasificación.

También podemos observar que ha habido solo una o dos activaciones como mucho por neurona.

-Mapa de distancias:

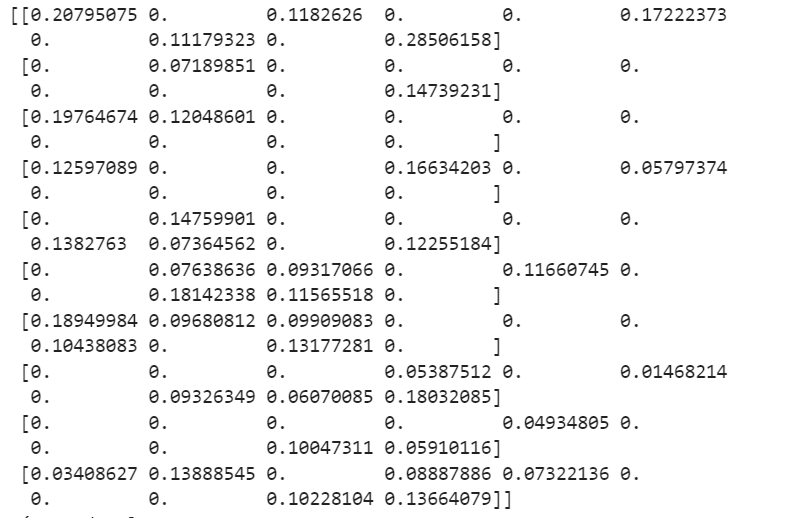


Ilustración 7 Mapa de distancias

El mapa de distancias nos permite ver la media de las distancias entre cada patrón de la clasificación y la neurona que lo ha reconocido. Es decir, se suman todas las distancias de cada neurona al patrón presentado y se guarda en la posición de la neurona, y por último se divide entre todas las veces que esa neurona ha sido ganadora o BMU.

Gracias a este mapa, podemos observar la distancia media de cada vector de la neurona a cada vector de los patrones presentados.

-Errores:

Para ver los errores de nuestro mapa, hay que acudir a la gráfica adjuntada como primera ilustración del documento.

Allí podemos ver lo siguiente:

El error conseguido con 10.000 de periodo y 10 de lado de Kohonen es 0.1

Esto quiere decir que el mapa está clasificando con un 90% de acierto.

Para calcular este error, debemos hacer el sumatorio de todas las distancias de cada patrón de datos a su BMU correspondiente al presentar el patrón.

Una vez hemos hecho eso, debemos dividir ese sumatorio entre el número de patrones distintos que le presentamos a la hora de clasificar, o lo que es lo mismo, el tamaño de nuestra muestra que le presentamos cuando clasifica.

El número correspondiente a esa división es el error del mapa. Para calcularlos todos, hemos hecho prueba y error.

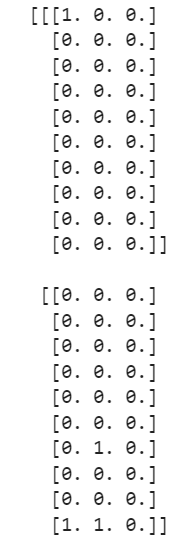
Es decir, hemos ido probando con distintos lados de mapa de Kohonen, así como con distintos Eta (learning rate) y distinto periodo.

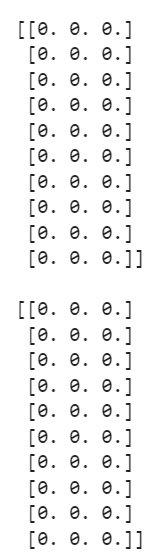
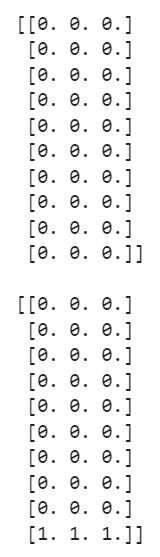
Hemos apuntado todos y cada uno de los resultados, para después poder plasmarlo en una gráfica y visualizar donde el error baja lo suficiente y el número de clases se estabiliza.

Cabe mencionar que, si hubiéramos aumentado el periodo y el lado del mapa, el error disminuiría; pero disminuiría tan poco que no es rentable en cuanto a computación se refiere. (Mayor tiempo de entrenamiento…)

## 1.3 Clasificar el data set de prueba

-Mapa de clasificación:





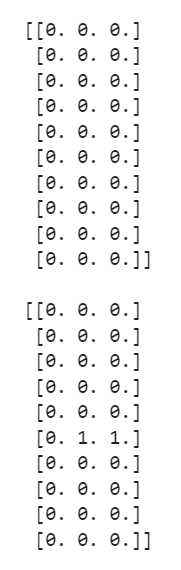
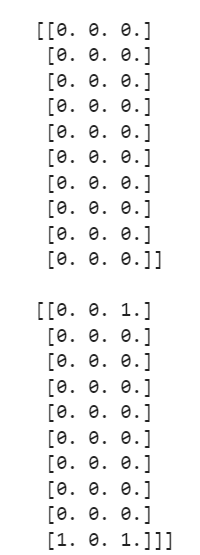


Ilustración 8 Matriz de clasificación de prueba



-Mapa de activaciones:

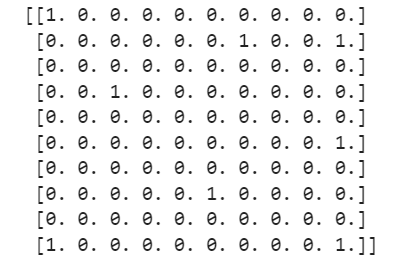


Ilustración 9 Mapa de activaciones de prueba

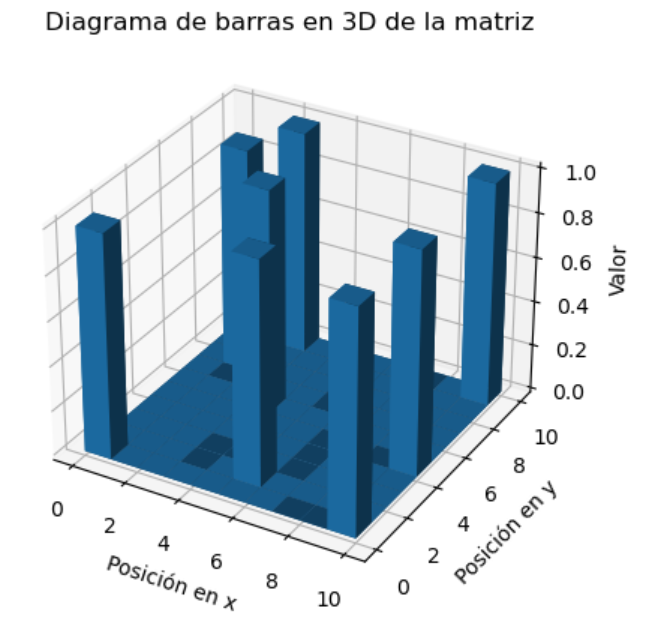


Ilustración Histograma 3D de la matriz de activación de la prueba

-Mapa de distancias:

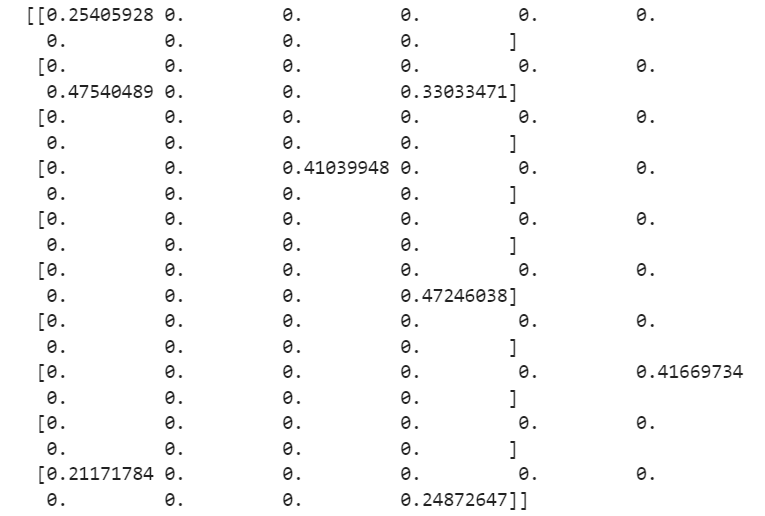


Ilustración 11 Mapa de distancias de prueba

Como hemos visto en el anterior apartado, el mapa de clasificación nos da el último patrón reconocido por esa neurona.

Como solo tenemos 8 patrones de entrada, cada uno corresponde a una neurona diferente, y por ello los 8 patrones de entrada están en neuronas diferentes.

En cuando al mapa de activaciones, nos indica que neuronas se han activado al presentar los patrones de prueba. Podemos ver que son solo 8 que corresponden a las posiciones del mapa de clasificación.

Por último, el mapa de distancias nos da la media de todas las distancias euclídeas entre la BMU y el patrón que ha reconocido.

En el caso de este mapa, simplemente nos da la distancia que había entre el vector de pesos de la neurona y el vector del patrón, puesto que cada patrón fue reconocido por una neurona distinta.

-Errores y número de clases:

Estos dos parámetros se pueden explicar juntos.

Tanto el número de clases como el error tiene un valor elevado para este mapa de Kohonen.

La explicación es simple, al utilizar patrones de entrada con valores límite, las clases que nos aparecen son 8, ya que hay una neurona que se activa por patrón.

El error en este caso es de 0.3, ya que a la hora de clasificar los patrones de entrada son pocos y son casos límite.

Asimismo, adjuntamos un mapa en el que se pueden visualizar donde ha clasificado el mapa los colores del enunciado:

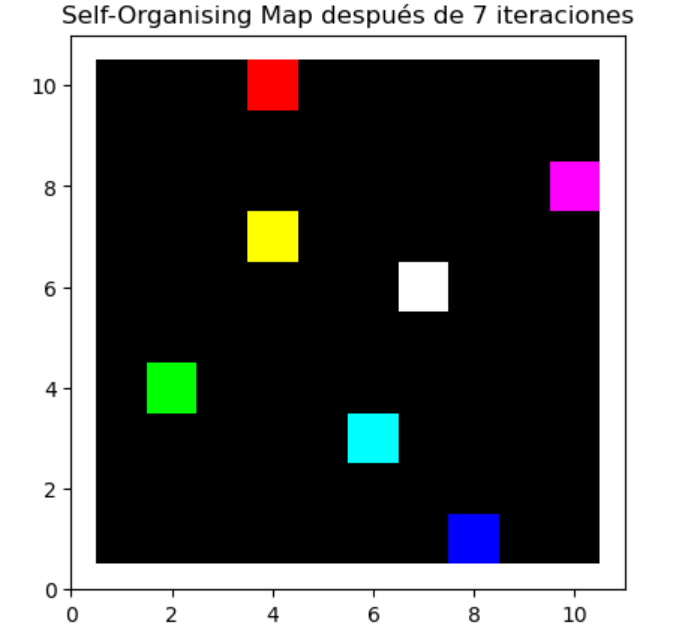


Ilustración 12 Mapa de prueba de los pesos

Como podemos ver, los colores clasificados coinciden con el mapa de pesos ordenado expuesto anteriormente. Los colores están en las posiciones en los que estaban los grupos de colores cuando ha hecho el clustering la red de Kohonen.

Esta es una manera visual de ver que la red de Kohonen ha clasificado los colores de prueba de manera correcta.

# Resultados de la práctica 2

## 2.1 ¿Cuáles son los valores de Lado\_Mapa, Periodo y Eta más adecuados?

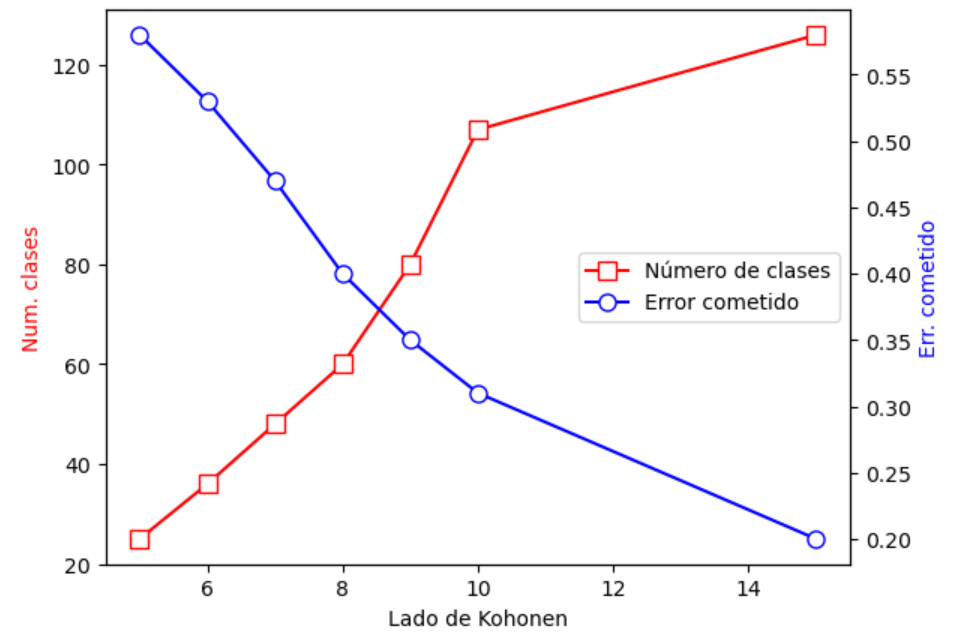


Ilustración 13 Gráfica del codo práctica 2

A partir de la gráfica podemos ver que el error conseguido es de 0.2 con un lado de mapa de Kohonen de 15, 30.000 de periodo y 0.1 de Eta.

Hemos hecho prueba y error, guardando los valores por cada lado de mapa, periodo y Eta para poder ver cuándo baja el error y se estabilizan tanto el error como el número de clases.

Al igual que la red de Kohonen pasada, si aumentamos el lado de Kohonen y el periodo, el error disminuirá, pero el cambio será tan pequeño que será insignificante.

## 2.2 Número de clases, mapa de clasificación, mapa de activaciones, mapa de distancias y errores.

-Mapa de clasificación:



Ilustración 14 Mapa de clasificación práctica 2

Esta matriz de clasificación está modificada respecto a la de la práctica anterior.

Hemos modificado la matriz y el código para poder añadir en cada posición de cada neurona el país que ha reconocido, esto nos sirve para hacernos una idea de los clusters que ha creado la red neuronal.

Si nos fijamos, tanto en la primera fila como en la última son los sitios donde más concentrados se encuentran los países.

Esto quiere decir que todos los países se pueden agrupar en esos dos grupos. Hay varios países por el medio de la matriz, pero unos están bastante separados de los demás.

Si vemos este mapa:

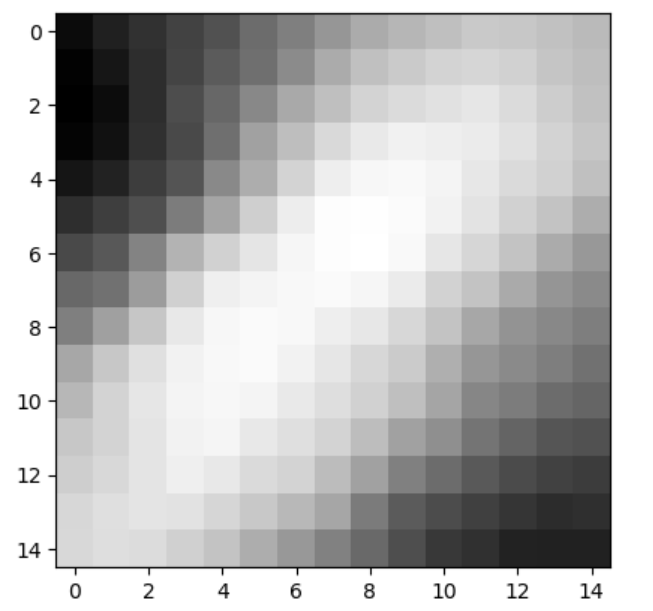


Ilustración 15 Mapa en escala de grises

Podemos observar, gracias a una gráfica hecha a partir de la matriz de pesos entrenada con una escala de grises.

En ella vemos la agrupación de los distintos países en clusters.

Podemos observar que hay dos grupos bien definidos y muchos países a su alrededor. En cambio, por el medio del mapa apenas hay países, igual que en la matriz de clasificación adjuntada anteriormente.

-Mapa de activaciones:

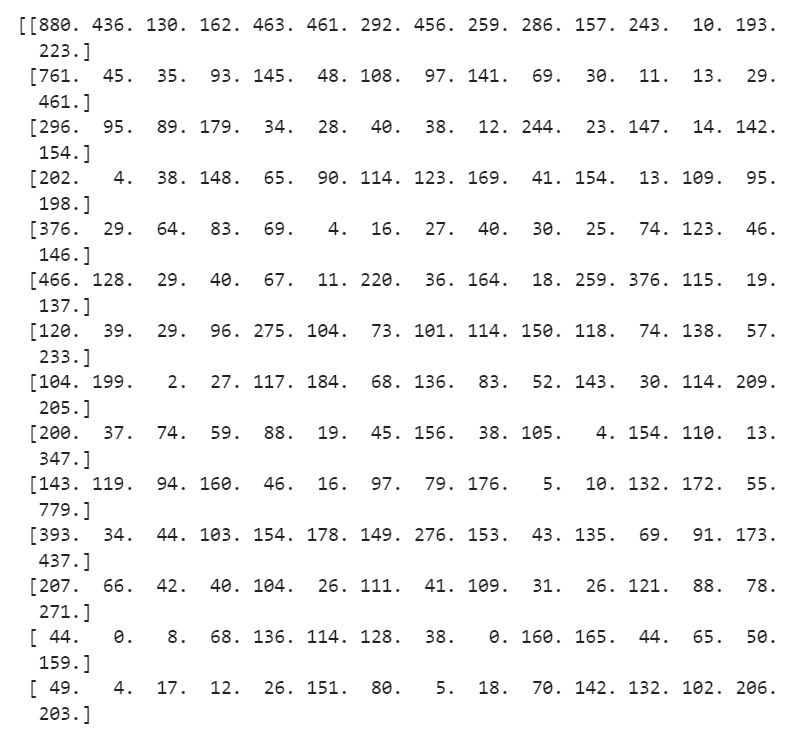


Ilustración 16 Mapa de activaciones práctica 2

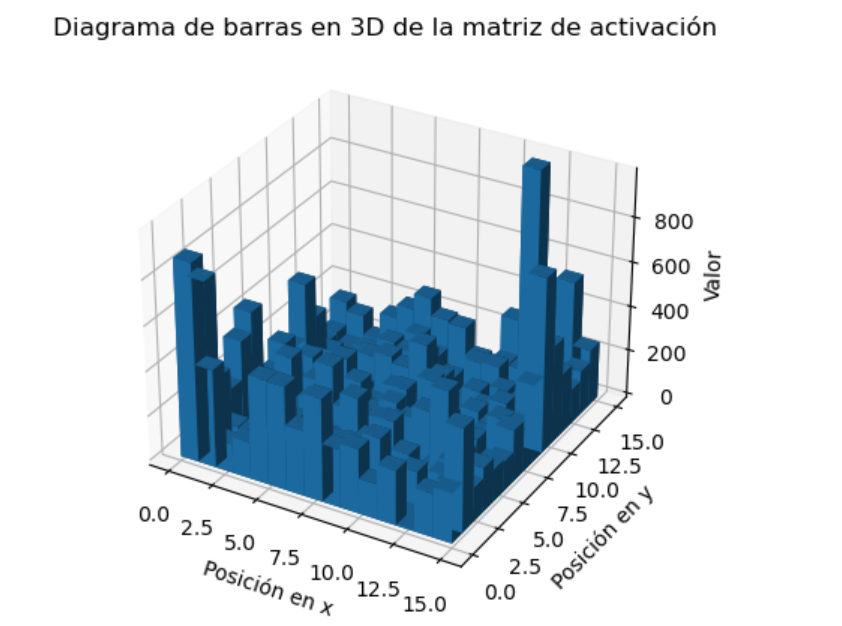


Ilustración Histograma 3D del mapa de activación

Como hemos visto en la práctica anterior, el mapa de activaciones nos muestra el número de veces que se ha activado cada neurona.

-Mapa de distancias:

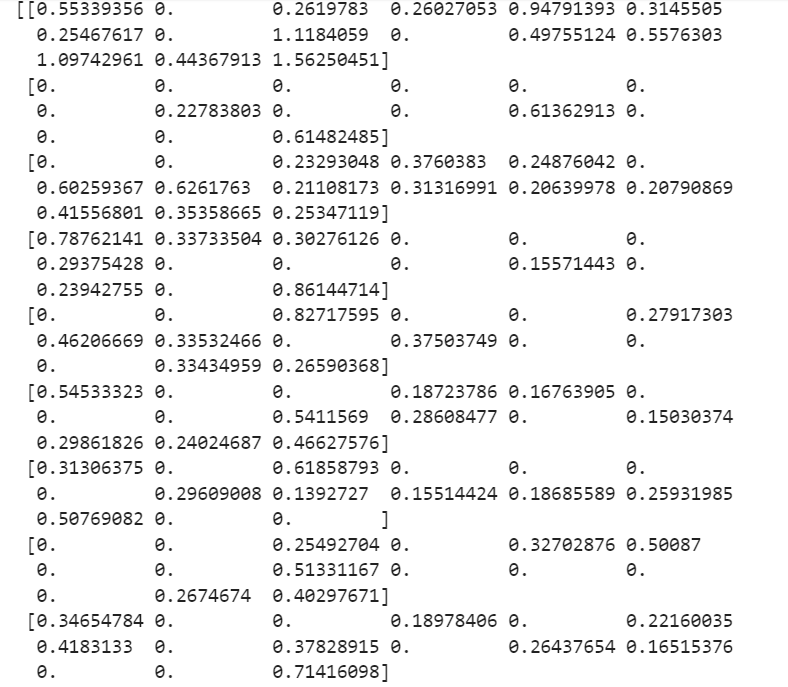
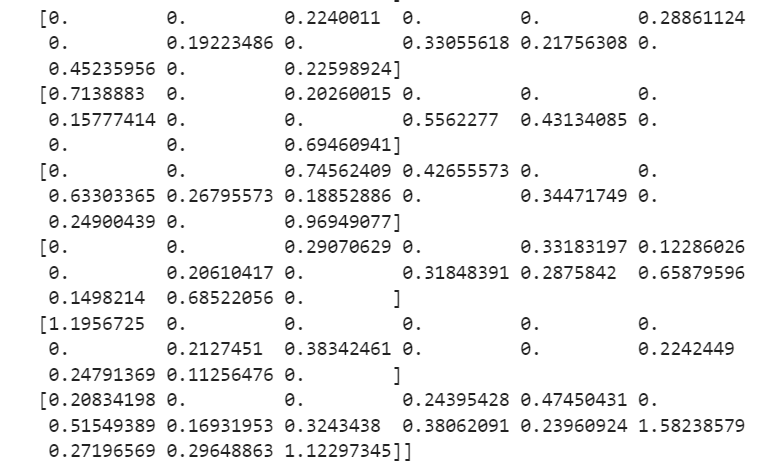


Ilustración 18 Mapa de distancias de la práctica 2

En el mapa adjuntado anteriormente, podemos ver la matriz del mapa de distancias de 15x15 (al igual que las 2 anteriores puesto que es igual que el lado del mapa de Kohonen).

Gracias a este mapa de distancias, podemos ver la distancia media de cada BMU al patrón que ha reconocido.

-Errores:

Los errores de este mapa los hemos comentado anteriormente con la gráfica del codo.

El error más bajo que hemos obtenido con unos hiperparámetros lógicos ha sido 0.2

Esto quiere decir que después de entrenar, el mapa de Kohonen tiene un porcentaje de acierto del 80%.

Sería posible obtener un error más bajo si aumentáramos el periodo y el lado de mapa de Kohonen. El problema de esto es la capacidad computacional de la máquina que entrena la red neuronal, ya que no es rentable en estos términos aumentar tantos los hiperparámetros para obtener un error un poco más bajo.

## 2.3 Observa el mapa de clasificación y contesta

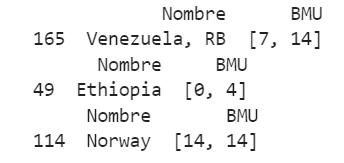
### 2.3.1 ¿Qué grupos se pueden establecer?

Viendo el mapa de clasificación y el mapa en escala de grises, podemos afirmar que se crean dos grupos a la hora de entrenar la red neuronal y hacer clustering, que podrían dividirse en subdesarrollados o en vías de desarrollo y los desarrollados; aunque hay algunos países lejos del núcleo de los clusters por ser países aún en desarrollo.

Estos dos grupos se agrupan en una zona de la matriz de pesos que es fácilmente reconocible al imprimir la matriz de clasificación y el mapa en escala de grises.

### 2.3.2 Analiza los vectores de pesos de las neuronas de Venezuela, Noruega y Etiopía

Aquí podemos ver la BMU que ha reconocido a los 3 países:



Y aquí el vector de pesos de la matriz de cada BMU.

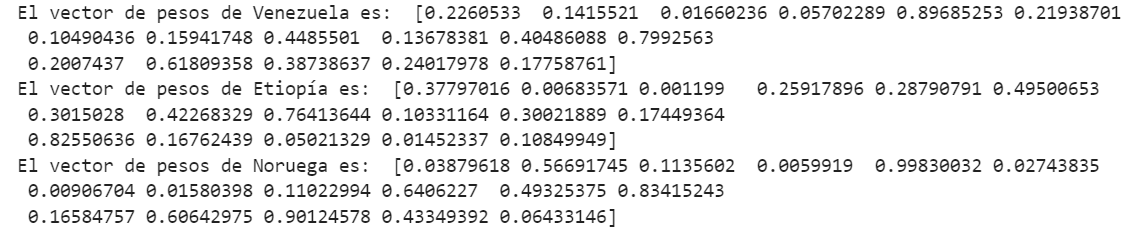


Ilustración 19 Vector de pesos de los países

Gracias a la información anterior, podemos saber que los 3 países están muy separados los unos de los otros.

Si nos fijamos en las BMU, vemos que los 3 países están muy alejados en una posible representación en 2D. Esto se debe a los distintos valores de las 17 características de entrada.

Si vemos los vectores de pesos de las neuronas, podemos ver que Etiopía tiene unos valores casi contrarios a los otros dos. Es decir, los valores más altos de Etiopía son los más bajos en los otros dos vectores.

Por otro lado, Venezuela y Noruega también tienen diferencias, a pesar de parecer que están cerca, la realidad es que están muy separados ambos países, es más, Etiopía está más cerca de Venezuela que de Noruega, y Venezuela está más cerca de Etiopía que de Noruega.

Por ello, Noruega está en un clúster distinto a Venezuela y Etiopía.

Aquí podemos ver un diagrama de dispersión en 3D de los 3 países, creado por el método de PCA aprendido en la asignatura de DataMinig, y que hemos incorporado para una mejor interpretación:

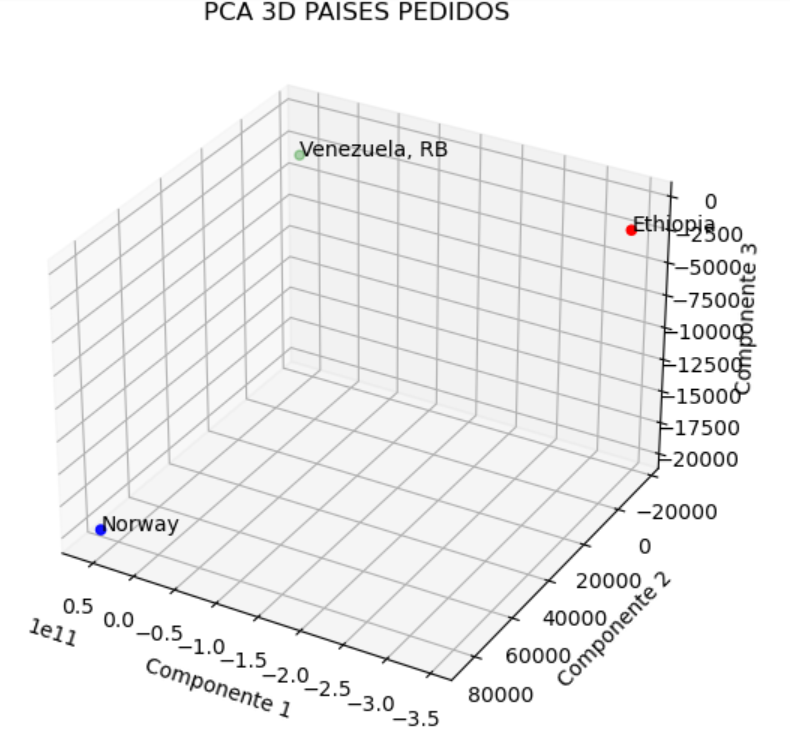


Ilustración 20 PCA de los 3 países

Gracias a esta representación, podemos ver que los 3 países están muy lejos de los demás, y que los 2 más cercanos son Venezuela y Etiopía.

Para ilustrar mejor el PCA de todos los países, hemos creado este gráfico:

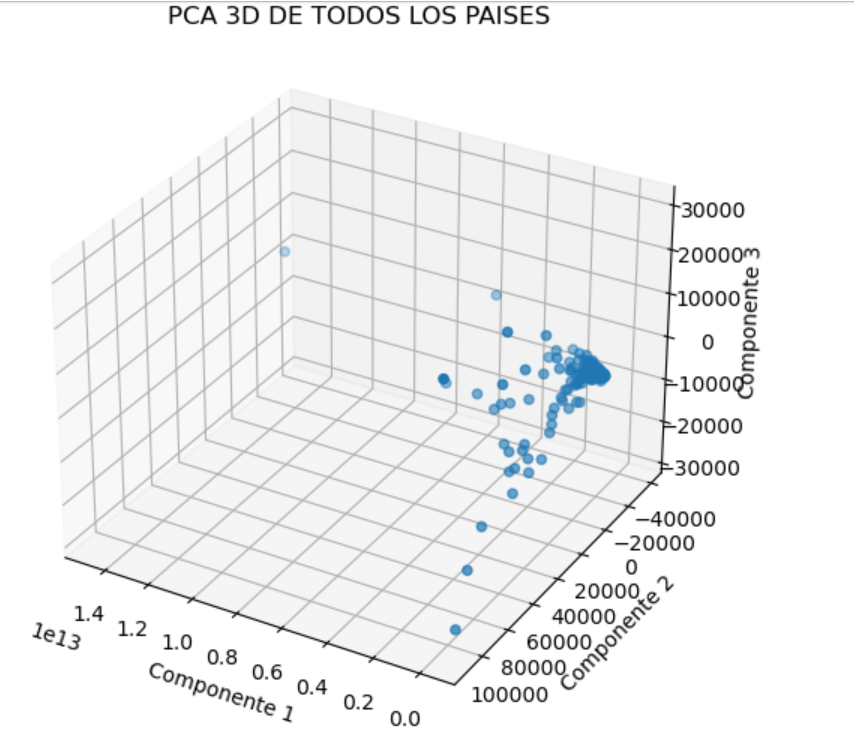


Ilustración 21 PCA de todos los países

Aquí podemos ver de manera más gráfica, la agrupación que hace el método de PCA con una exactitud del 73%. (Tabla con frecuencias autovalores realizada en el código).

Agrupa los países en una zona de la gráfica, y algunos de los mismo salen del grupo puesto que tienen alguna característica con un valor extremo.

### 2.3.3 Colorea el mapa de clasificación

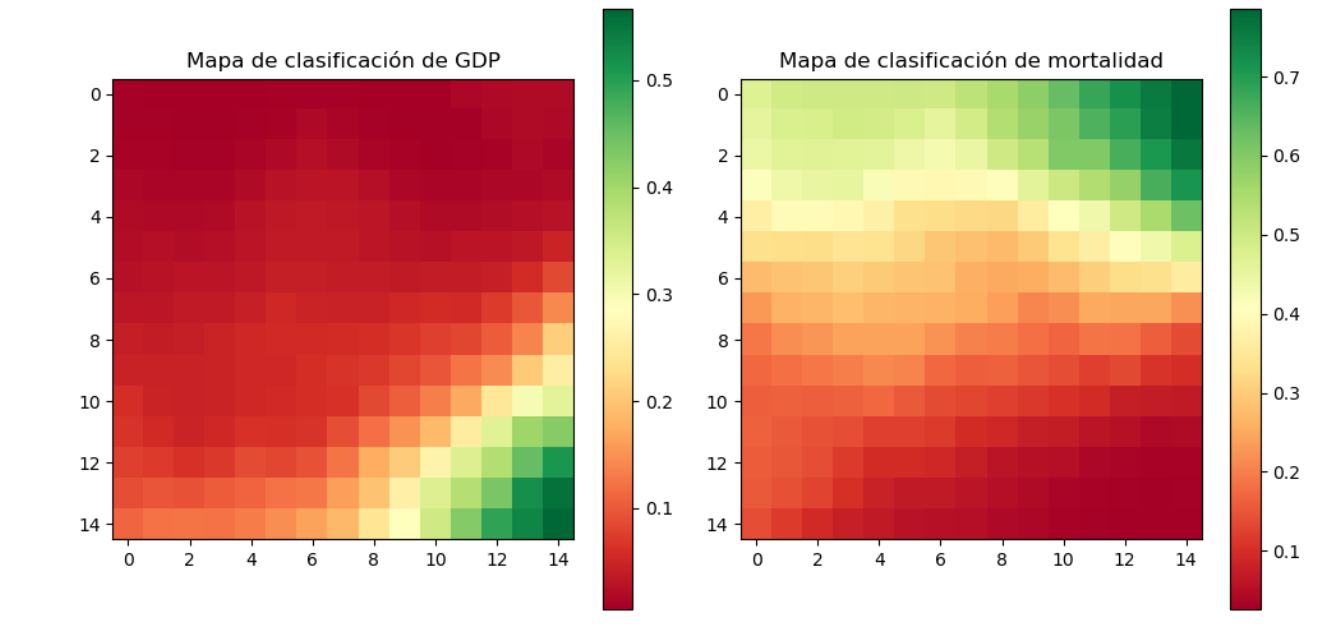
Aquí tenemos el mapa de clasificación de GDP per cápita y de mortalidad neonatal:

Ilustración 22 Mapas de clasificación

Podemos observar en los mismos una zona verde, referente a donde el valor es más alto.

Gracias a esto, podemos ver una cosa clara:

Donde mayor GDP per cápita hay (es decir, donde más dinero hay), hay una menor mortalidad y viceversa.

Atendiendo a este mapa, podemos ver una zona muy roja en el mapa de GDP per cápita en la zona de arriba del mapa. Esta misma zona es verde en el mapa de mortalidad.

Asimismo, la zona roja del mapa de mortalidad (abajo a la derecha), es la zona ver del mapa de GDP.

### 2.3.4 ¿Cómo se encuentran los países normalmente considerados subdesarrollados o en vías de desarrollo en tu mapa?

Los países menos desarrollados o en vías de desarrollo del mapa los vamos a encontrar en:

-En el mapa de clasificación de GDP per cápita los encontraremos en la zona roja, puesto que es la zona que tiene menos valor, en este caso el capital.

-En el mapa de clasificación de mortalidad los encontraremos en las zonas más claras, amarillas y verdes, ya que en estos países los índices de mortalidad neonatal suelen ser elevados.

## 2.4 Crea un mapamundi y coloréalo con el GDP per cápita

El mapamundi creado es el siguiente:

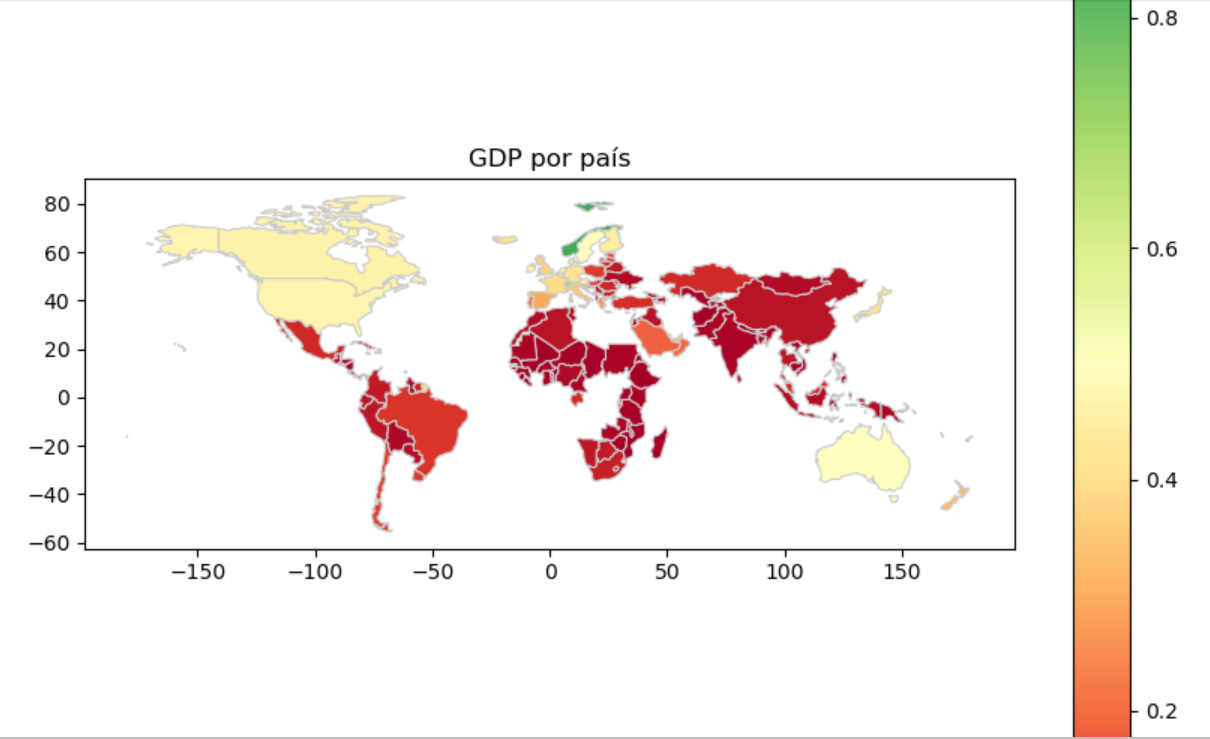


Ilustración 23 Mapamundi por GDP

Para la creación de este mapamundi hemos usado las librerías de geopandas y mapclassify.

Esto está convenientemente señalado en el código.

Gracias a este mapamundi, podemos ver por donde se distribuyen los valores mayores y menores respecto al GDP per cápita.

Los países con menor GDP están en rojo y los mayores en verde.

Viendo el mapa, África es el continente con mayor rojo, seguido muy de cerca de América del sur y Asia.

En cambio EE. UU, Canadá y gran parte de Europa occidental, tienen un color claro haciendo referencia a un GDP per cápita bueno.

Por último, destacar a Noruega, que es el país más verde, y por ende, el que mayor GDP per cápita posee.

# Bibliografía

-Colaboradores de Wikipedia. (2022, 15 julio). *Mapa autoorganizado*. Wikipedia, la enciclopedia libre. <https://es.wikipedia.org/wiki/Mapa_autoorganizado>

*-SOM Toolbox: Intro to SOM by Teuvo Kohonen*. (s. f.). <http://www.cis.hut.fi/projects/somtoolbox/theory/somalgorithm.shtml>

-Temario aula virtual asignatura de inteligencia artificial 2.

-Temario aula virtual asignatura Data Mining.