




SIA 2024 - 2C

Aprendizaje no supervisado

Grupo 4





SIA 2024 - 2C

EJERCICIO 1



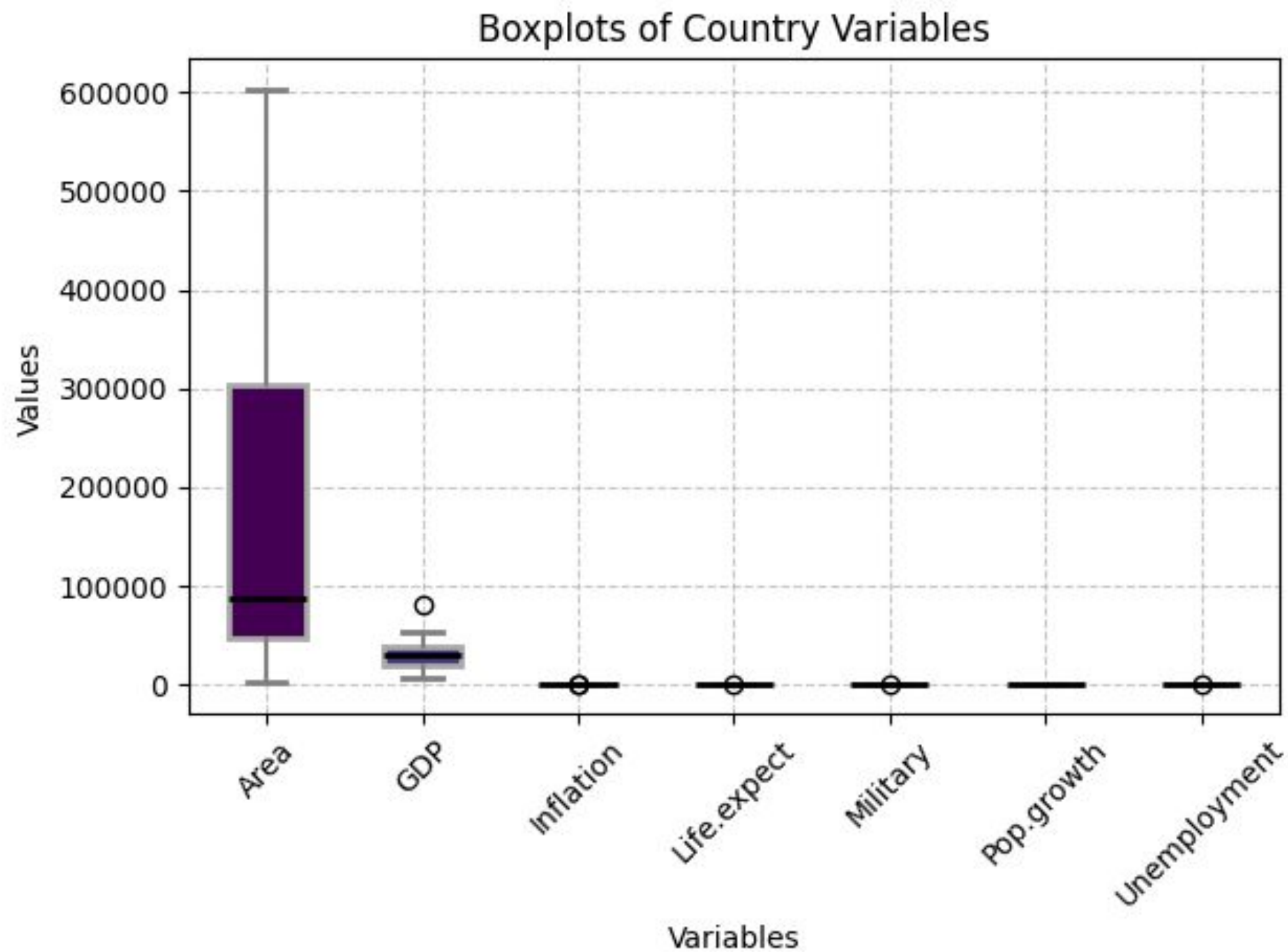
Datos

- Datos de diversos países europeos:
 - Country
 - Area
 - GDP
 - Inflation
 - Life expect
 - Military
 - Population growth
 - Unemployment

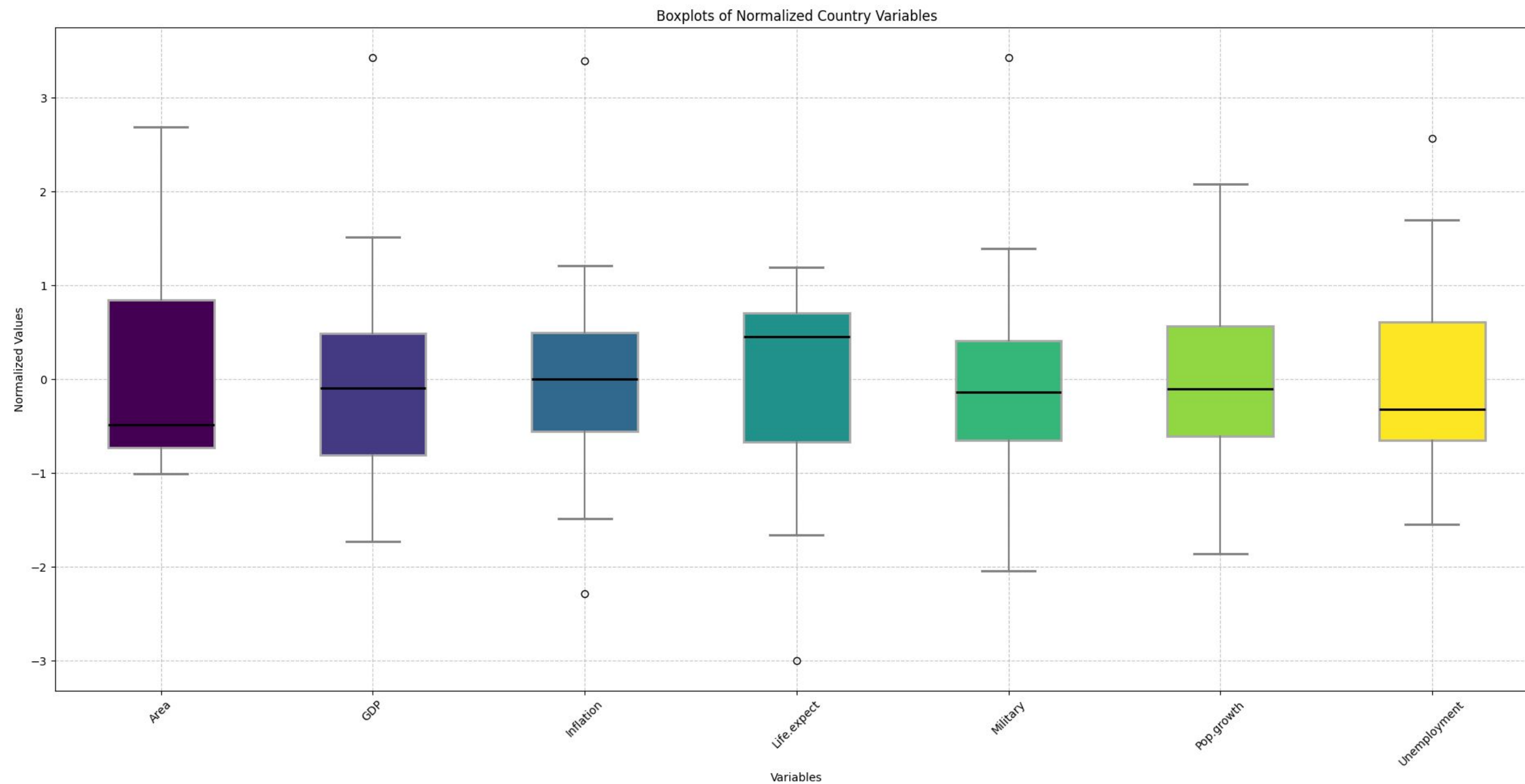
Estos datos son de distintos órdenes de magnitud. **ESTANDARIZAR**. Nosotros vamos a utilizar Z-Score.

$$Z = \frac{x - \mu}{\sigma}$$

Datos - Estandarización



Datos - Estandarización





SIA 2024 - 2C

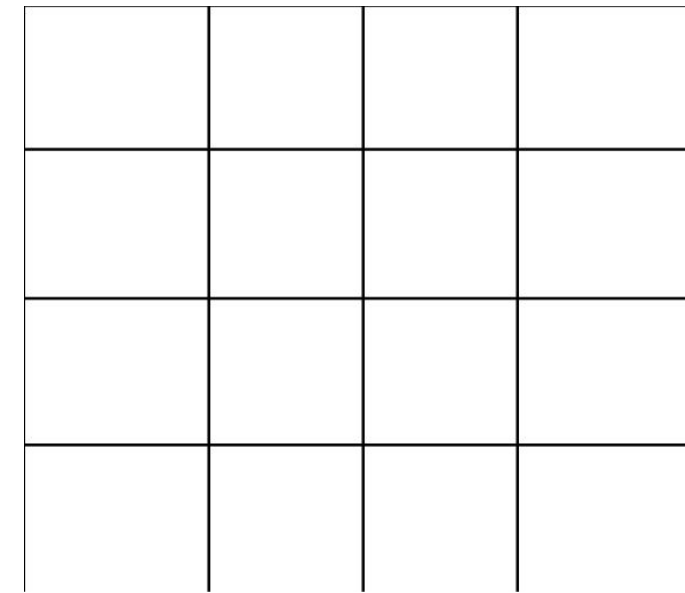
Red de Kohonen



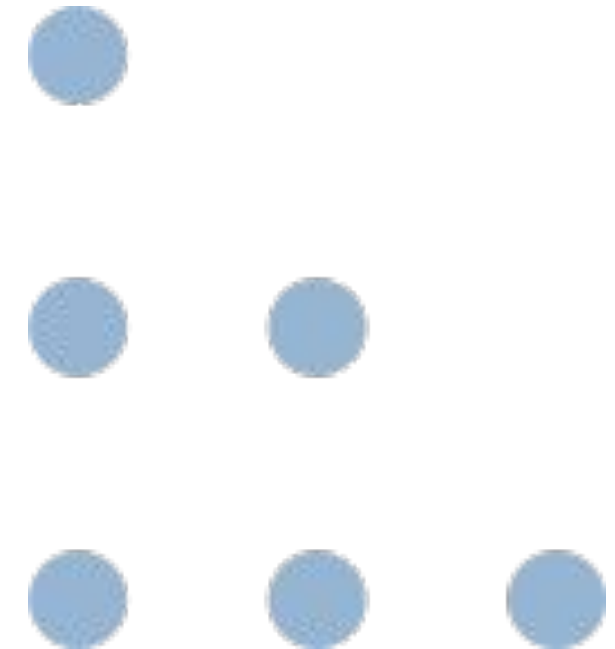
Ejercicio 1

Parámetros iniciales

- Learning rate = 0.1
- Sigma = 3
- K = 4



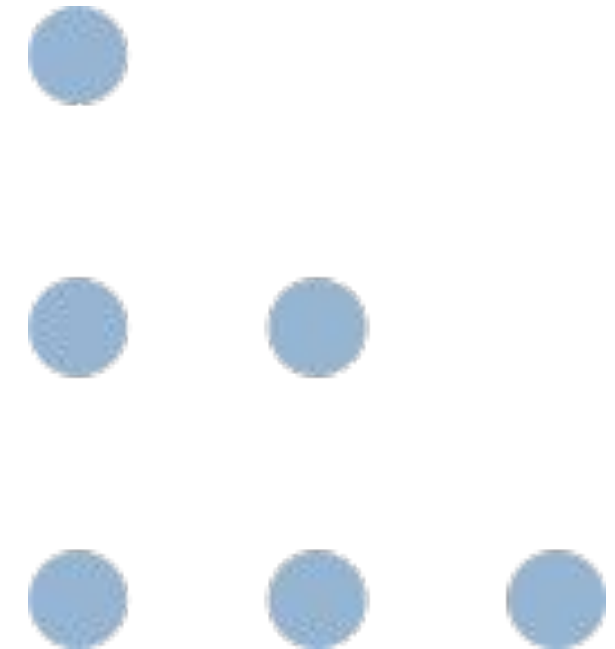
Grilla rectangular



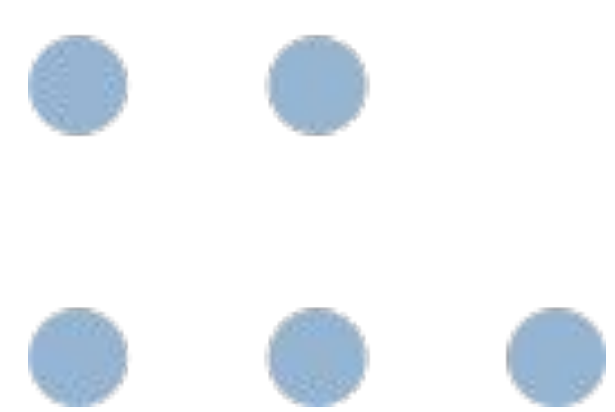
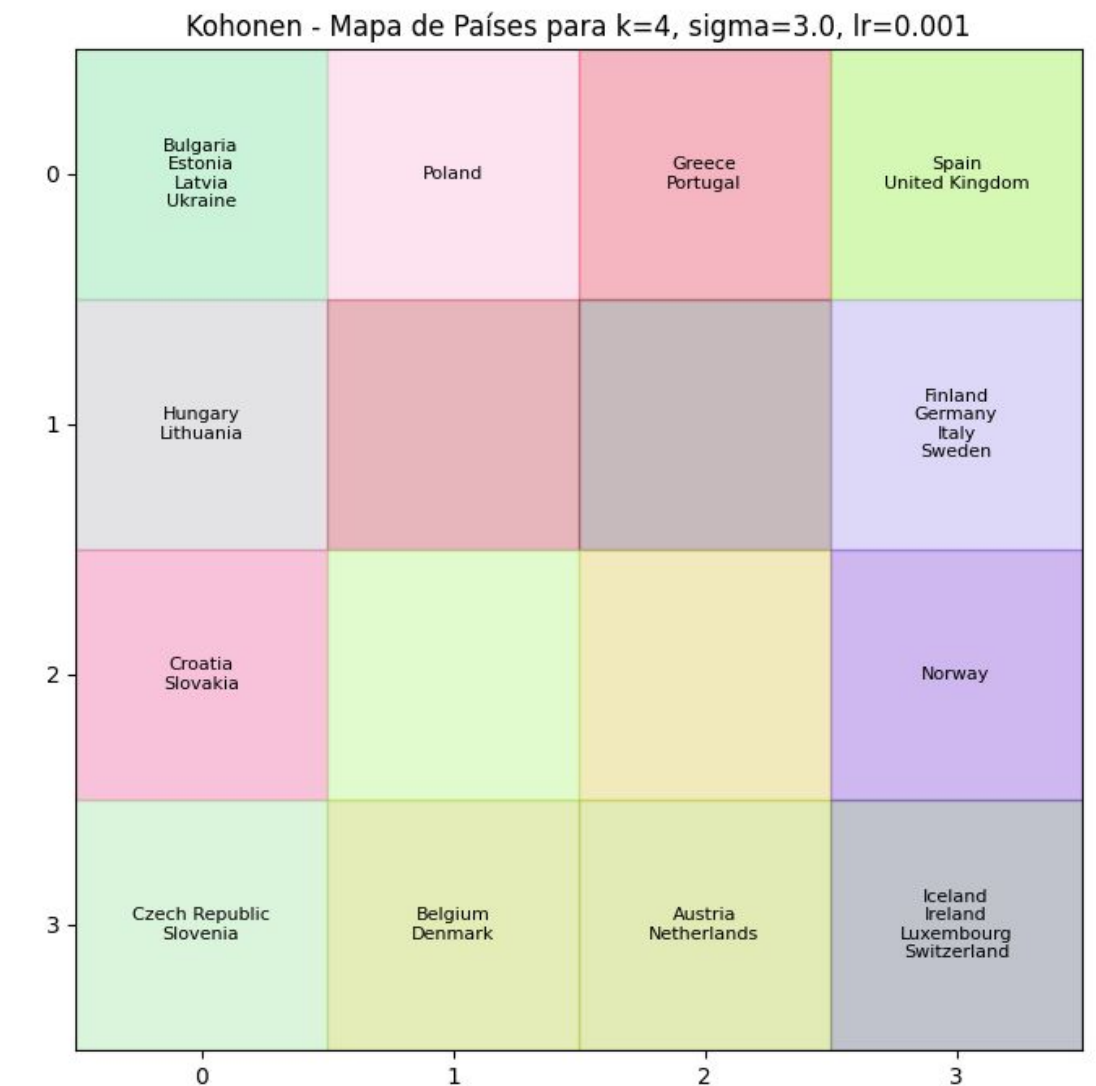
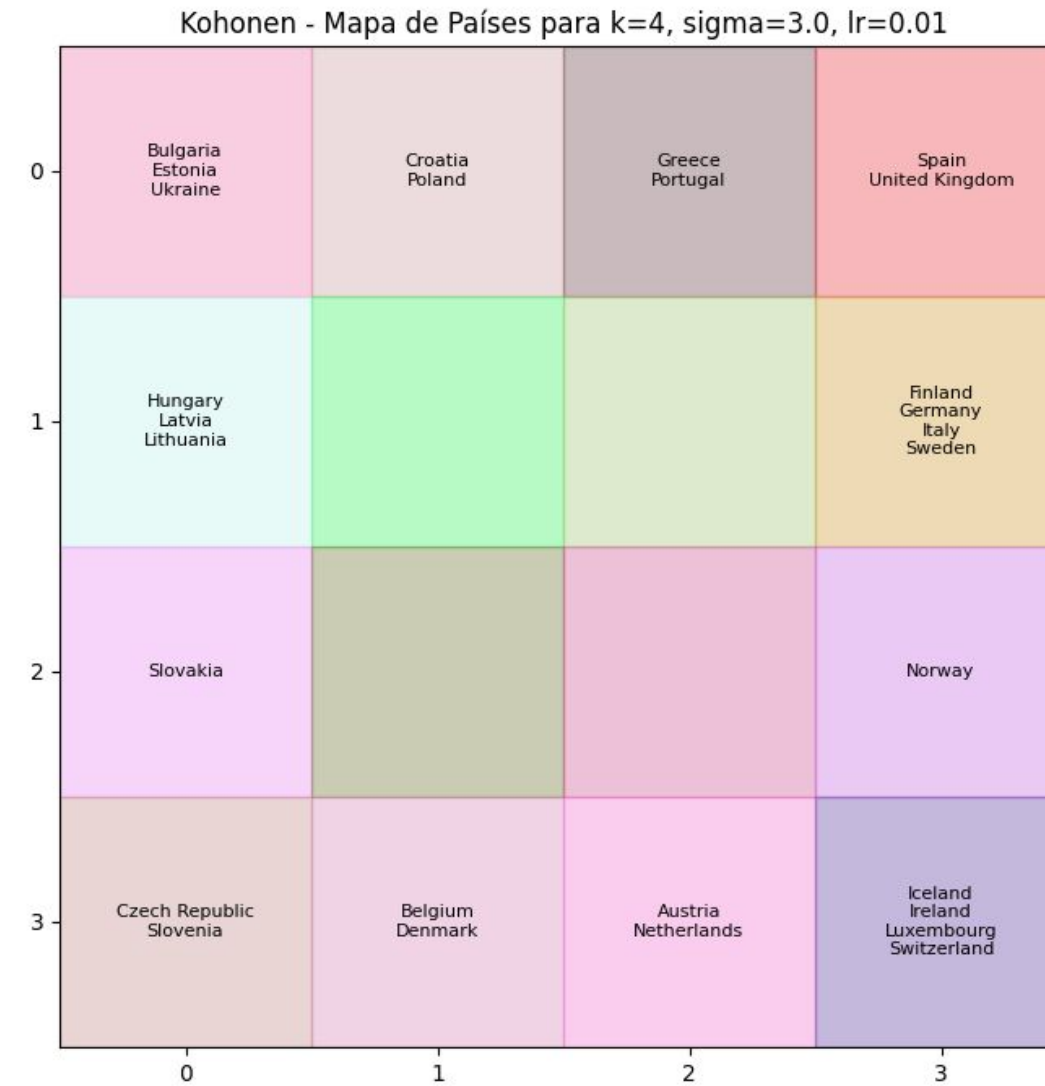
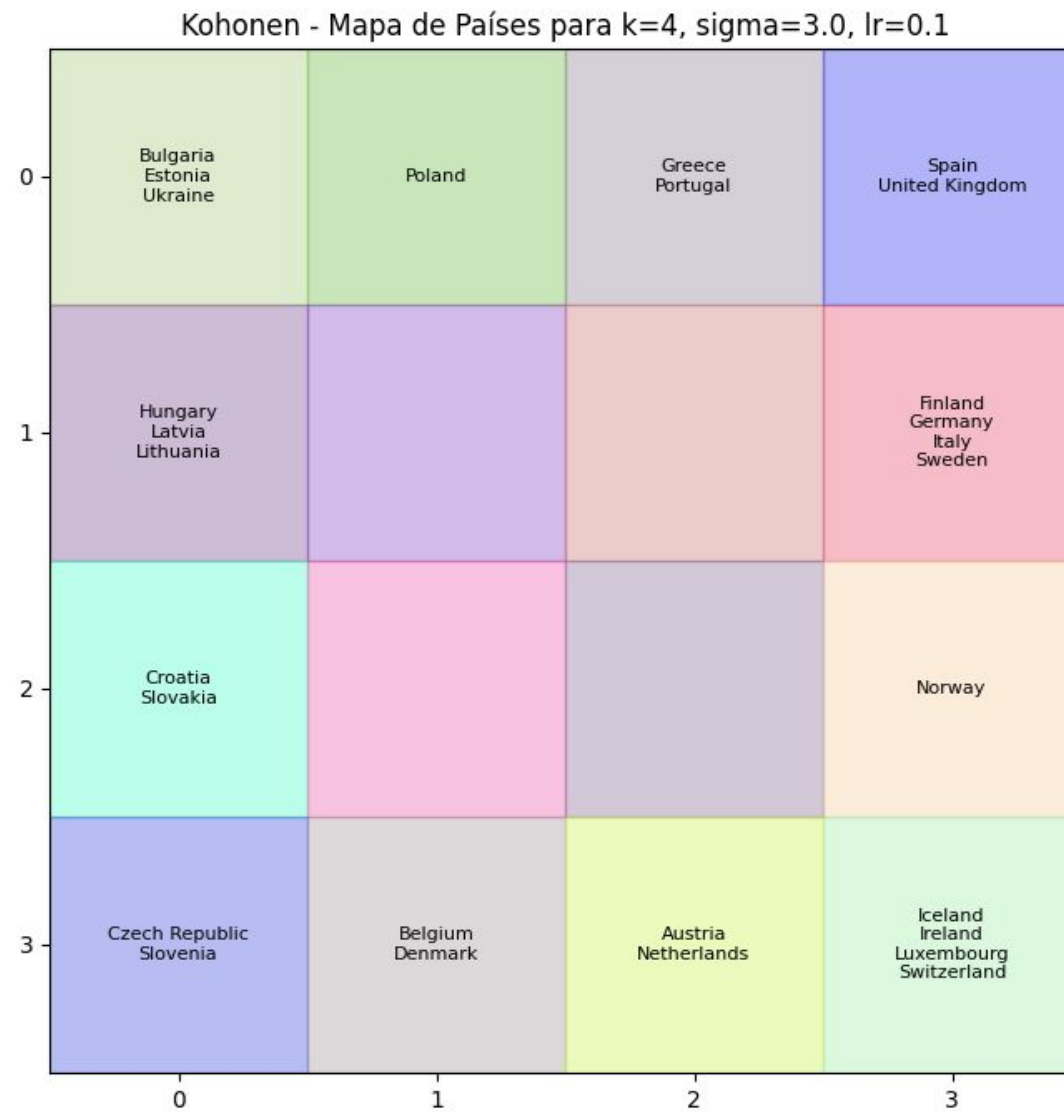
Selección de mejor learning rate

Buscaremos obtener un learning rate que agrupe de manera eficiente los países.

- Learning rate = 0.1
- Learning rate = 0.01
- Learning rate = 0.001



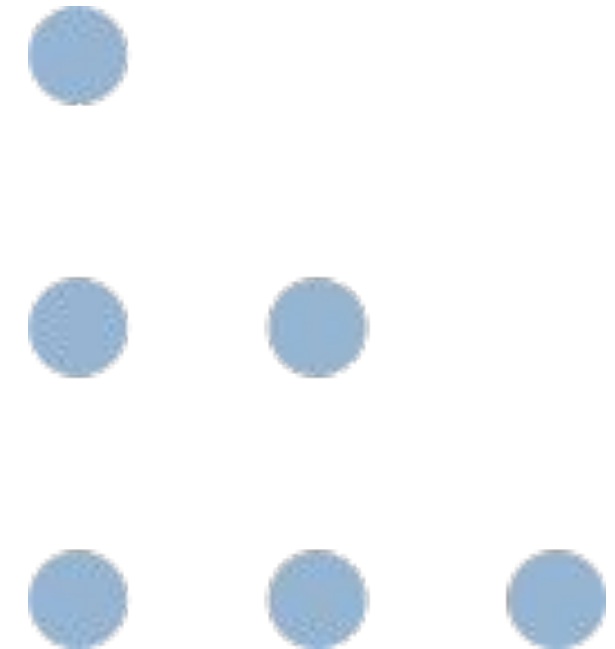
Selección de mejor learning rate



Selección de mejor learning rate

No se notaron diferencias por lo que nos quedamos con un learning rate de 0.1

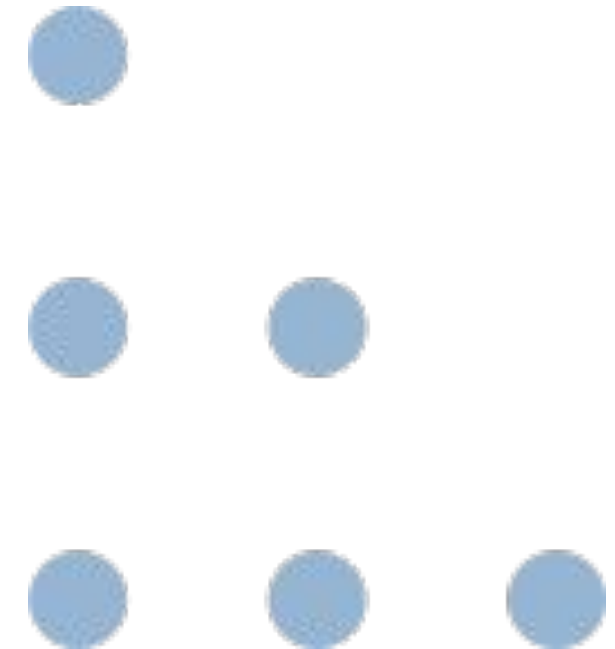
- Learning rate = 0.1
- Learning rate = 0.01
- Learning rate = 0.001



Selección de mejor Sigma

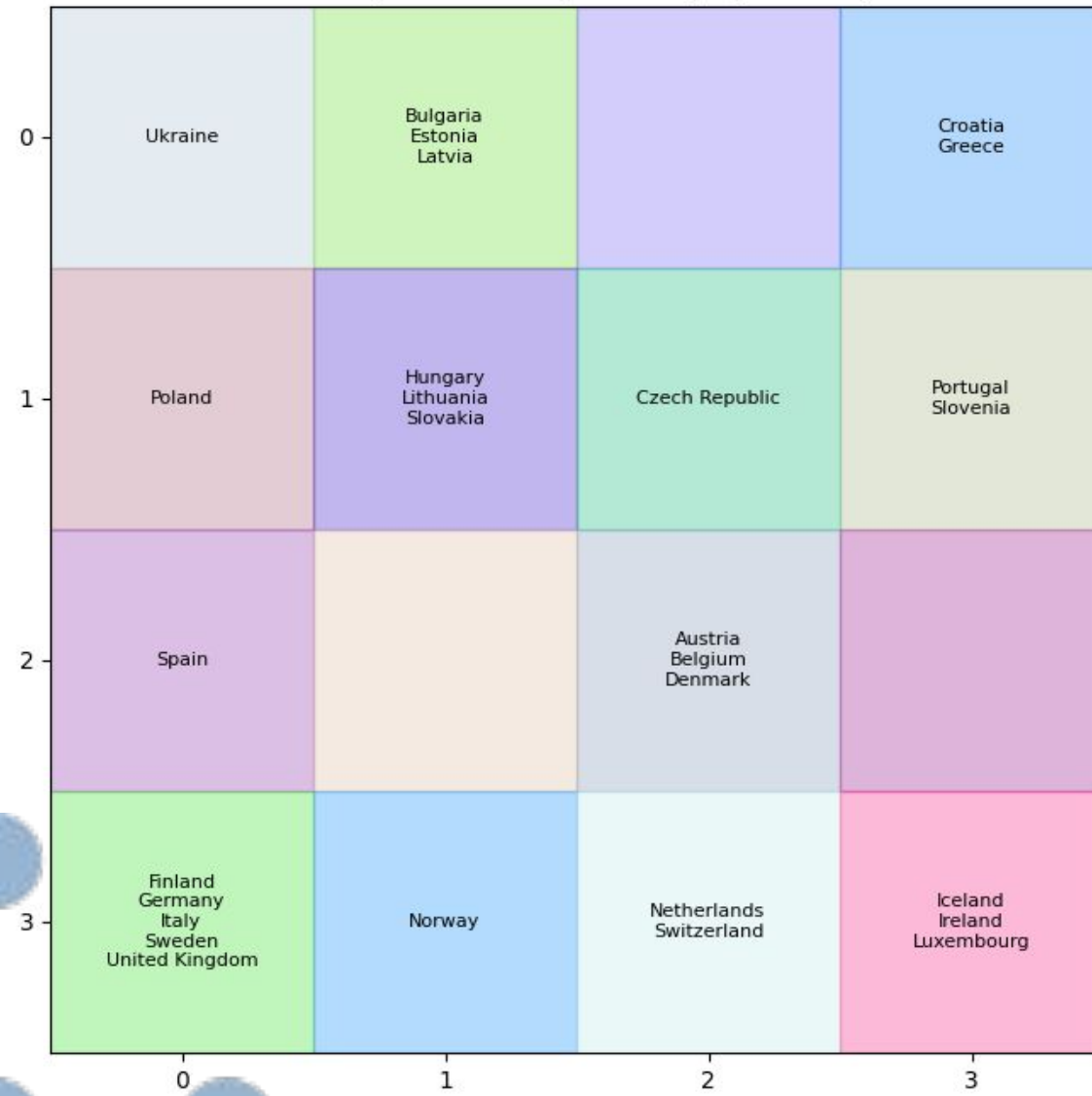
Buscaremos obtener un radio de vecindad que agrupe de manera eficiente los países.

- Sigma = 1
- Sigma = 2
- Sigma = 3

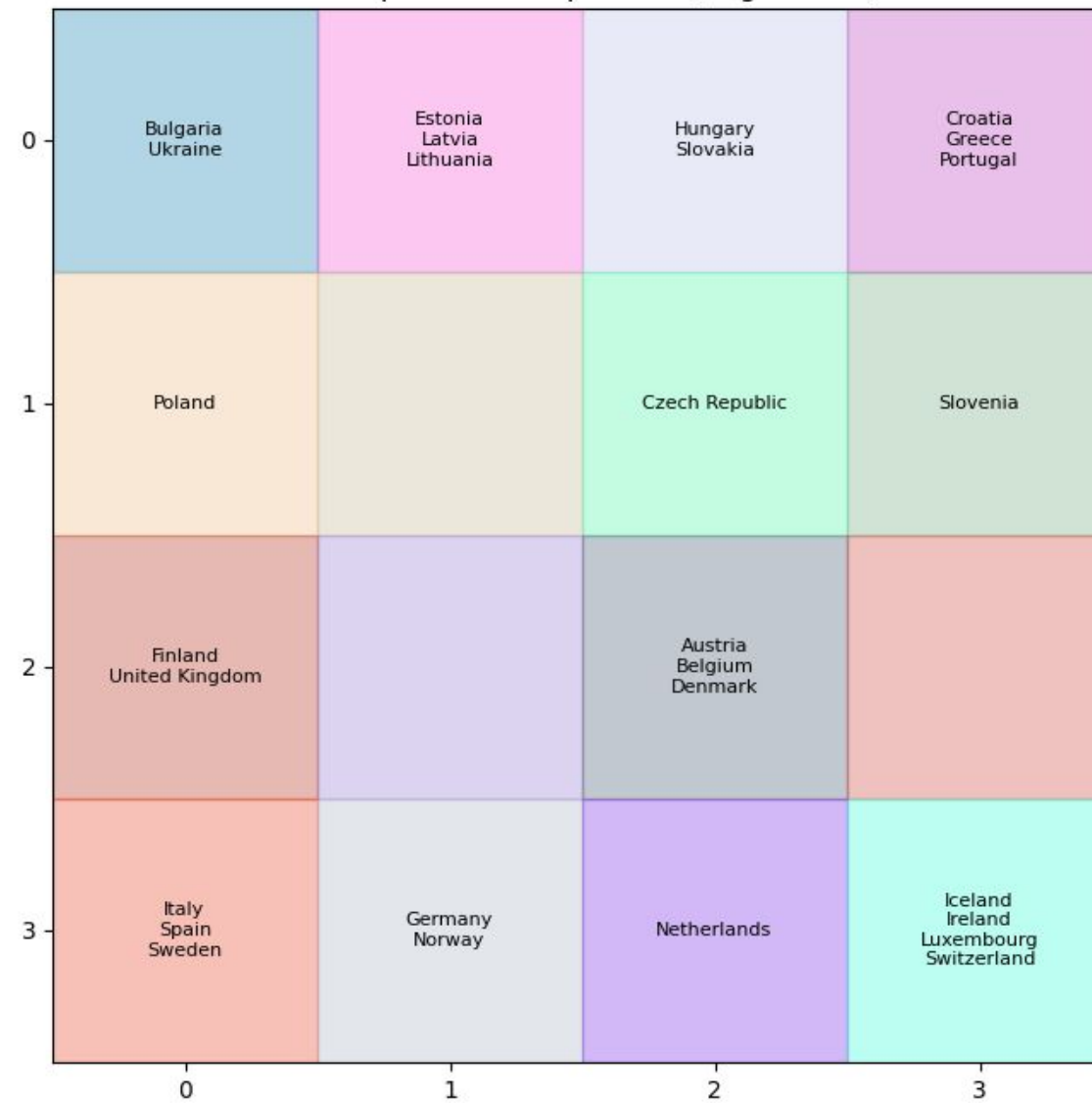


Selección de mejor Sigma

Kohonen - Mapa de Países para k=4, sigma=1.0, lr=0.1



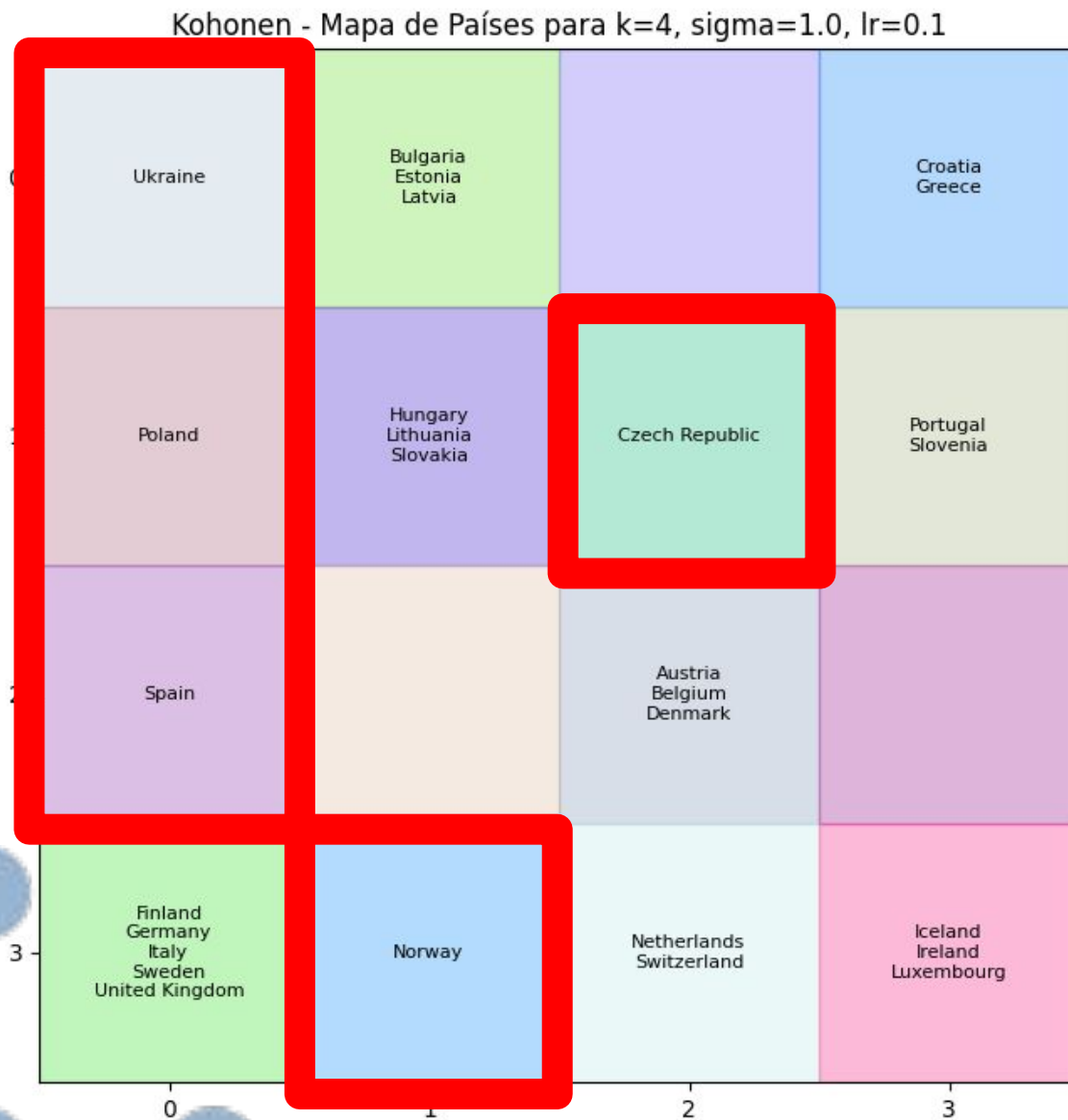
Kohonen - Mapa de Países para k=4, sigma=2.0, lr=0.1



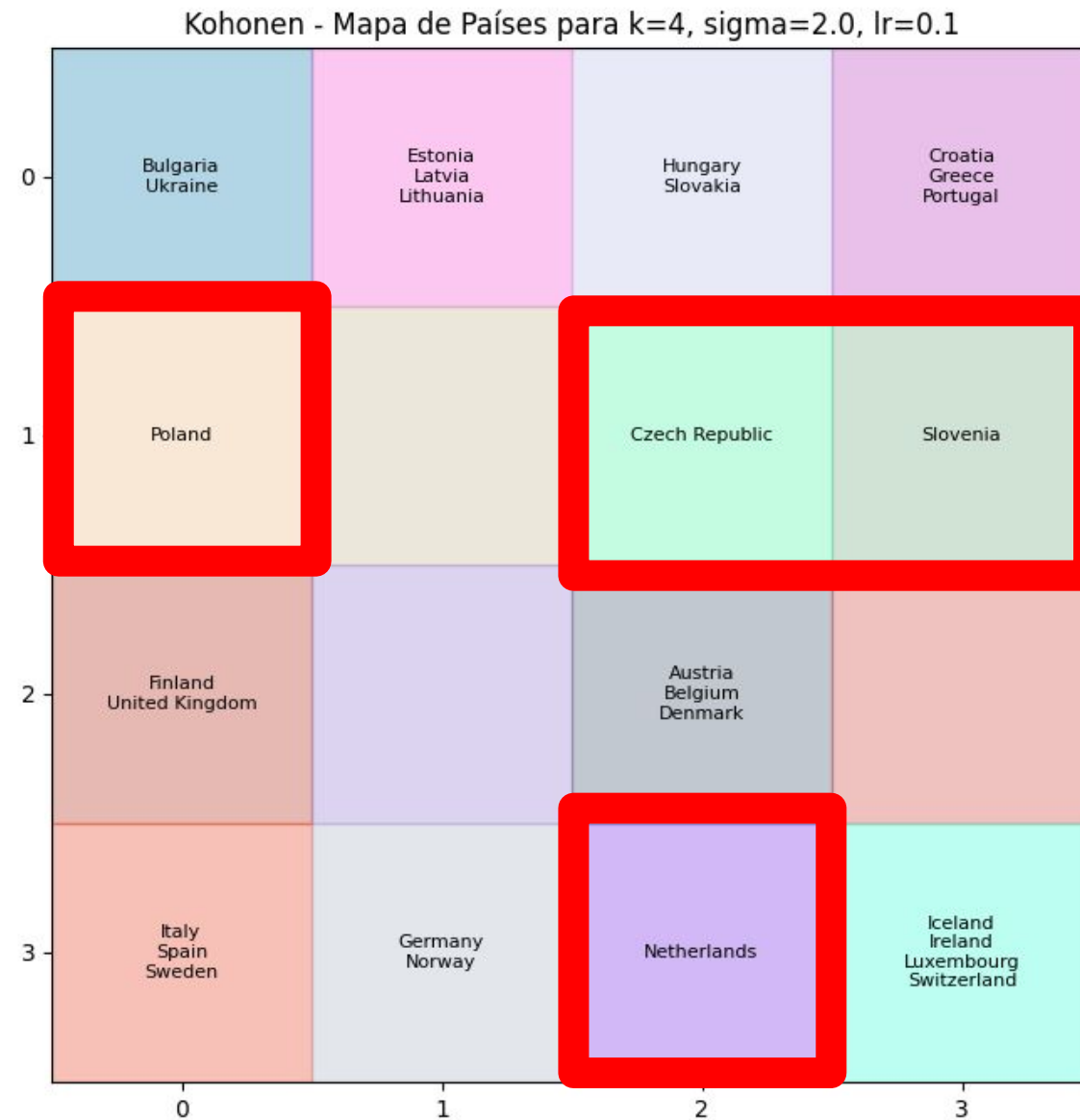
Kohonen - Mapa de Países para k=4, sigma=3.0, lr=0.1



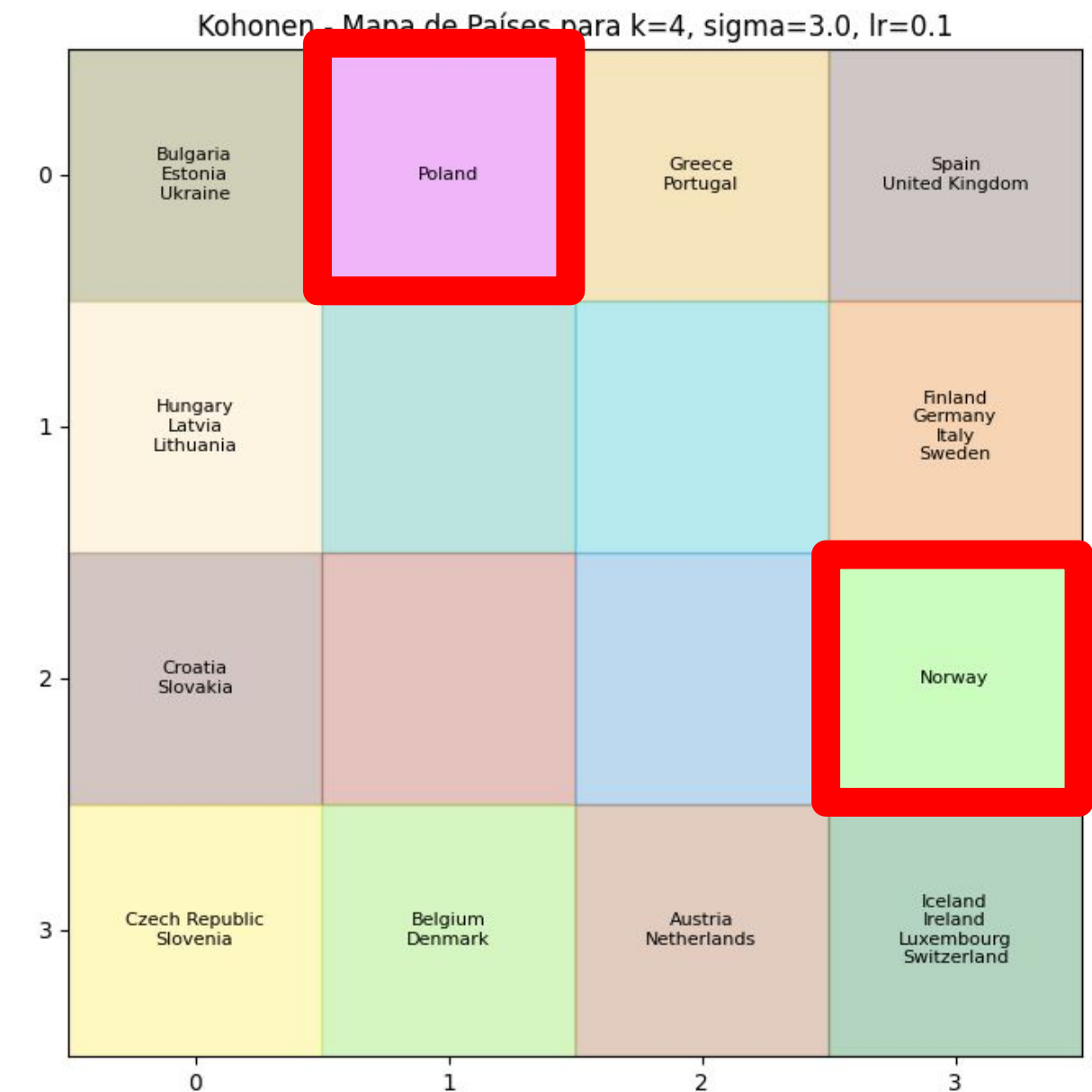
Selección de mejor Sigma



5



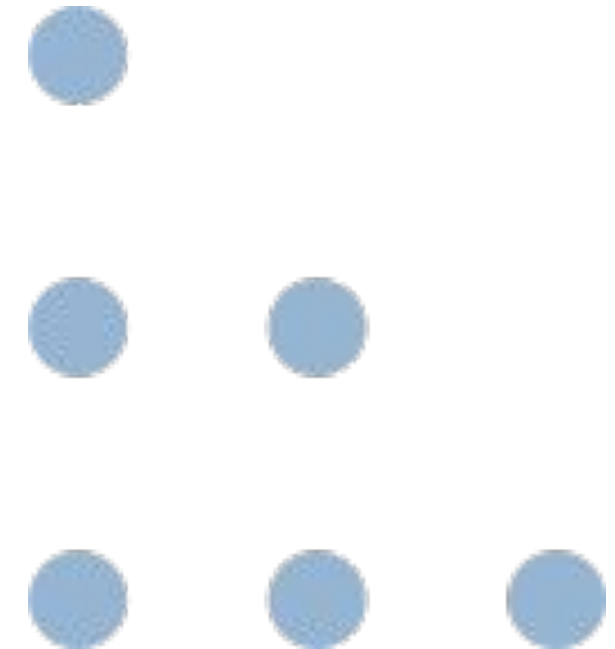
4



2

Selección de mejor Sigma

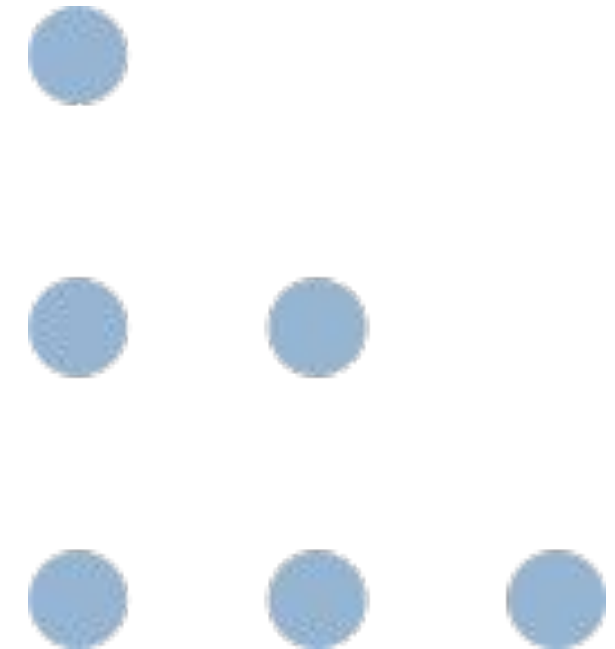
- Sigma = 1 (5 grupos con solo un elemento)
- Sigma = 2 (4 grupos con solo un elemento)
- Sigma = 3 (2 grupos con solo un elemento)



Selección de mejor K

Buscaremos obtener un K que agrupe de manera eficiente los países.

- $K=3$
- $K=4$
- $K=5$

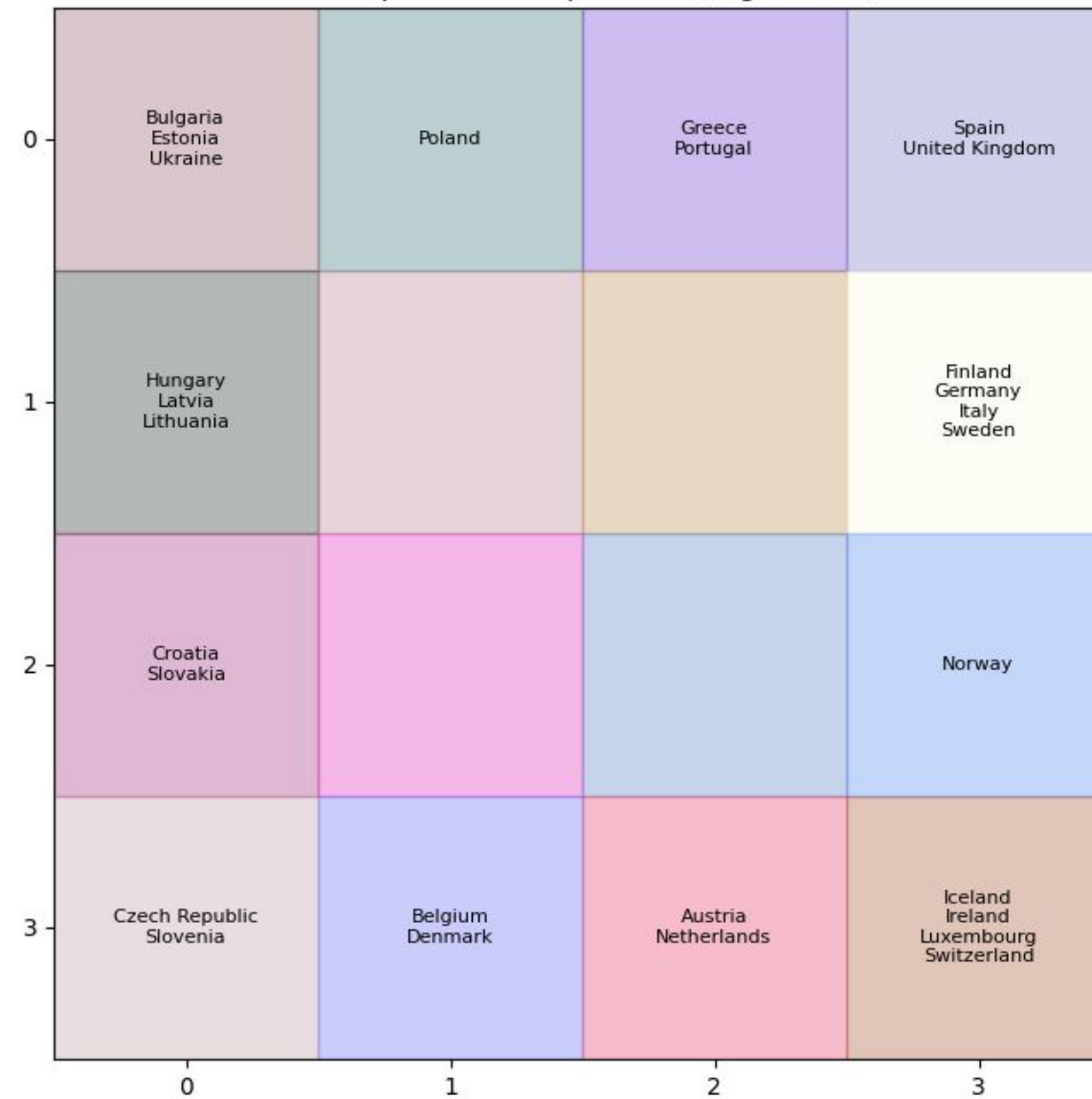


Selección de mejor K

Kohonen - Mapa de Países para k=3, sigma=3.0, lr=0.1



Kohonen - Mapa de Países para k=4, sigma=3.0, lr=0.1

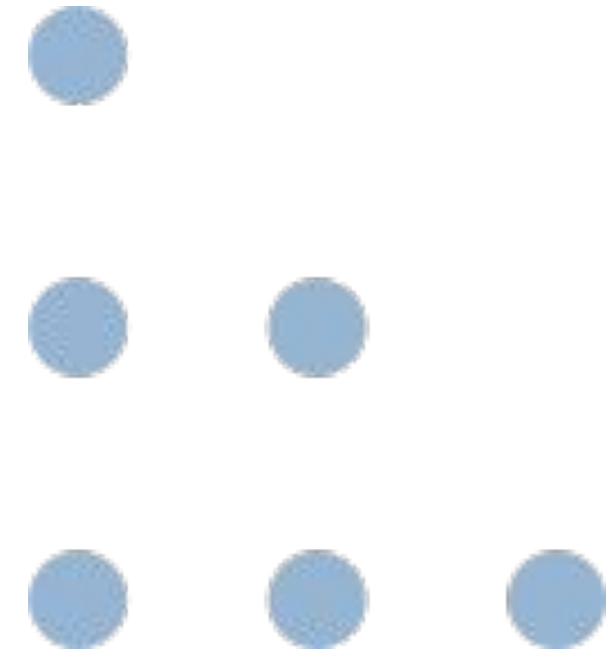


Kohonen - Mapa de Países para k=5, sigma=3.0, lr=0.1



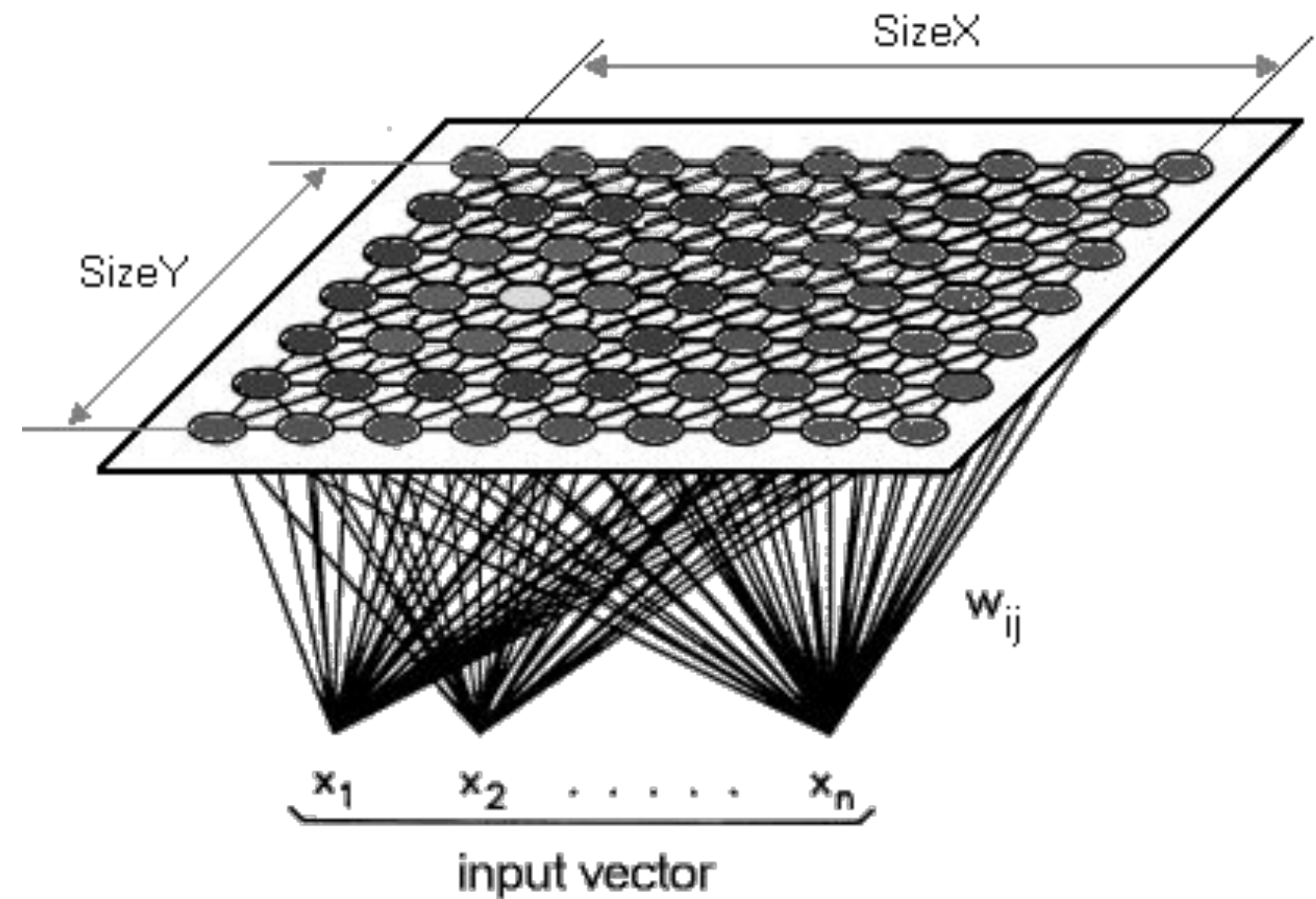
Selección de mejor K

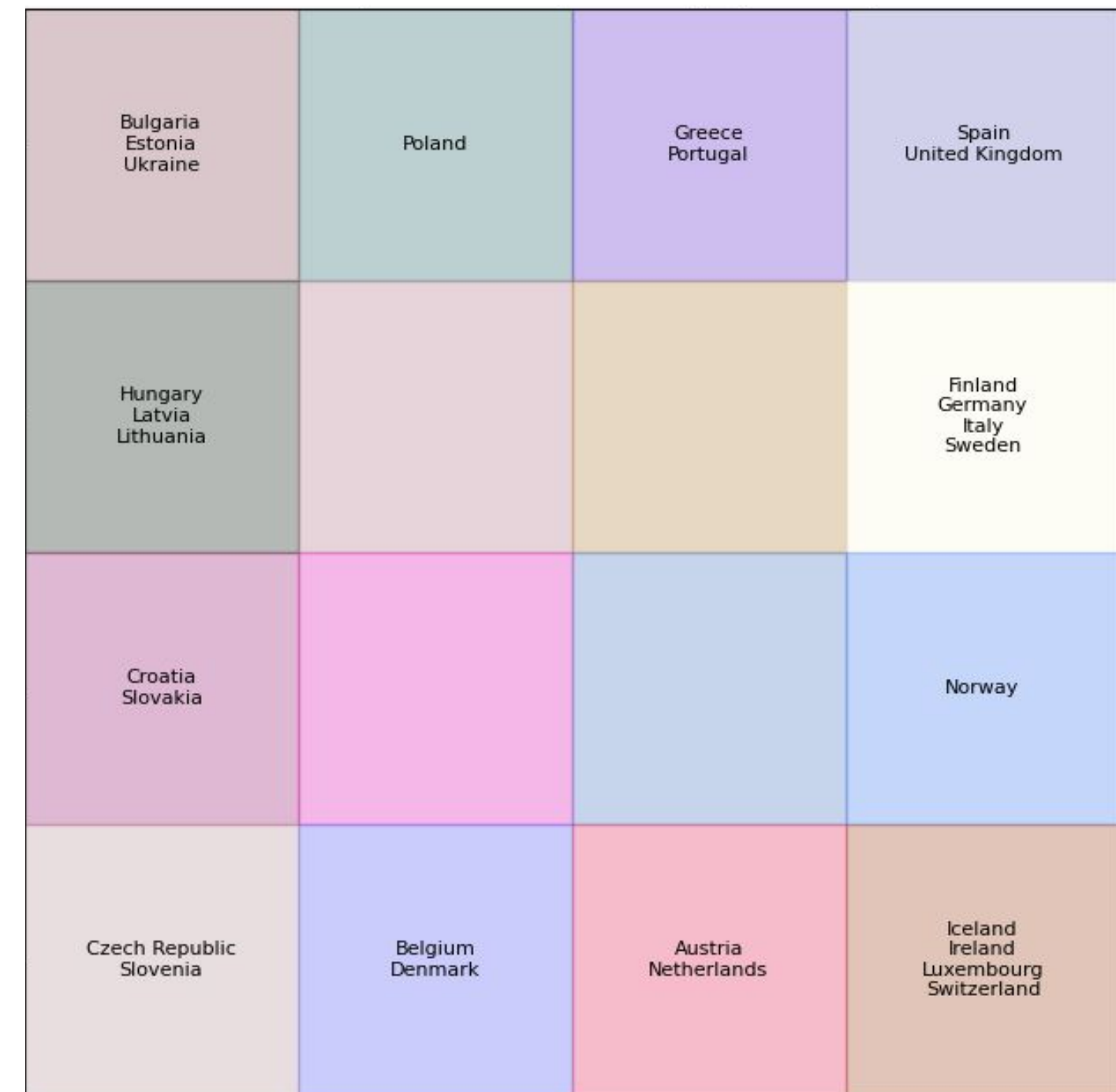
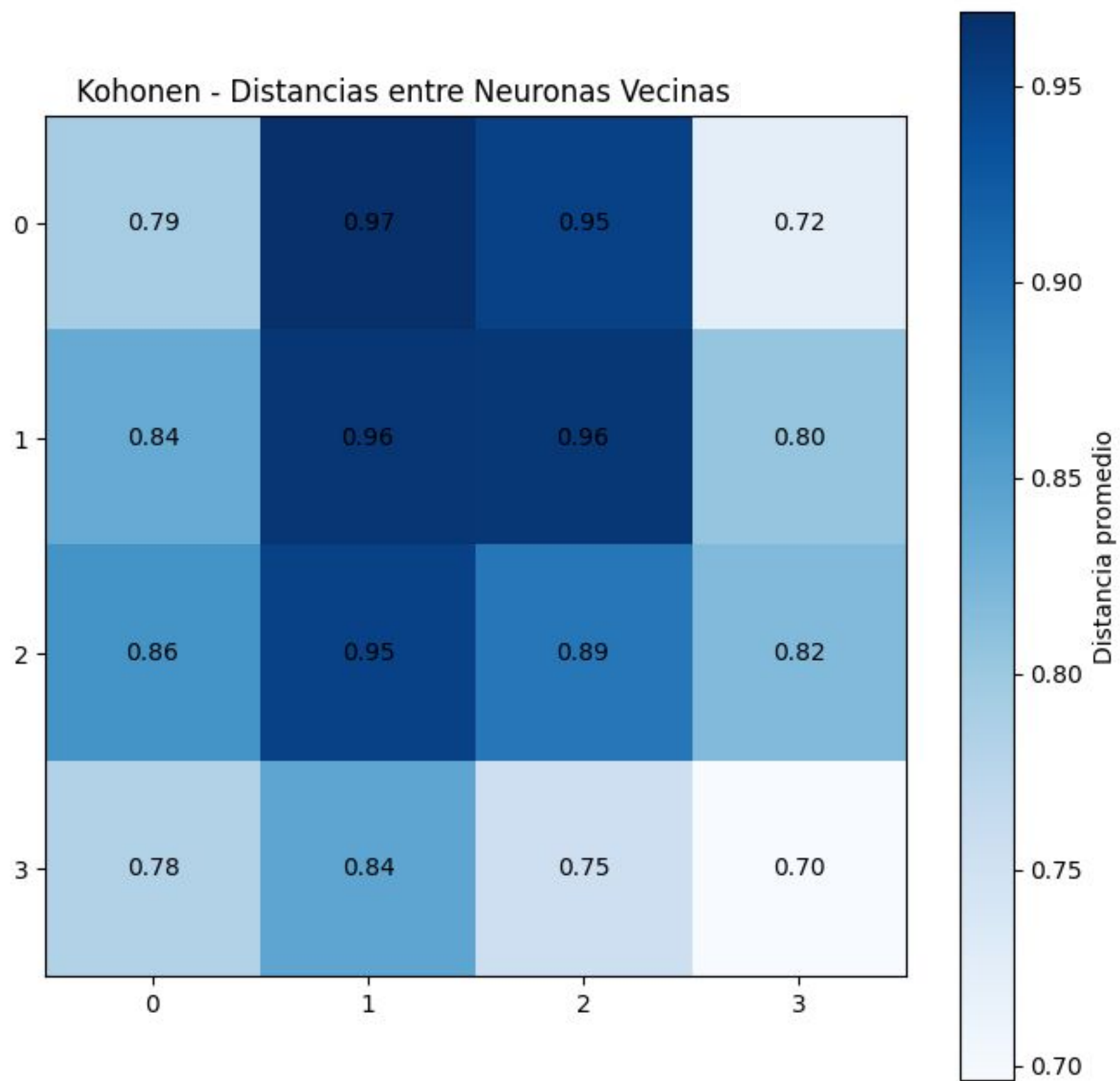
- $K=3$ (Muy pocos grupos distinguibles)
- $K=4$ (Ideal)
- $K=5$ (Demasiados grupos, varios con solo un elemento)

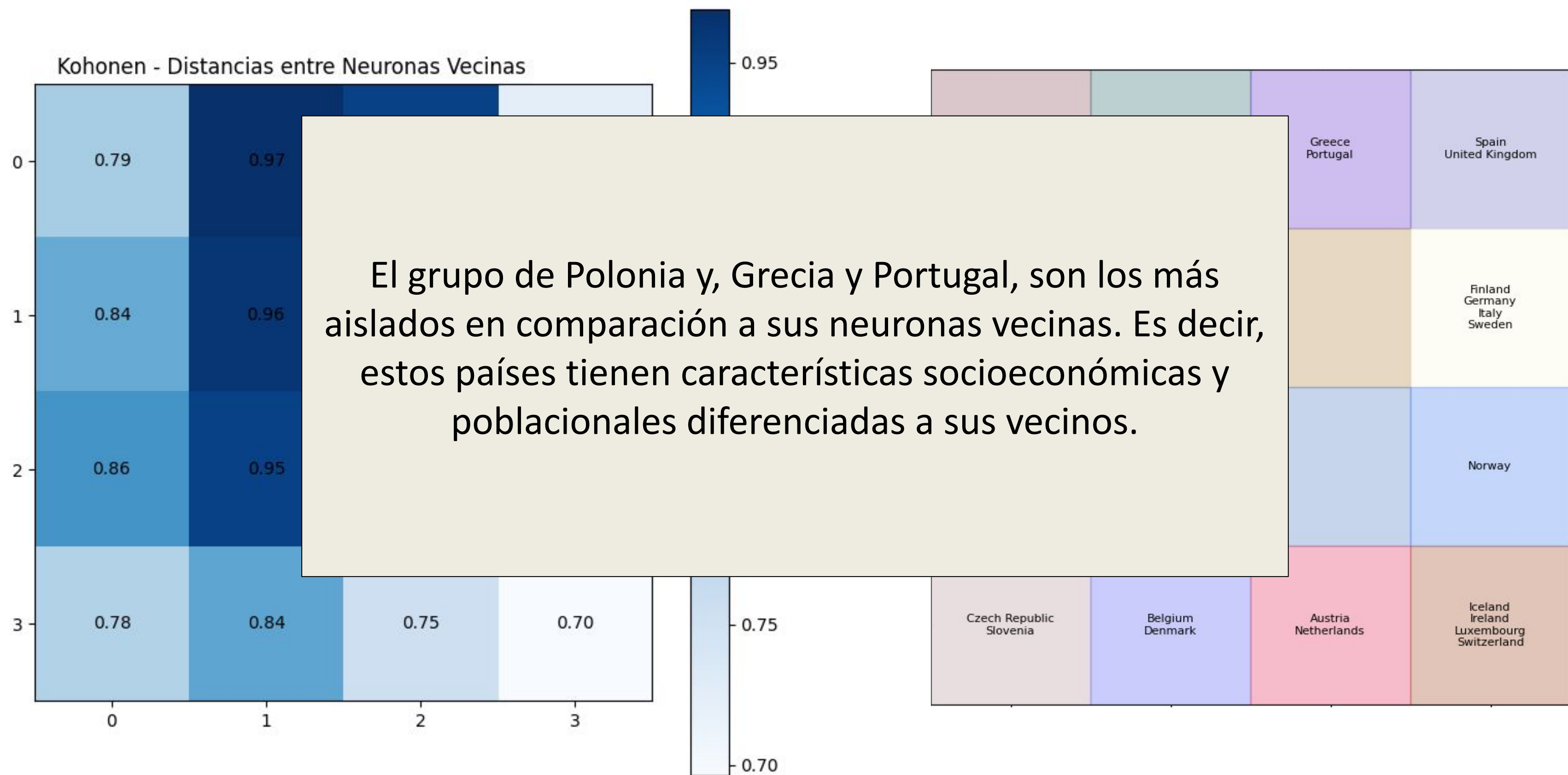


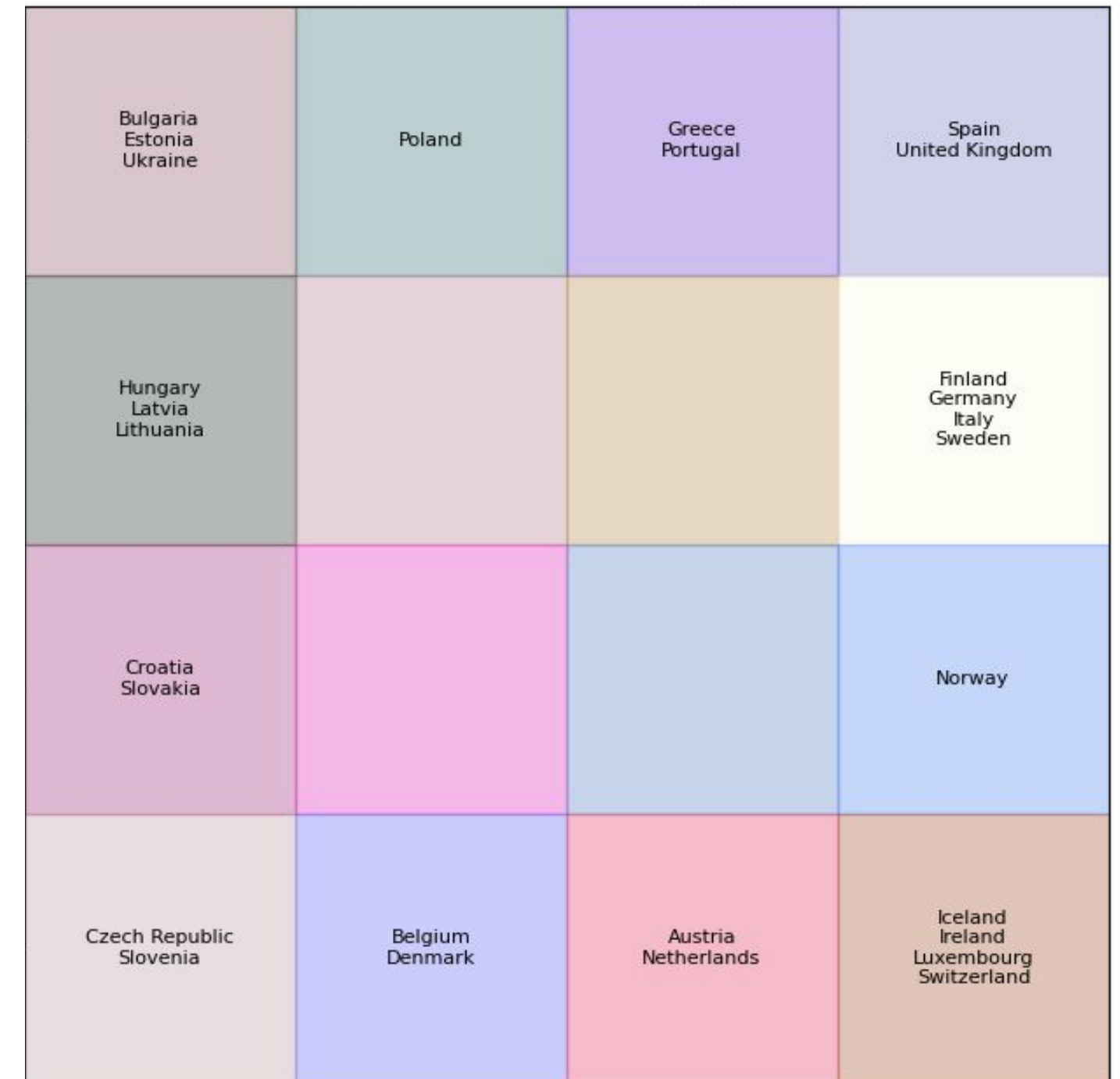
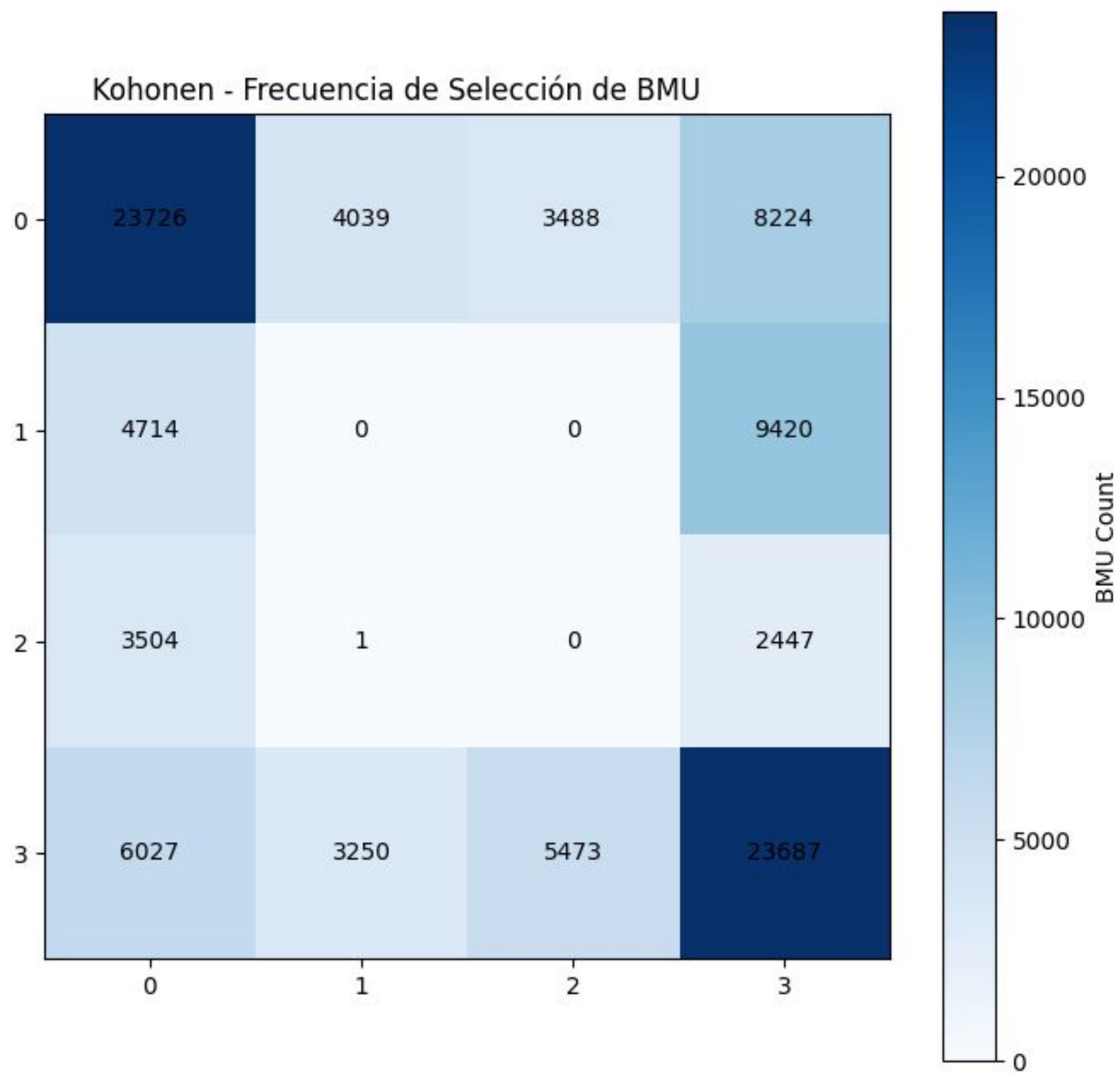
Parámetros establecidos

- Learning rate = 0.1
- Sigma = 3
- K = 4

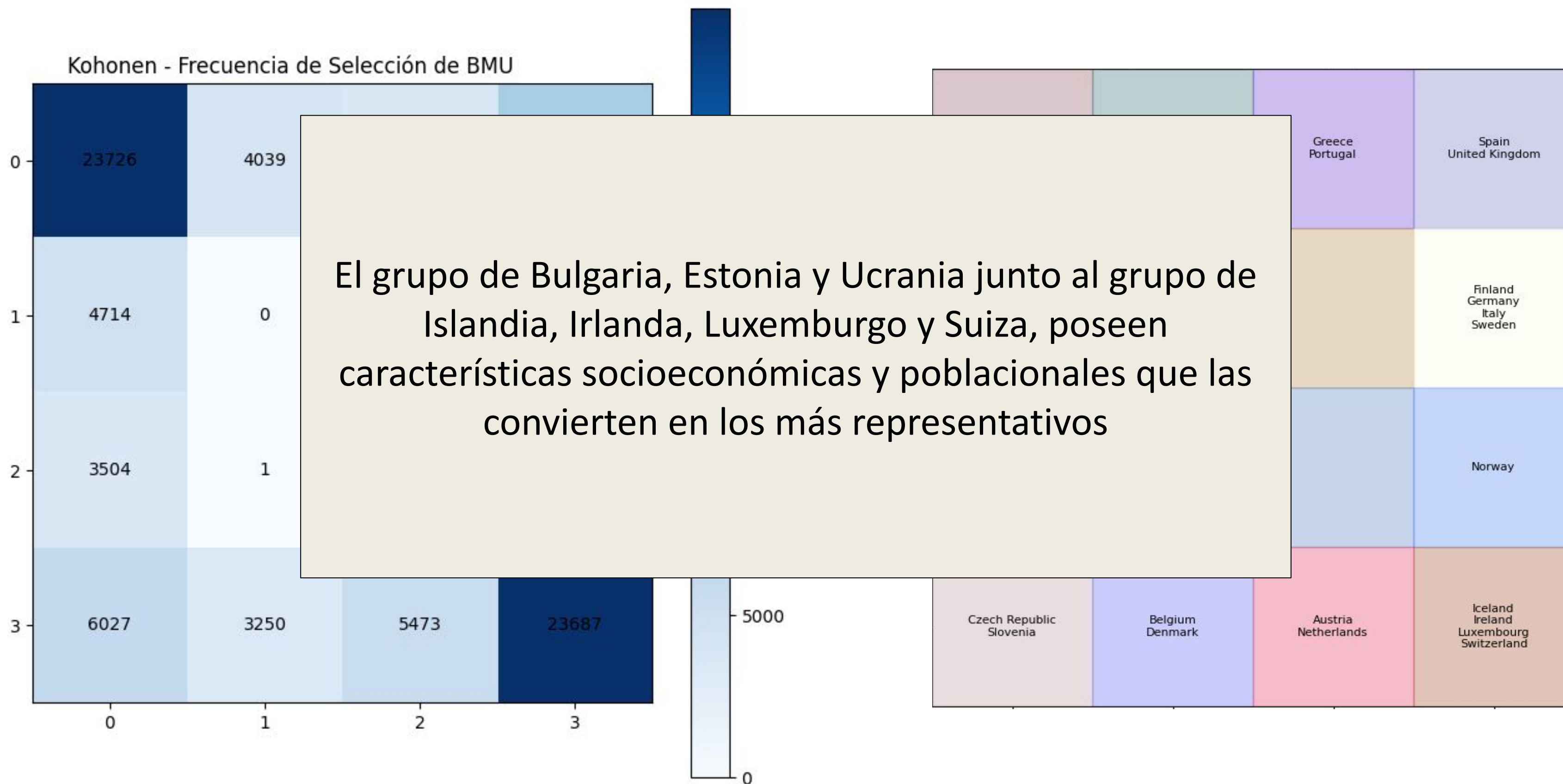






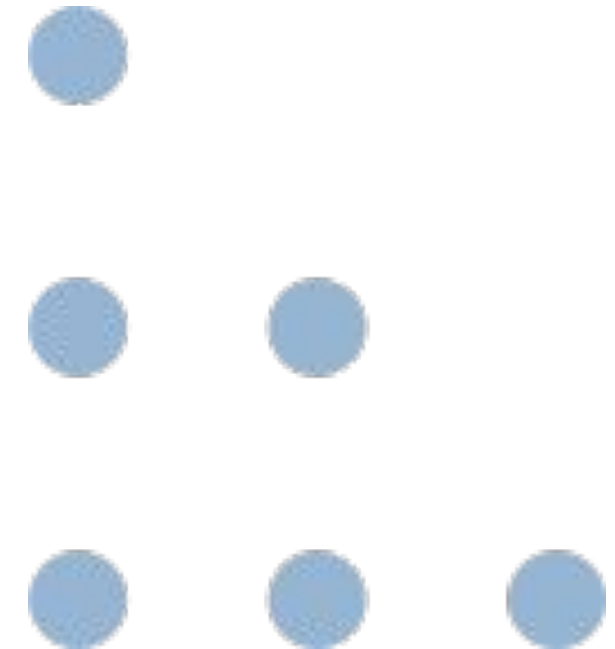


Kohonen - Frecuencia de Selección de BMU

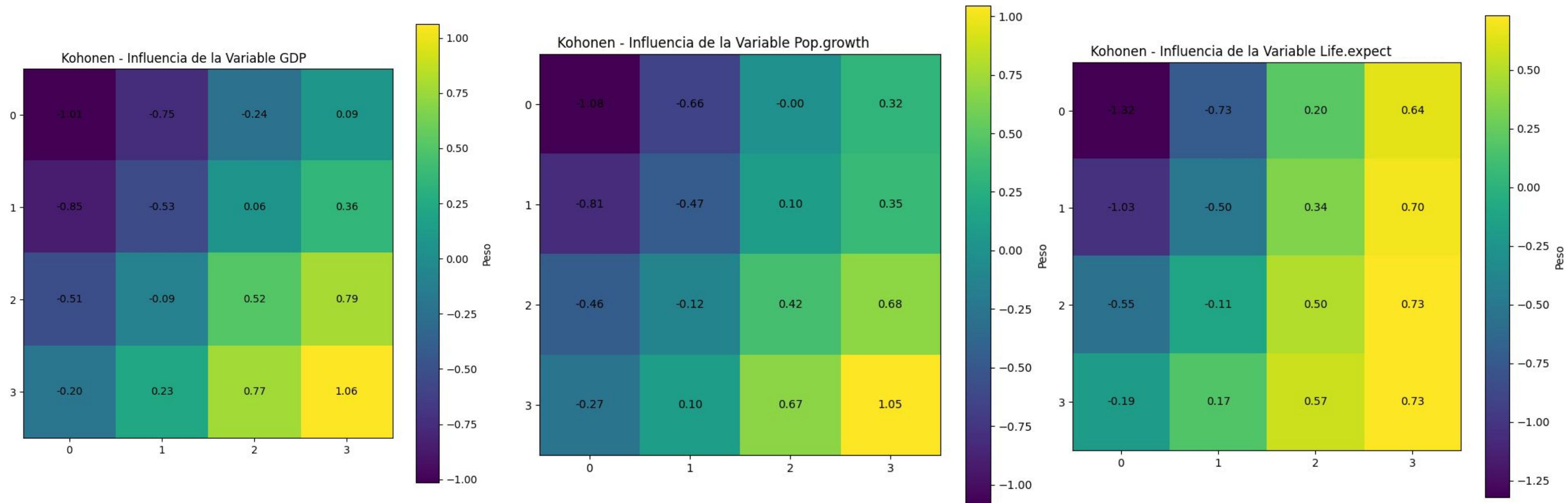


Análisis de variables independientes

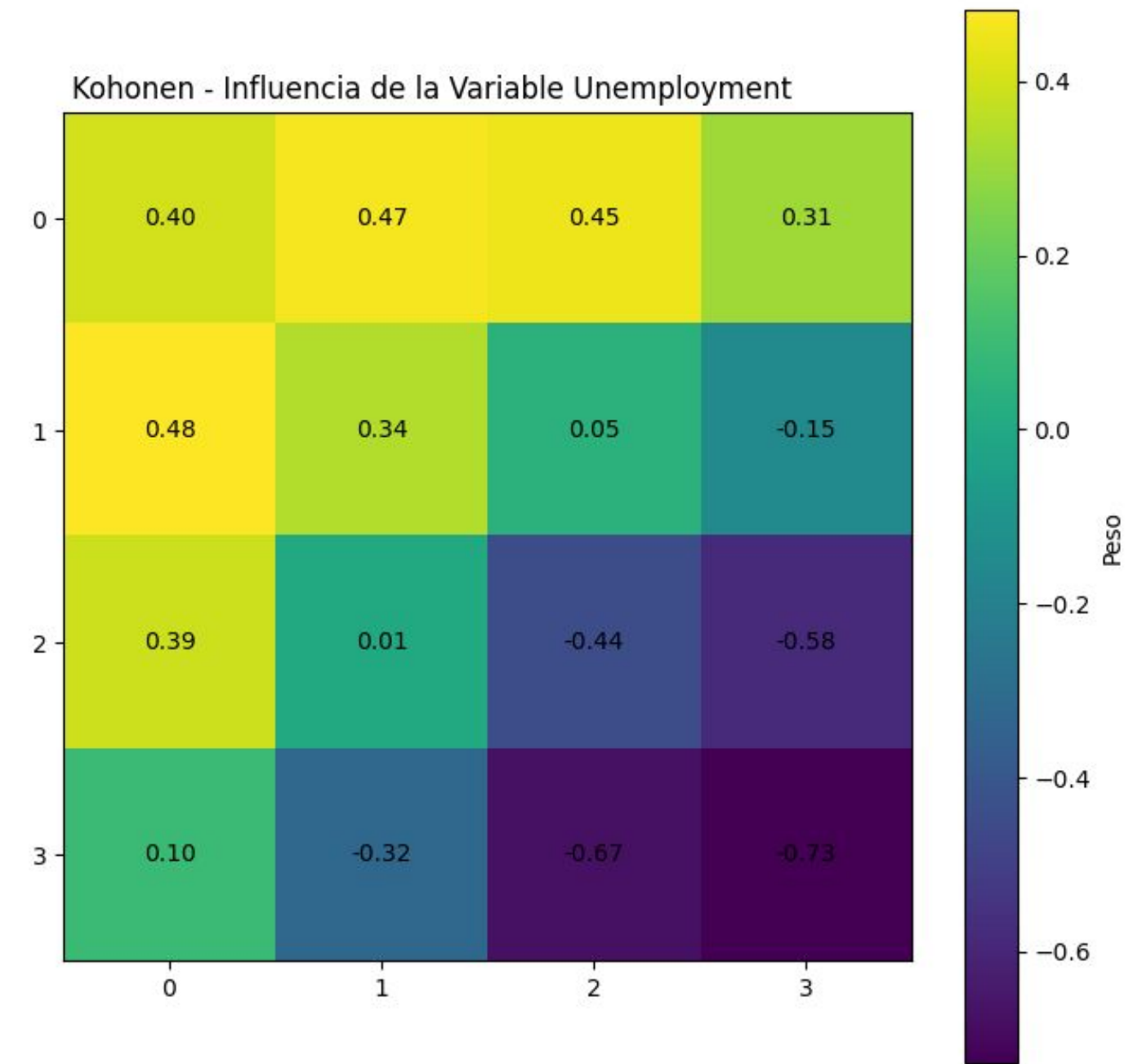
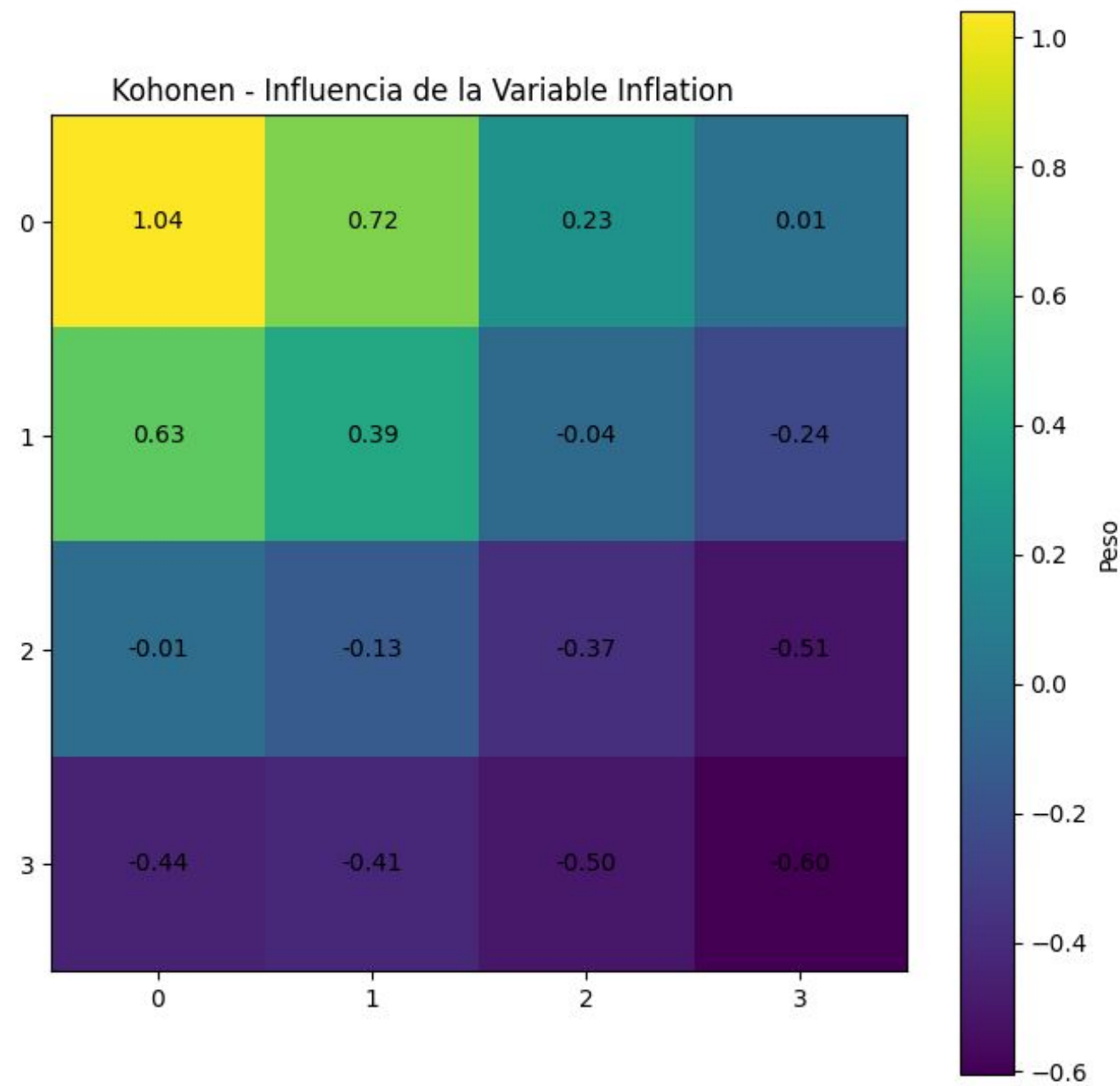
Veamos qué conclusiones podemos tener si aislamos las variables socioeconómicas para cada grupo y nos fijamos su respectivo peso aprovechando la salida matricial de Kohonen.



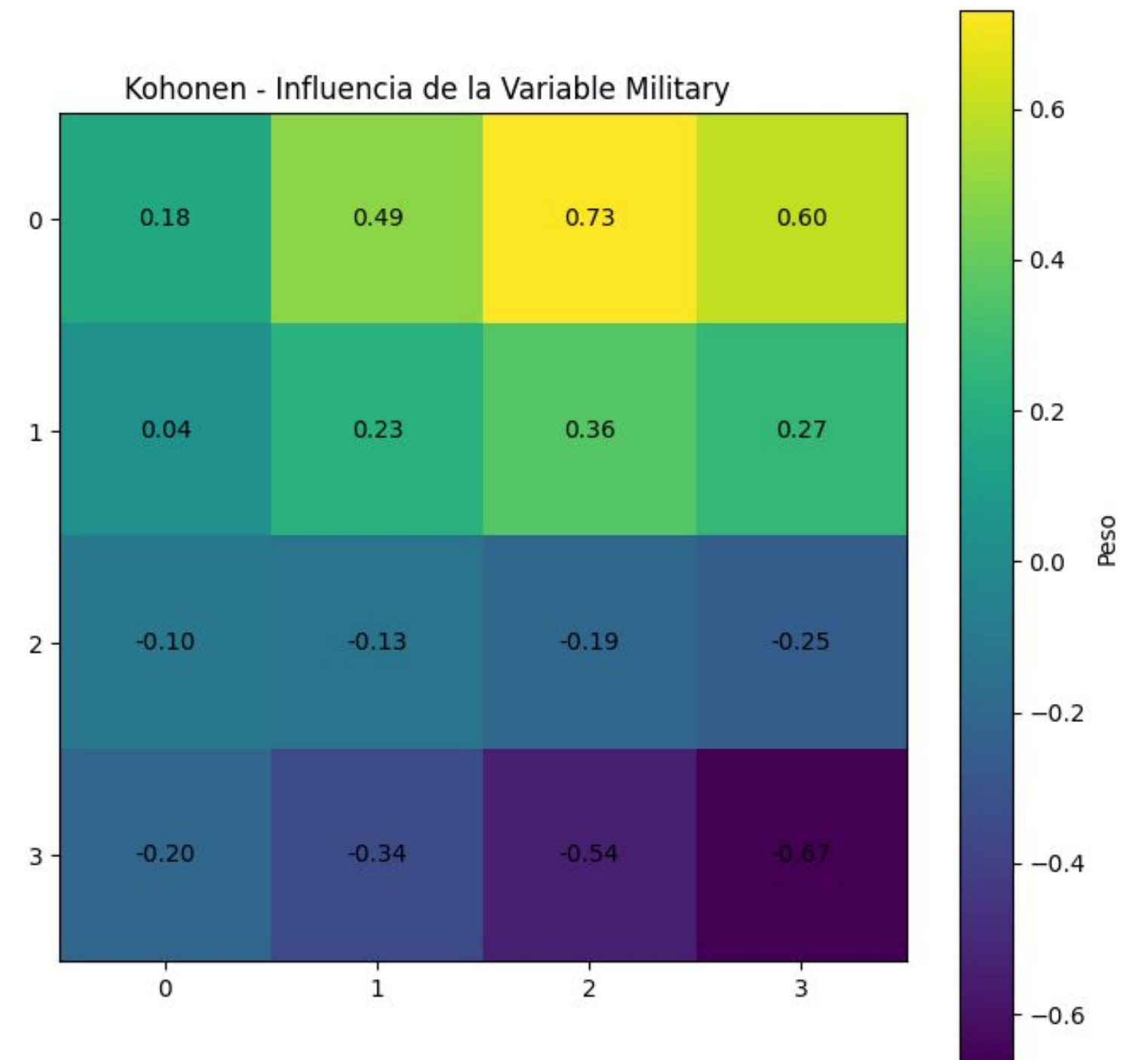
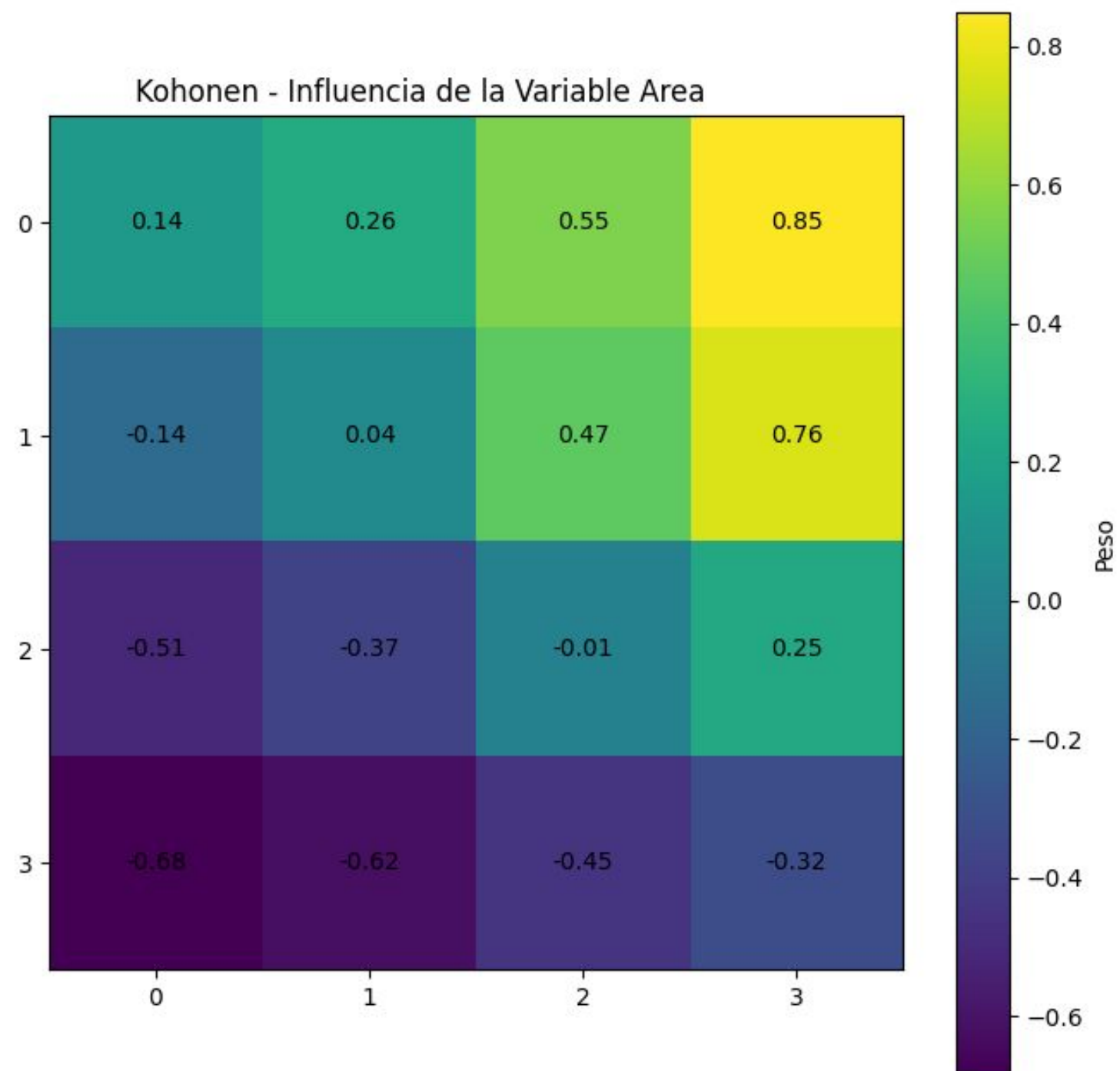
GDP - Pop.growth - Life.expect



Inflation - Unemployment

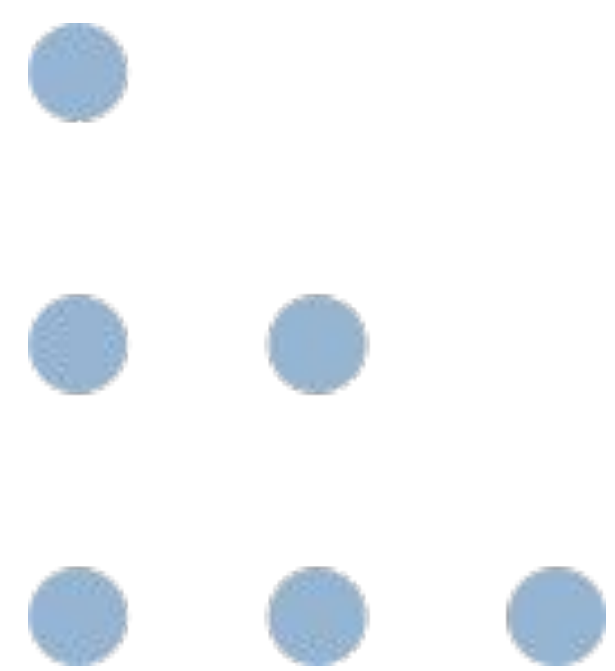


Area - Military



Analisis de variables

- El **GDP**, **crecimiento poblacional** y **expectativa de vida** de un país crecen y disminuyen simultáneamente
- La **inflación** y **desempleo** de un país crecen y disminuyen simultáneamente
- **GDP - Pop.growth - Life.expect** son inversas a **Inflation - Unemployment**
 - Países con un gran **Área** territorial tienden a tener más gasto **Militar**





SIA 2024 - 2C

Modelo de Oja

Ejercicio 1.2



Implementación

train

```
for epoch in range (epochs)
    for i=1 to N
        y = inner(xi, w )
        w +=  $\eta$  * y * (xi - y * w )
```

test

```
for i=1 to N
    y = inner(xi, w )
    pc1_list.append(y)
```

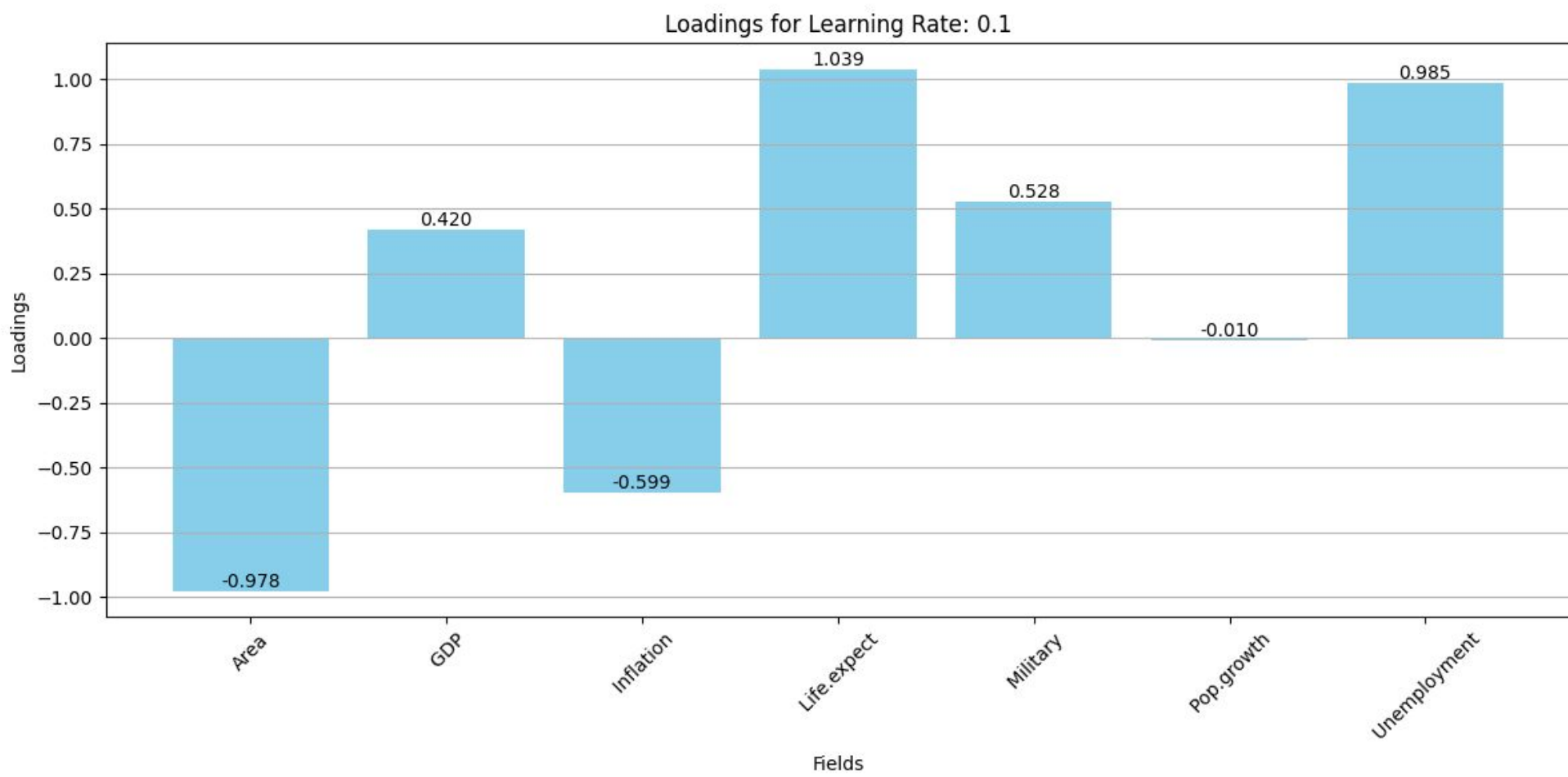
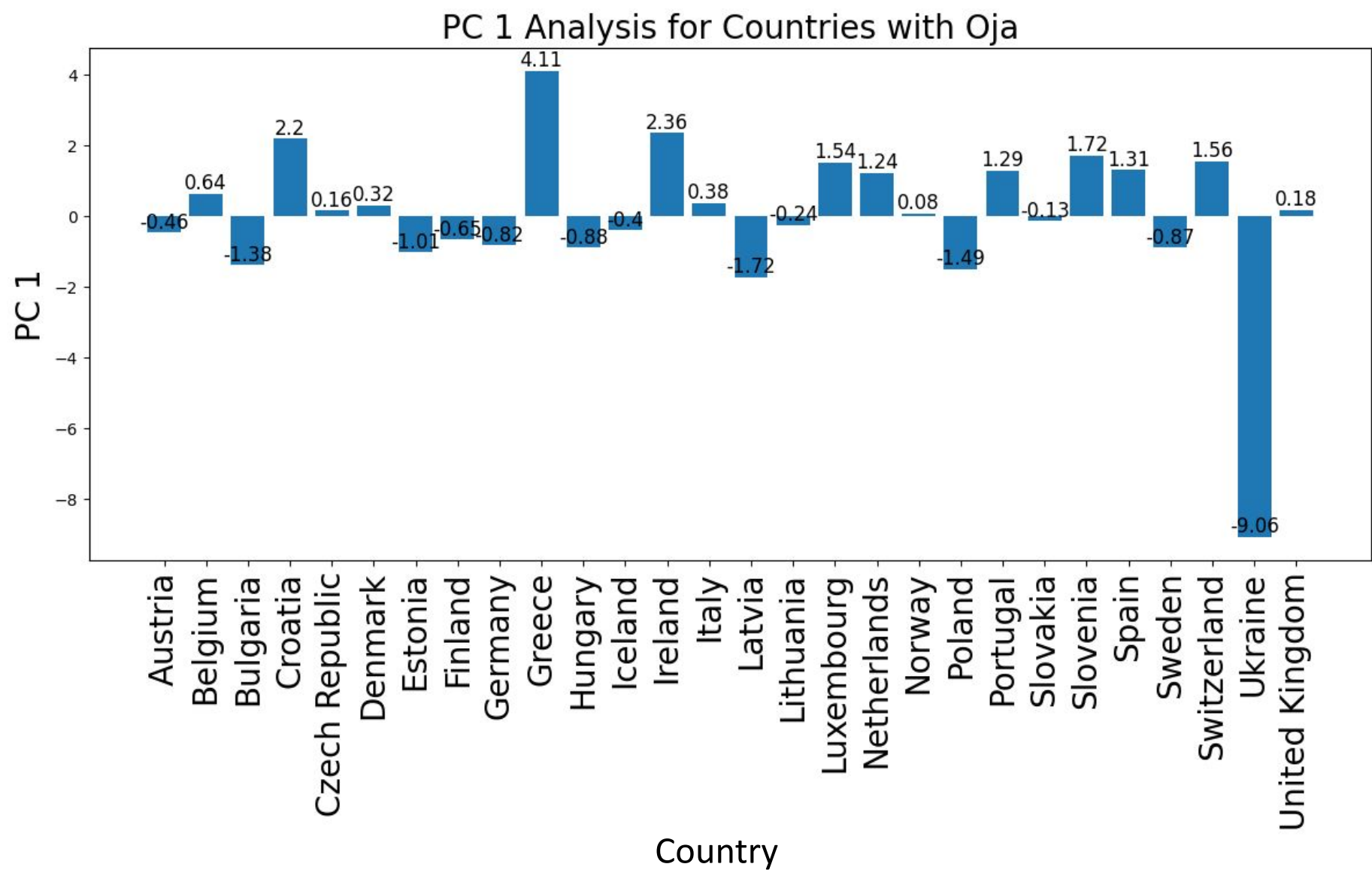
- Pesos inicializados de manera random.
- Se utilizó una seed fija a lo largo de las ejecuciones.
- Se optó por 10k épocas.
- Se estudian distintos learning rate.

Preguntas disparadoras del análisis

- ¿Qué obtenemos del cálculo de la primer componente (PC1) para el conjunto de datos?
- ¿Qué representa la PC1?
- ¿Qué diferencias encontramos entre PCA y Oja?

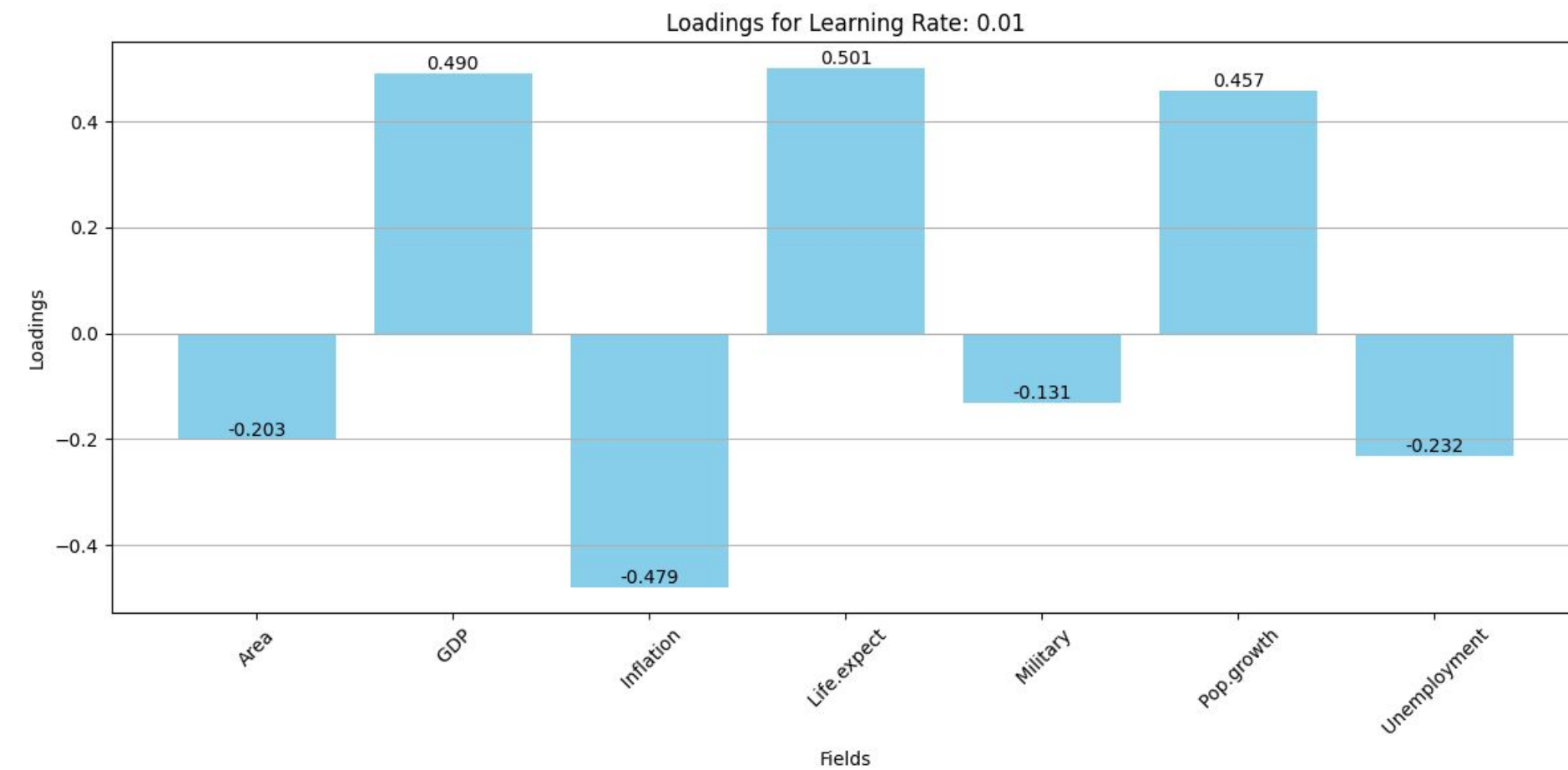
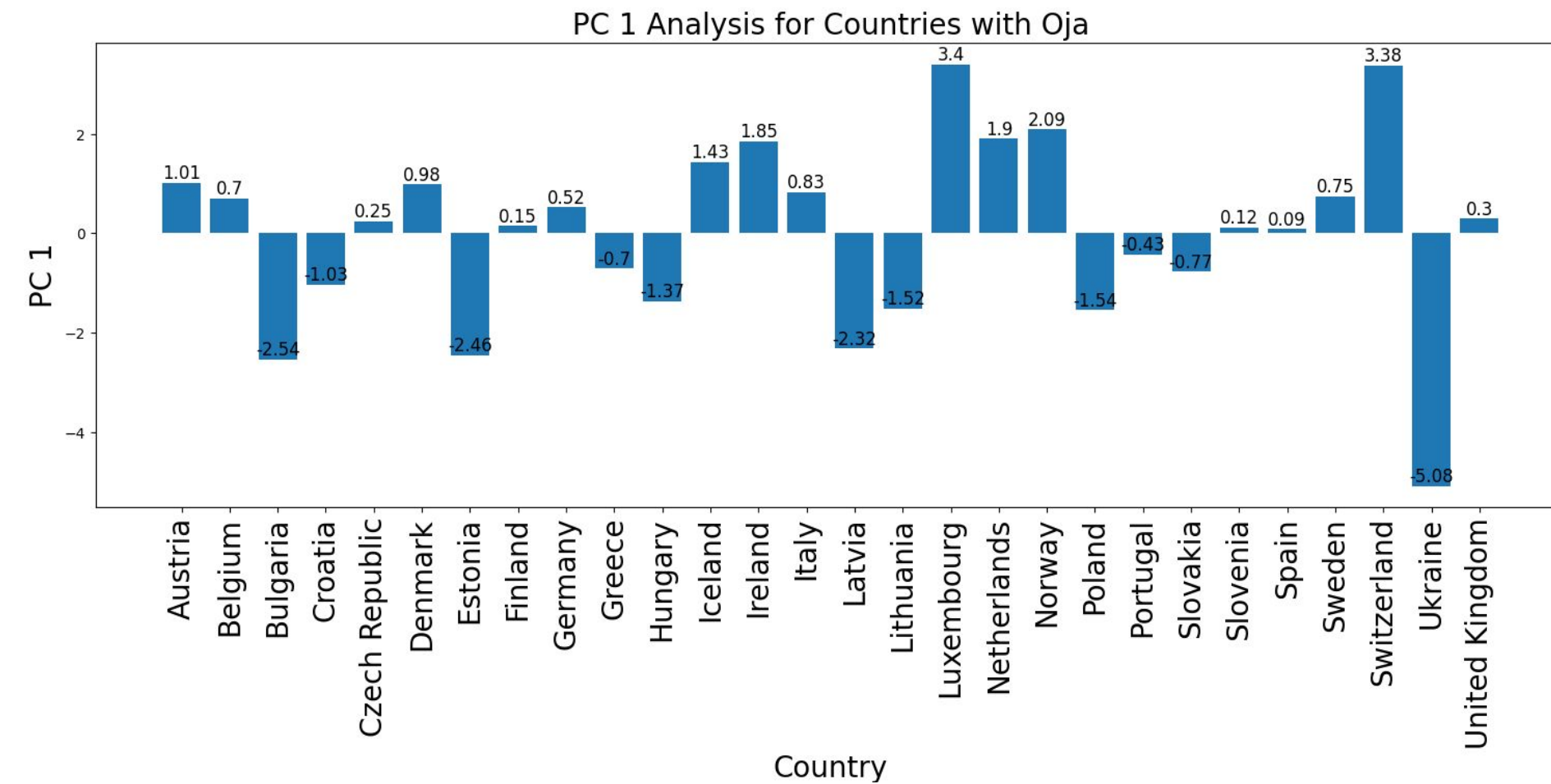


10k épocas y $\eta = 0.1$



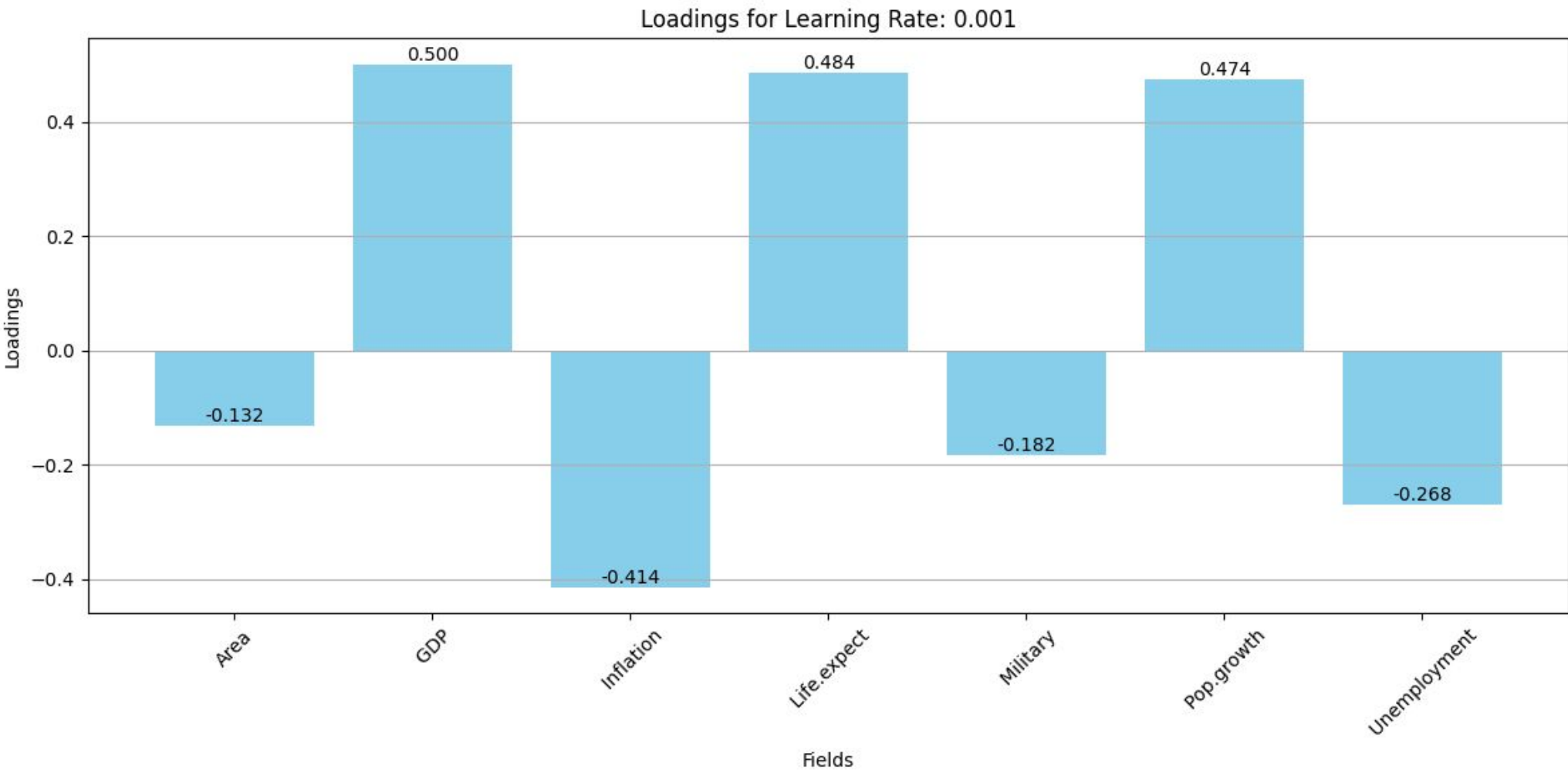
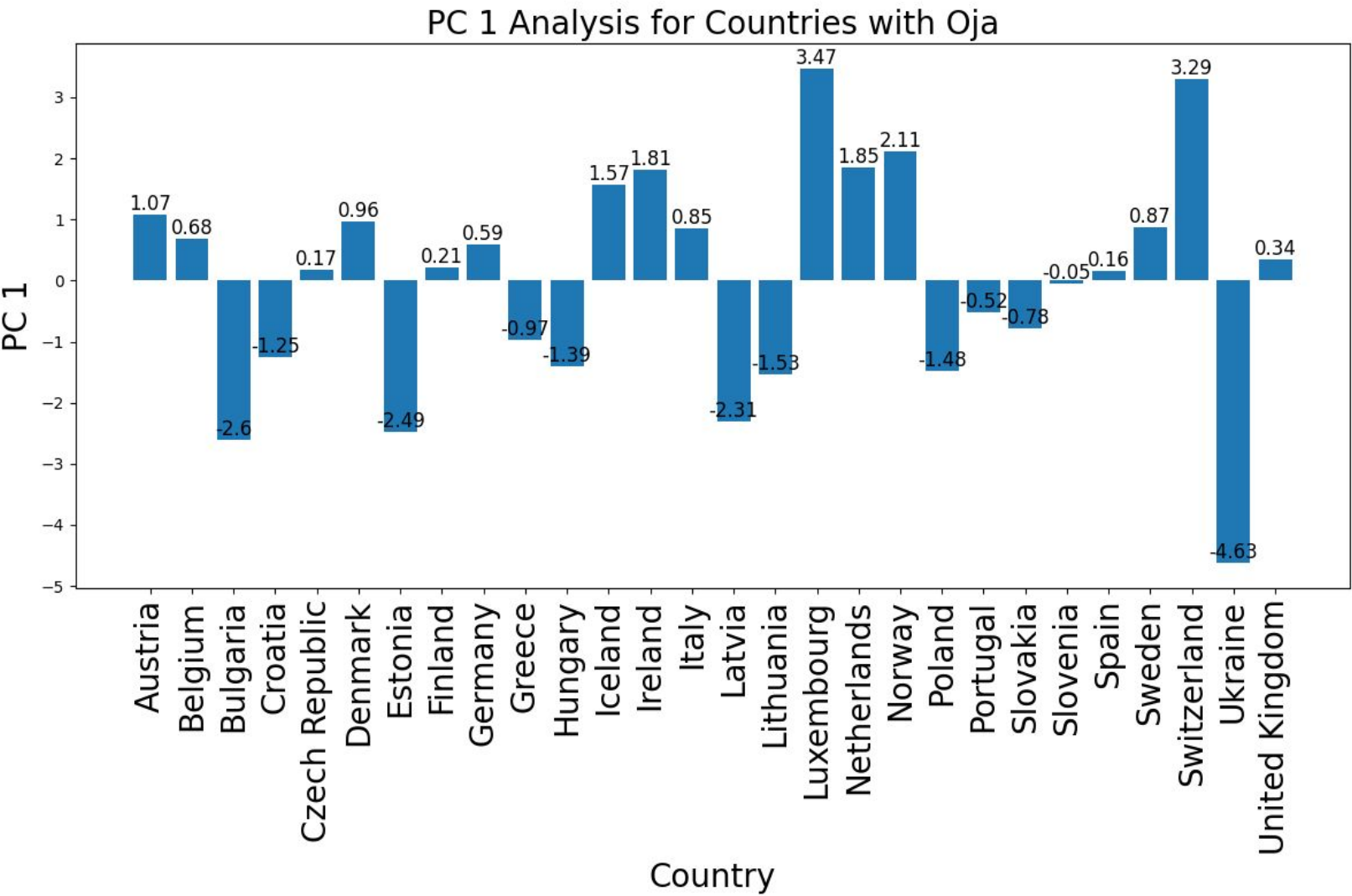
Area	GDP	Inflation	Life Expectancy	Military	Pop Growth	Unemployement
-0.978	0.419	-0.598	1.039	0.528	-0.009	0.985

10k épocas y $\eta = 0.01$



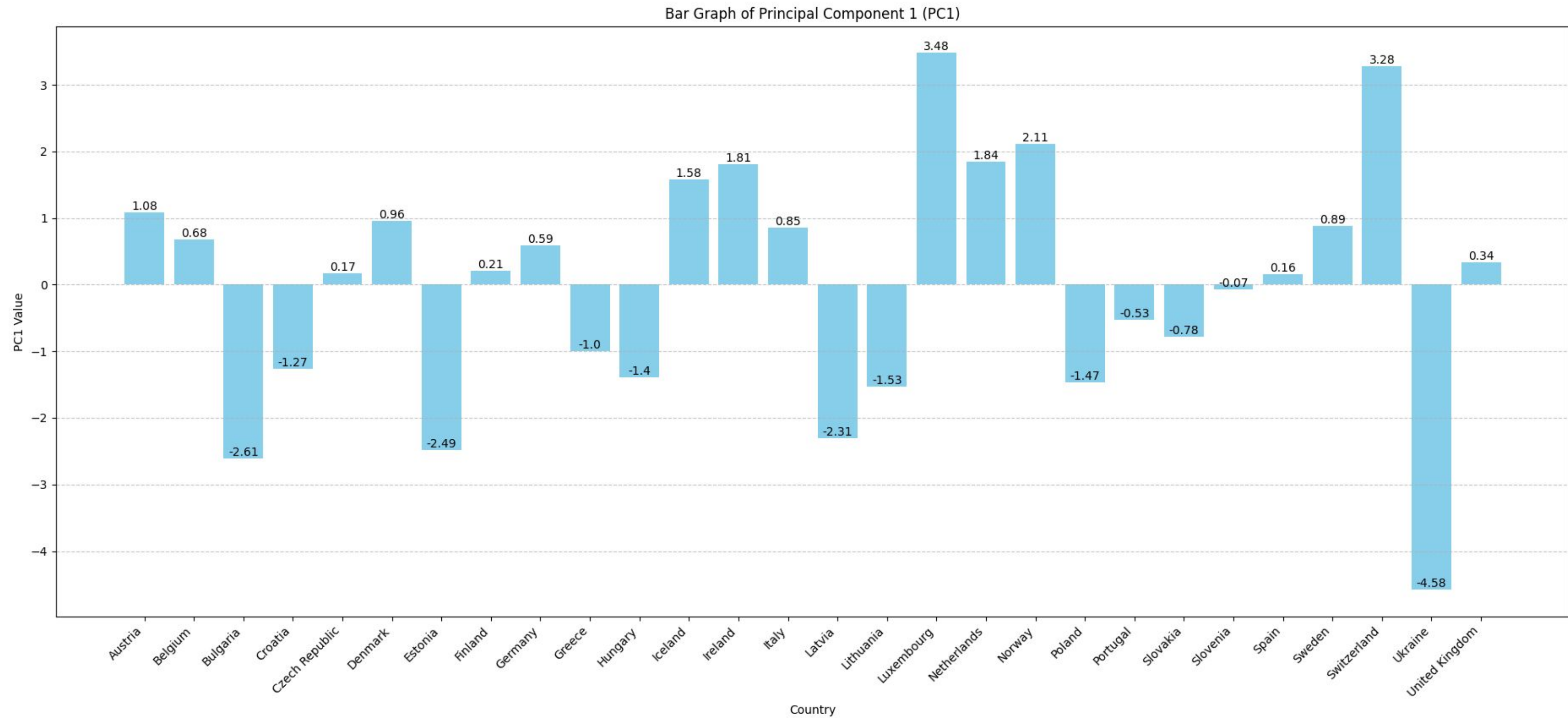
Area	GDP	Inflation	Life Expectancy	Military	Pop Growth	Unemployme nt
-0.203	0.500	-0.480	0.500	-0.131	0.457	-0.232

10k épocas y $\eta = 0.001$

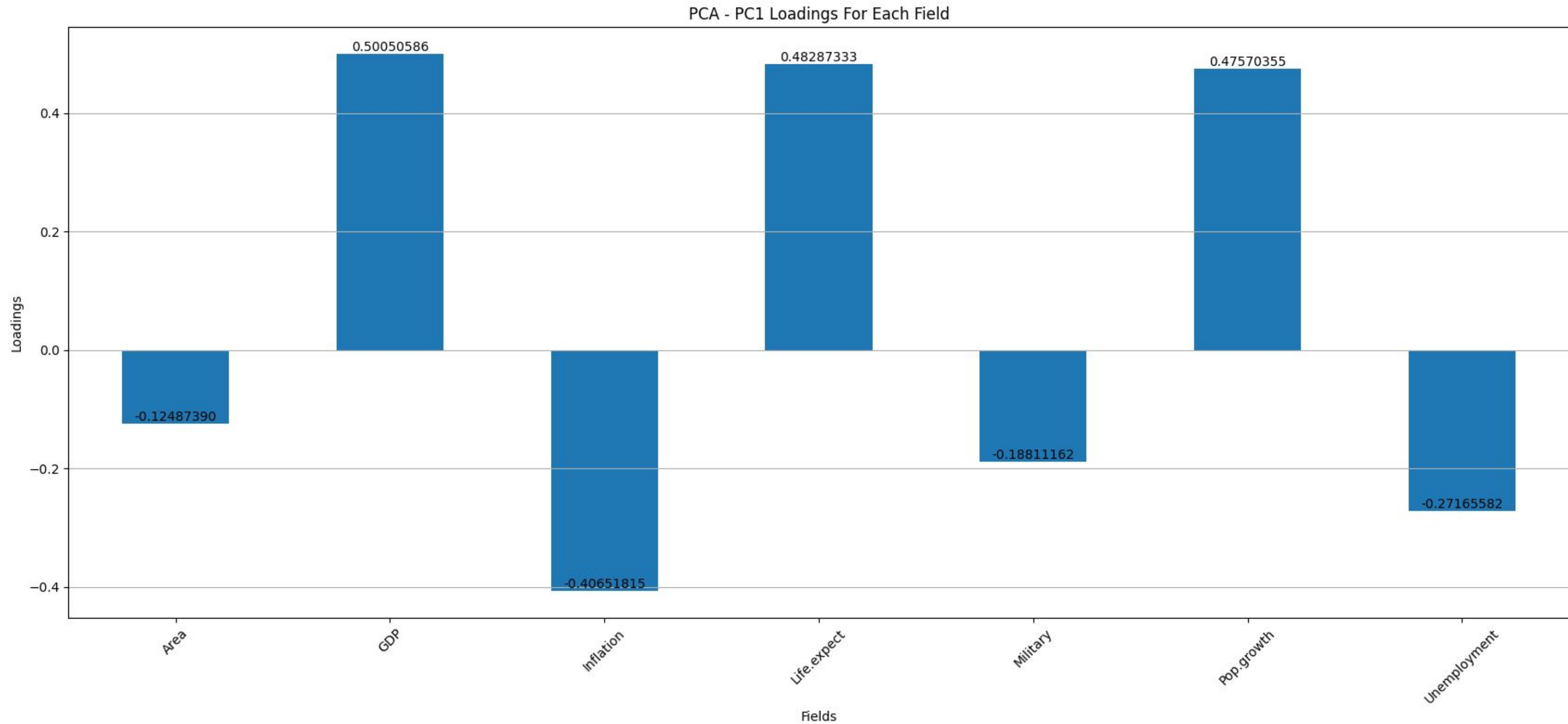


Area	GDP	Inflation	Life Expectancy	Military	Pop Growth	Unemployme nt
-0.132	0.500	-0.414	0.484	-0.182	0.474	-0.268

PCA (sklearn)



PCA-PC1 Loadings (sklearn)



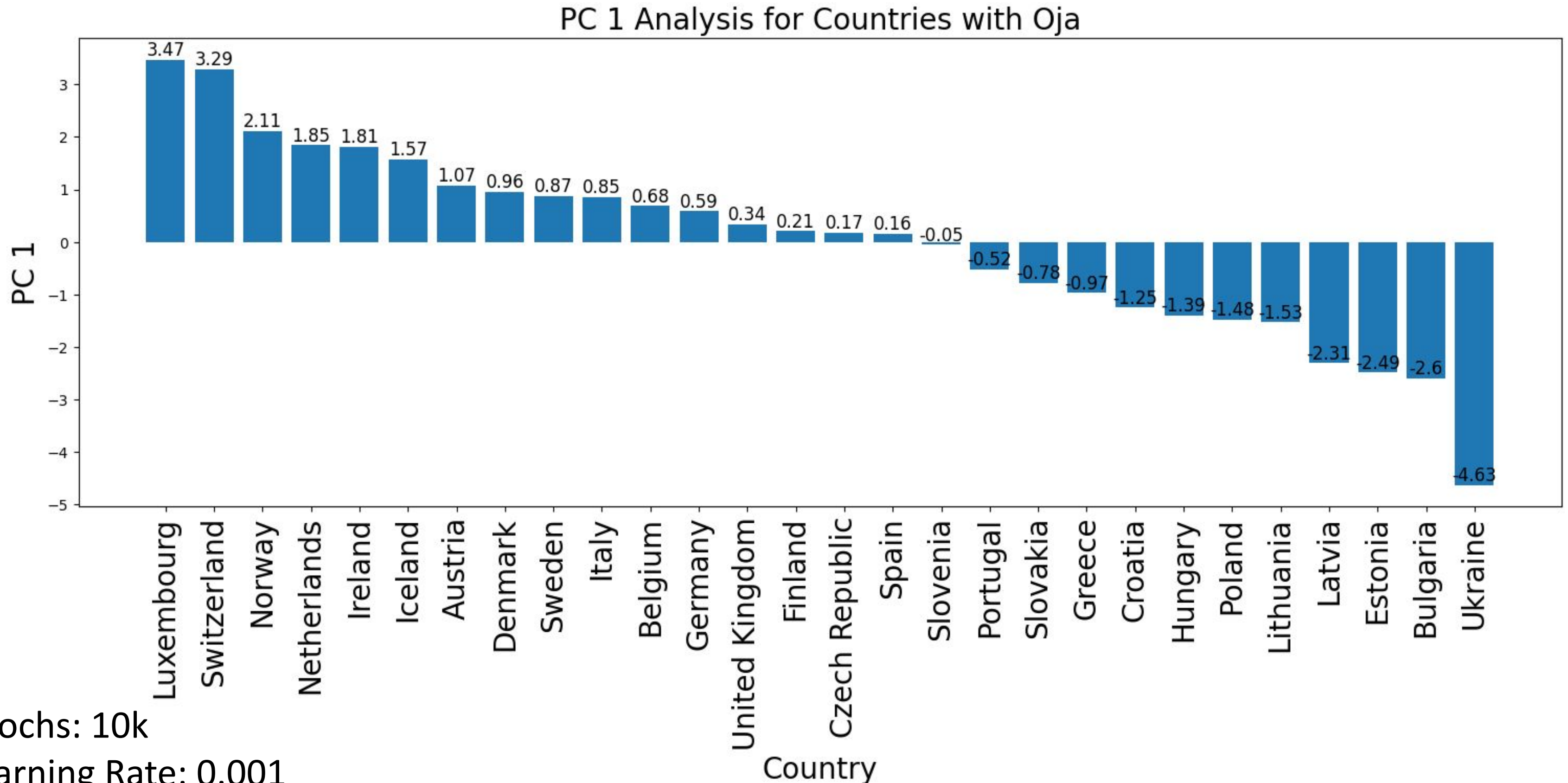
Comparacion OJA - PCA

	Area	GDP	Inflation	Life Expectancy	Military	Pop Growth	Unemployment	MSE vs PCA
PCA	-0.125	0.501	-0.407	0.483	-0.188	0.476	-0.272	-
Oja (0.1)	-0.978	0.419	-0.598	1.039	0.528	-0.009	0.985	0.487
Oja (0.01)	-0.203	0.500	-0.480	0.500	-0.131	0.457	-0.232	2.435e-03
Oja (0.001)	-0.132	0.500	-0.414	0.484	-0.182	0.474	-0.268	2.214e-05

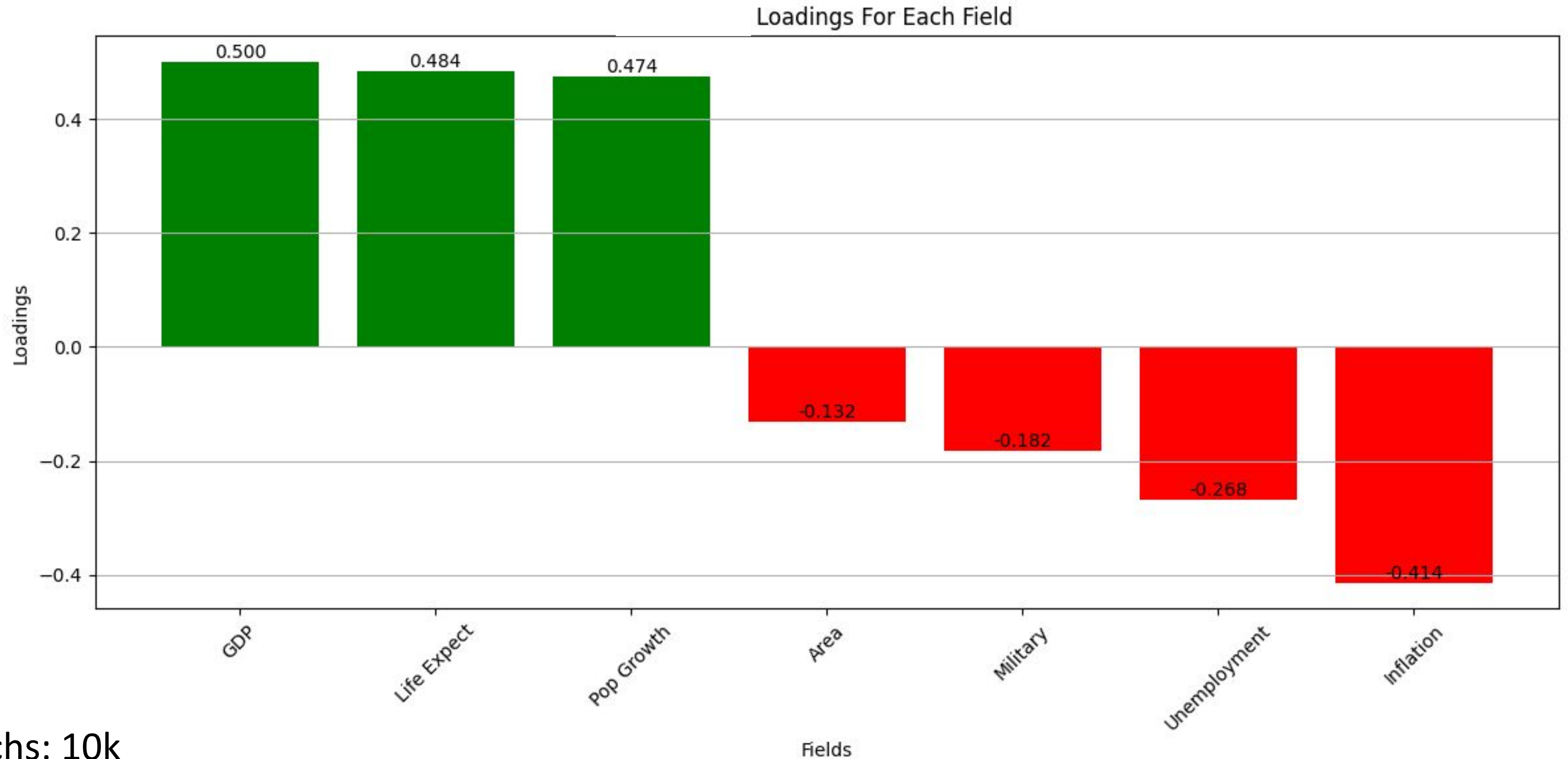
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

n : cantidad de columnas

Countries by PC1 with Oja



Fields by PC1 with Oja



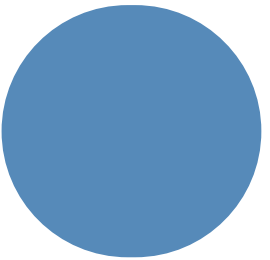
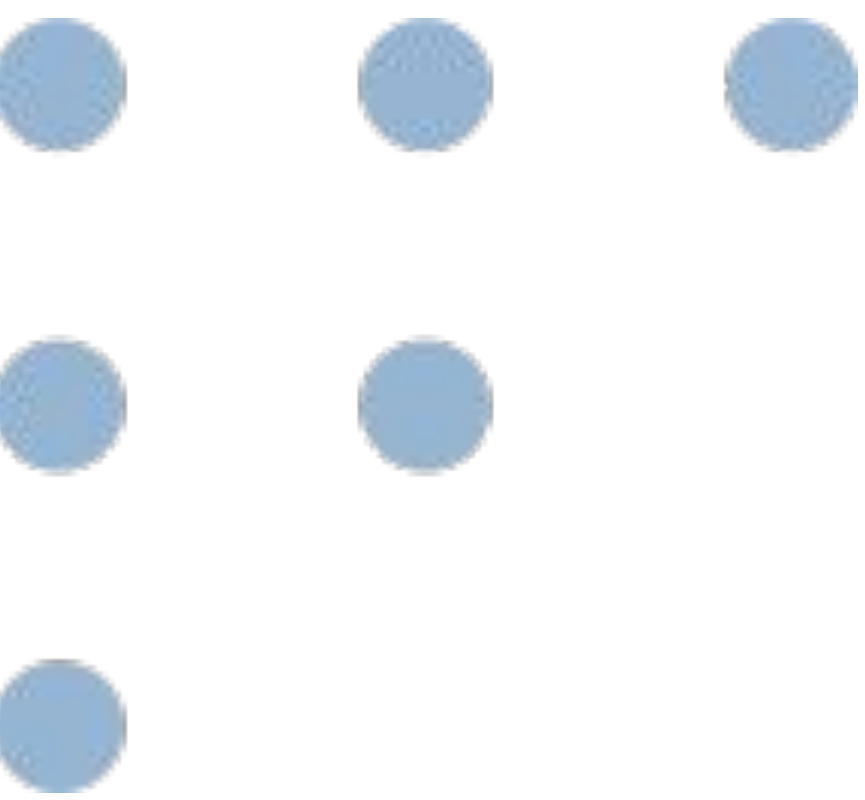
Epochs: 10k

Learning Rate: 0.001

Conclusiones

- Se puede inferir que el learning rate utilizado debe ser bajo así se logra la convergencia a la PC1.
- Se interpreta PC1 como un indicador del desarrollo socioeconómico de los países.
- Con Oja se puede hacer una buena aproximación del primer componente principal, con error muy bajo en comparación a PCA (sklearn)
- Como el dataset no posee muchas variables, el Oja funciona bien ya que no se pierde mucha información al reducir la dimensión.

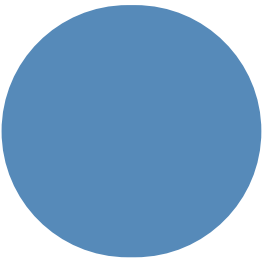




SIA 2024 - 2C

EJERCICIO 2





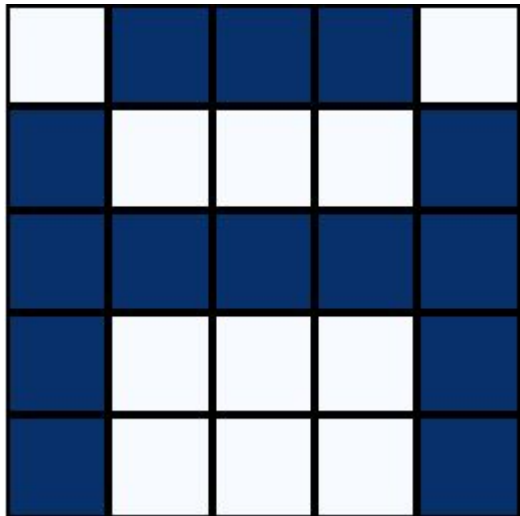
SIA 2024 - 2C

Modelo de Hopfield

Grupo 4



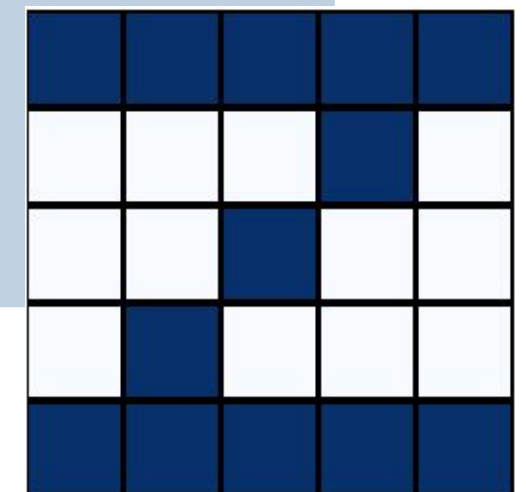
Problema a resolver



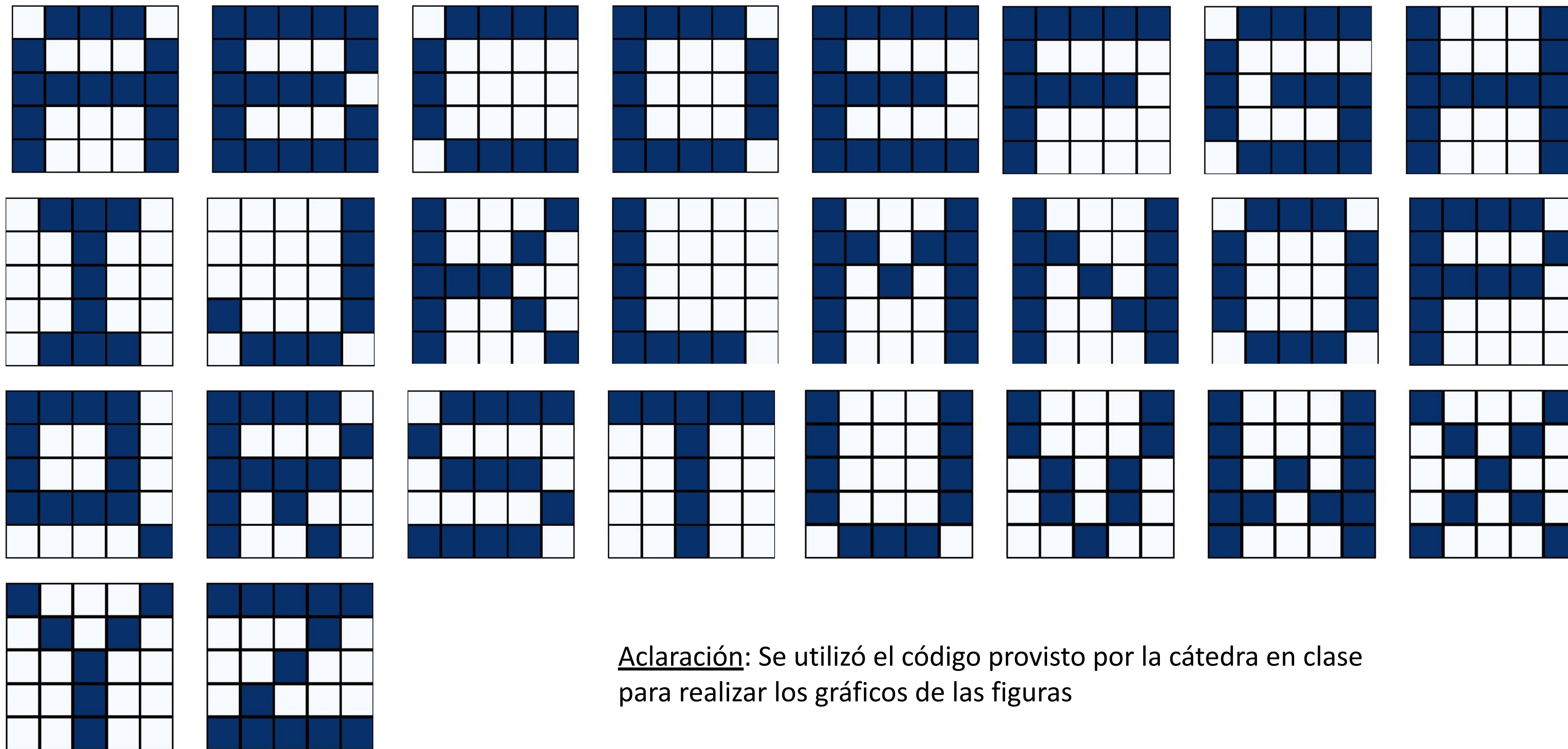
Se tiene un diccionario de letras cuyas figuras se representan únicamente con 1 y -1 en matrices de 5x5

Asociar matrices ruidosas de las letras al patrón original de dicha letra a través del modelo de Hopfield

Solo se entrenará el modelo con grupos de 4 matrices para no superar la capacidad del modelo



Diccionario



Aclaración: Se utilizó el código provisto por la cátedra en clase para realizar los gráficos de las figuras

Preguntas disparadoras del análisis

- ¿Cómo consideramos cuáles son el mejor y el peor patrón de nuestro diccionario?
- ¿Cuál es el mejor subconjunto de patrones de nuestro diccionario? ¿Y cuál es el peor?
- ¿Cómo se comporta la función de energía a través de las épocas?
- ¿Cómo responden estos subconjuntos al ruido?



Consideraciones

Para obtener el mejor y el peor subconjunto de patrones tendremos en cuenta el producto interno entre las matrices, cuanto más cercano a 1 sea el resultado, mayor distinción tendrán los patrones entre sí.

Prod. Interno

-1	1	1	1	1
1	-1	-1	-1	-1
1	-1	-1	-1	-1
1	-1	-1	-1	-1
-1	1	1	1	1

,

1	1	1	1	-1
1	-1	-1	-1	1
1	-1	-1	-1	1
1	-1	-1	-1	1
1	1	1	1	-1

$= (-1)*1 + 1*1 + \dots + 1*(-1) = 5$

Para cada subconjunto hacemos una matriz con los productos internos entre los distintos pares de matrices. Luego obtenemos el promedio de los productos internos y para no perder información recopilaremos el mayor producto interno obtenido entre los pares junto a su frecuencia.

Hacemos una búsqueda exhaustiva entre los $26 C 4 = 14950$ subconjuntos sin ruido

Resultados

Mejores

Combinación	Promedio	Prod. interno máx. (PIM)	Frecuencia P.I.M
(B, Q, T, V)	1	1	6
(L, Q, T, X)	1	1	6
(B, J, T, V)	1.33	3	1
(G, L, Q, T)	1.33	3	1
(L, Q, T, V)	1.33	3	1
(O, Q, W, Z)	1.33	3	1
(B, L, Q, T)	1.66	5	1
(B, N, Q, V)	1.66	5	2

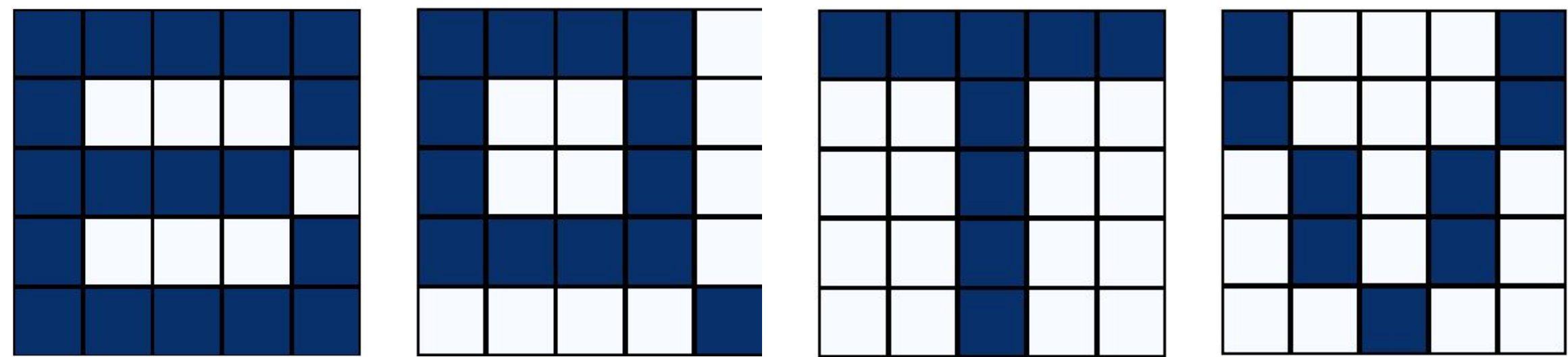
Peores

Combinación	Promedio	Prod. interno máx. (PIM)	Frecuencia P.I.M
(H, M, N, W)	18.33	21	2
(I, M, N, W)	17.33	21	2
(A, F, P, R)	17	21	2
(E, F, P, R)	17	21	2
(H, I, M, N)	16.66	21	1
(H, I, N, W)	16.66	21	1
(B, E, F, P)	16.33	21	2
(B, F, P, R)	16.33	21	2

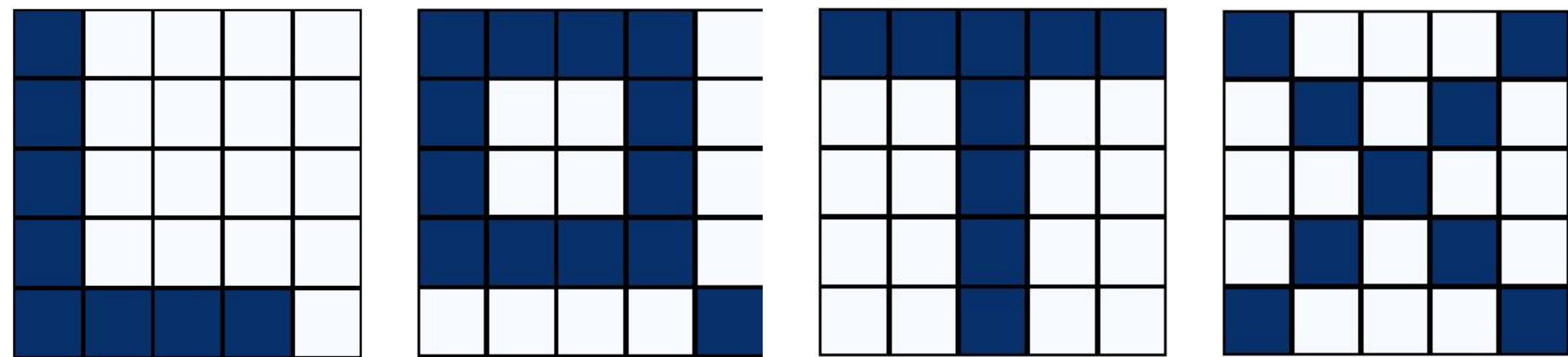
Mejores subconjunto

En este caso llegamos a un empate técnico ya que hay 2 subconjuntos que dieron el mejor caso posible donde el producto interno entre todas las matrices del conjunto dan 1 y por ende el promedio da 1

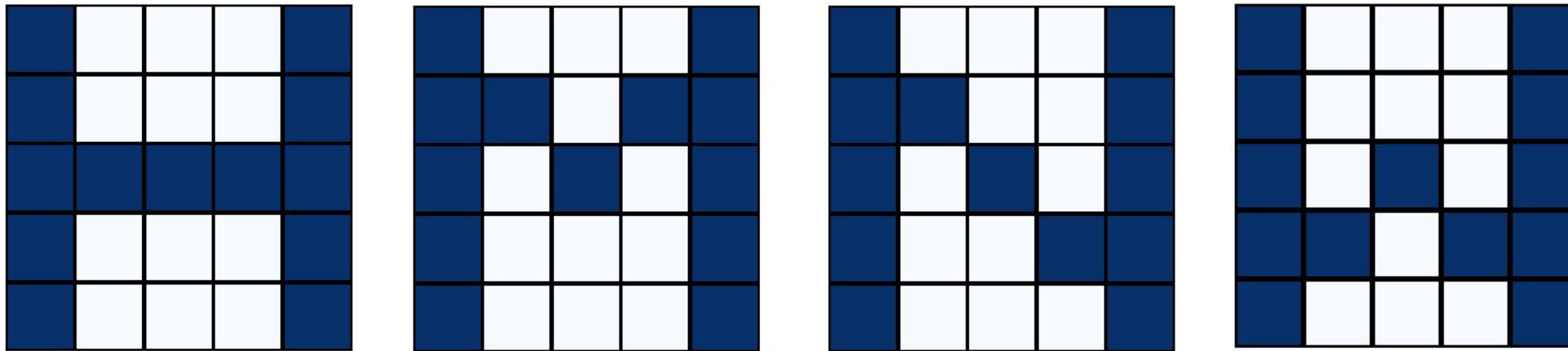
Subconjunto 1



Subconjunto 2



Peor subconjunto



Producto interno promedio: 18.33

Máximo producto interno: 21

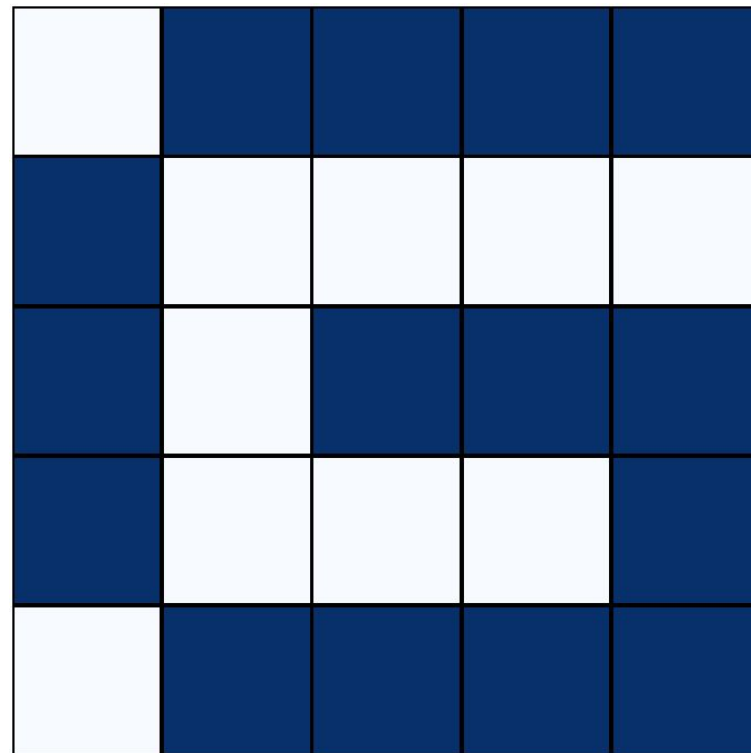
Frecuencia del máximo producto interno entre los pares: 2

Función de energía

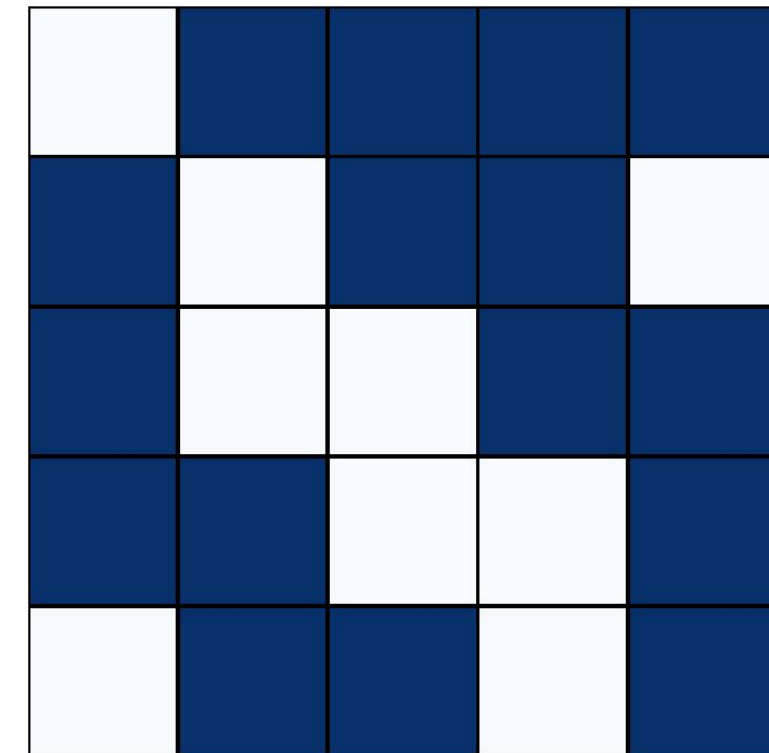
- Sirve para medir el estado de estabilidad de una red
- Cada configuración de neuronas tiene un valor de energía asociado
- El modelo de Hopfield minimiza esta función durante el funcionamiento
- Los mínimos locales de esta función corresponden a estados estables de la red o atractores

Ruido

Nuestra función de ruido toma un patrón e invierte “píxeles” aleatoriamente. Esta función modifica el porcentaje de “píxeles” que le pasemos como argumento. Por ejemplo, si aplicamos ruido 0.2, se modificarán $25 * 0.2 = 5$ posiciones del patrón.

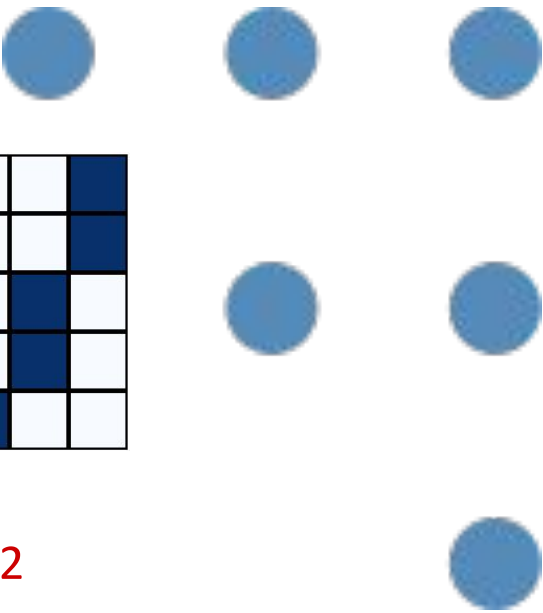
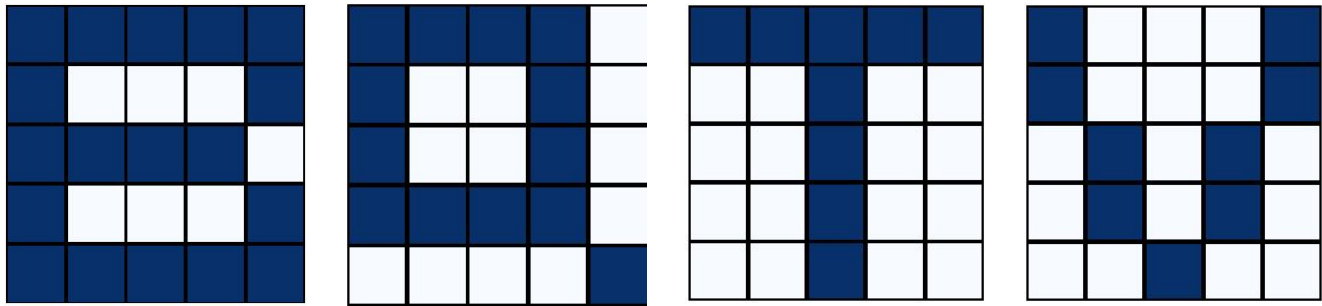


Letra G sin ruido



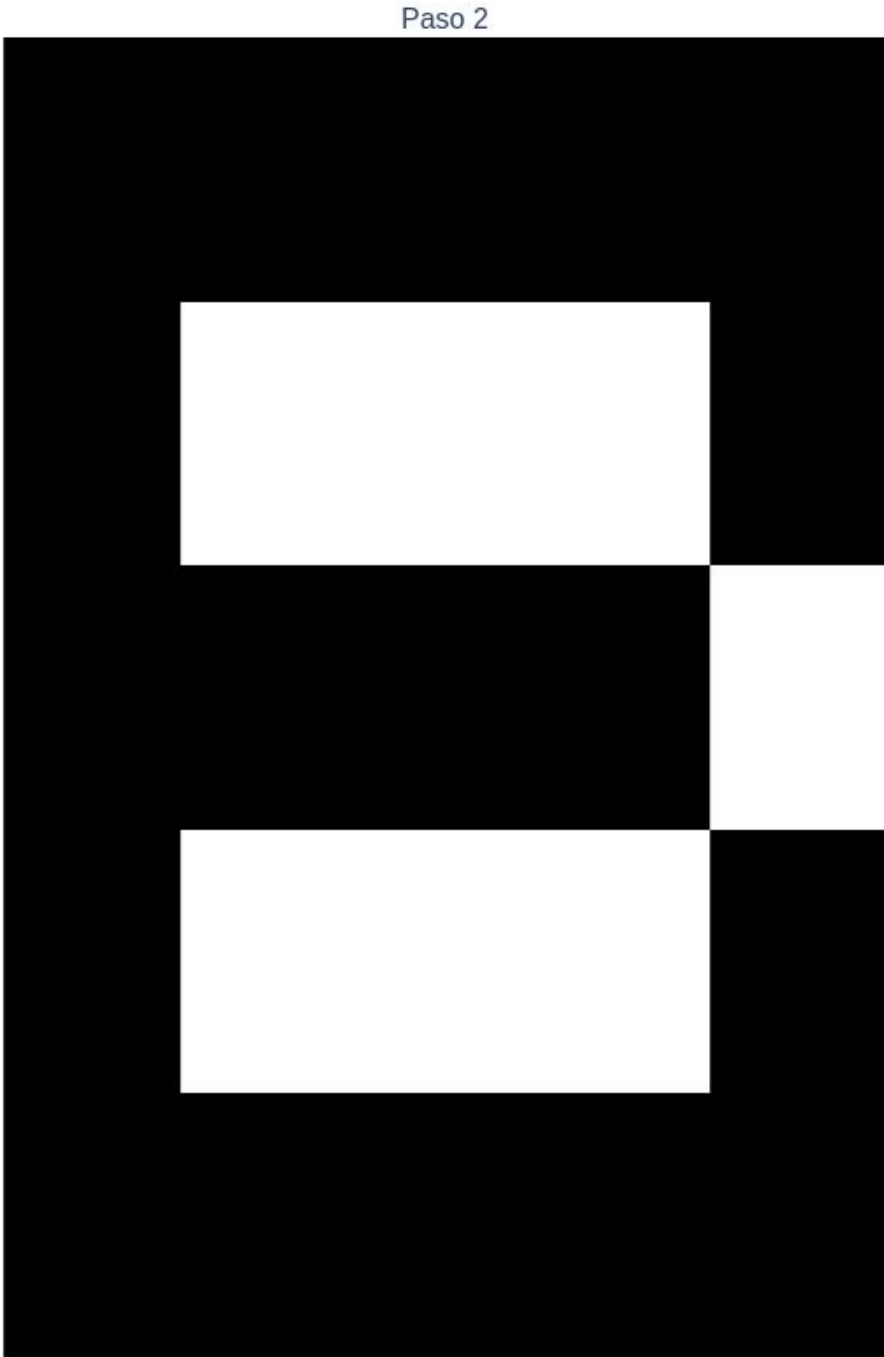
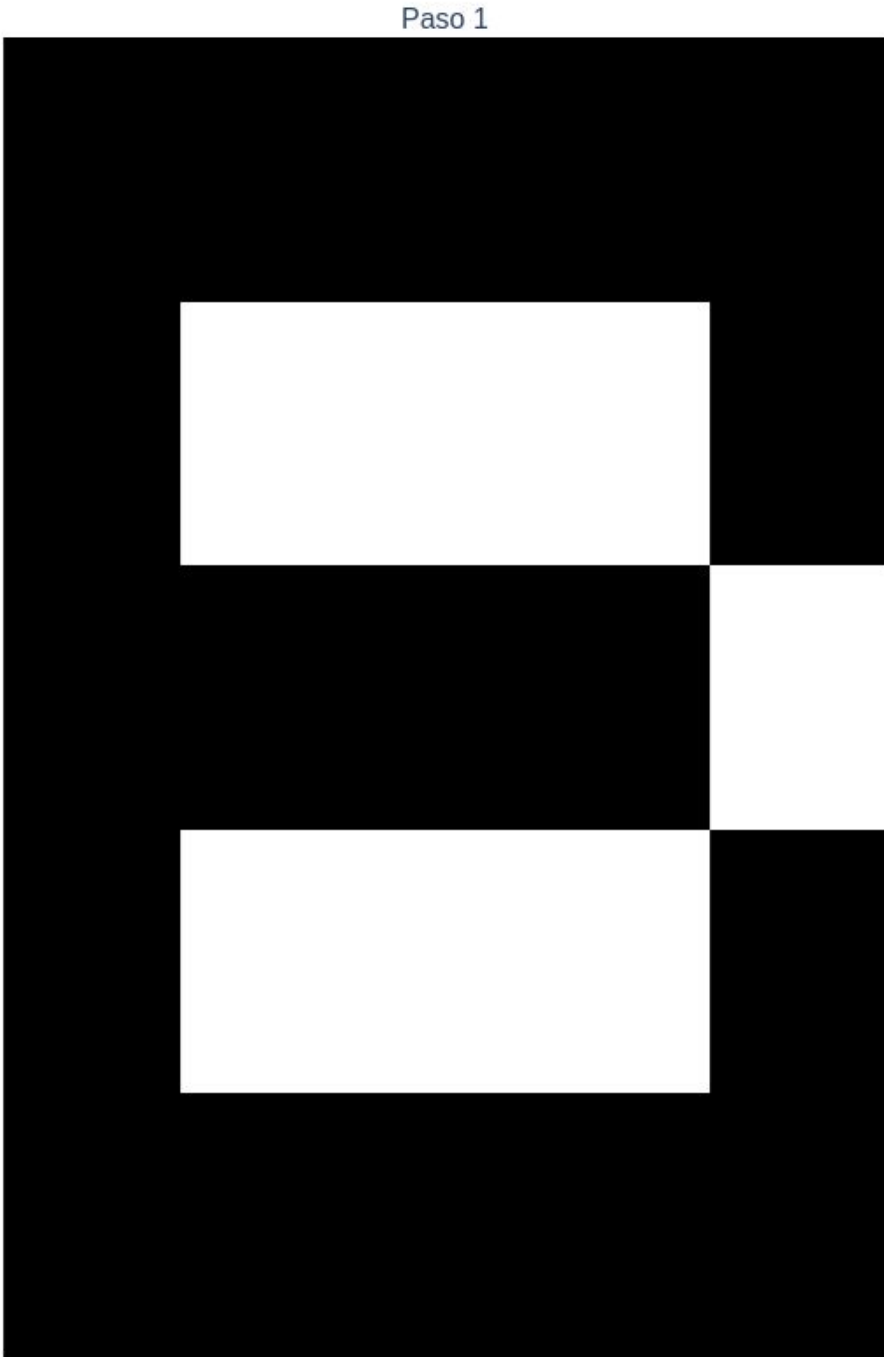
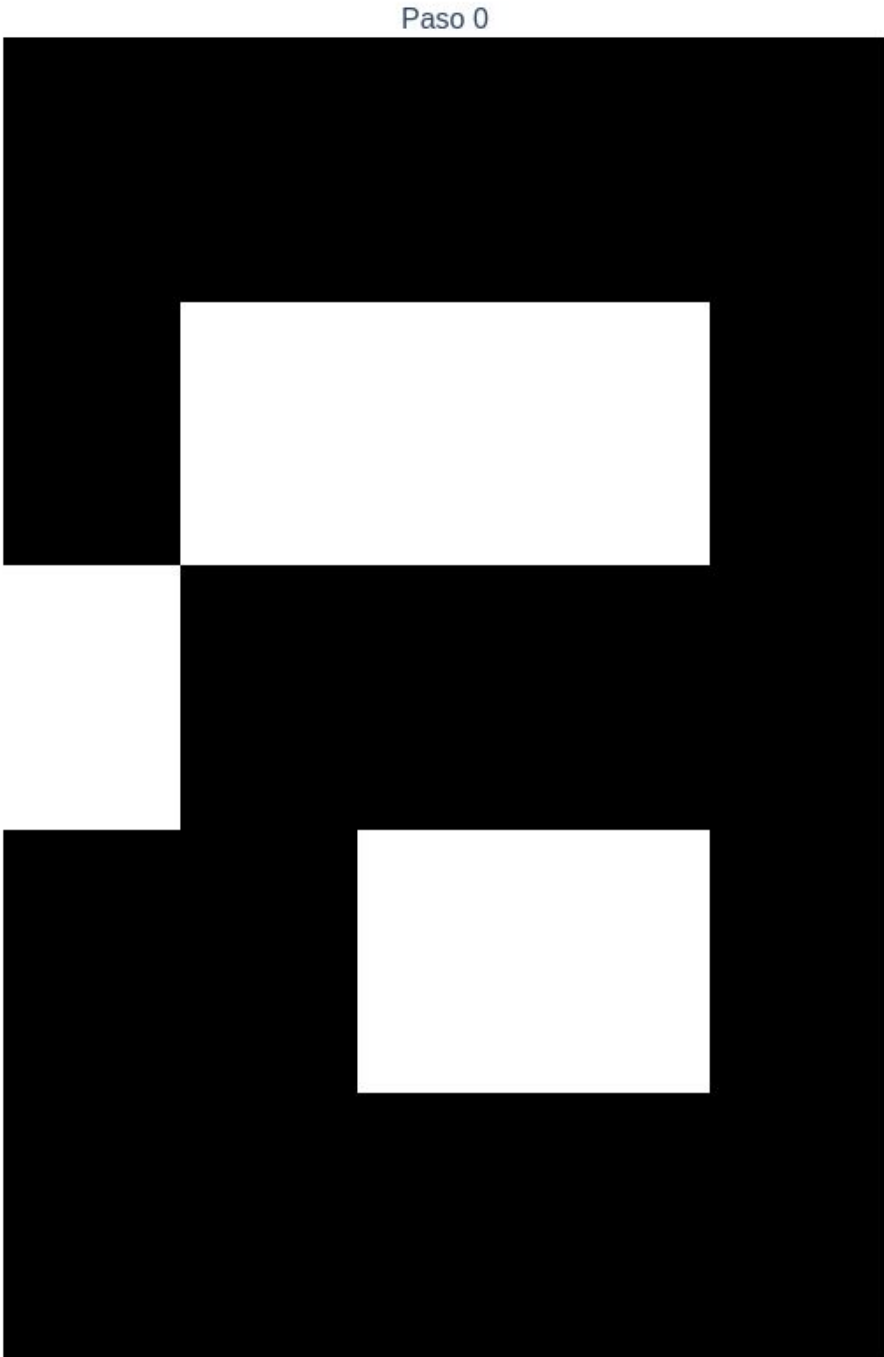
Letra G con ruido 0.2

Evaluación del mejor subconjunto con ruido

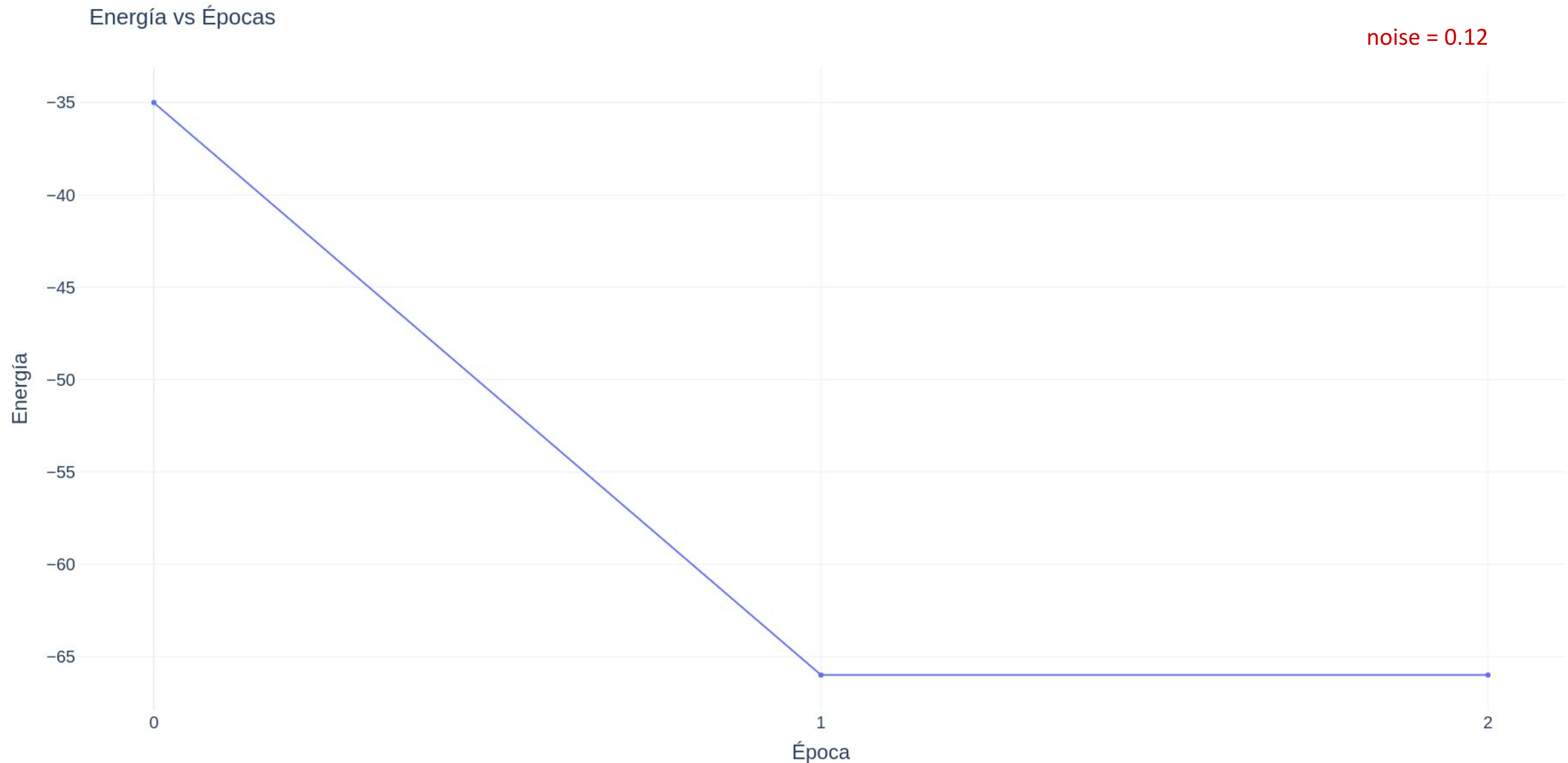
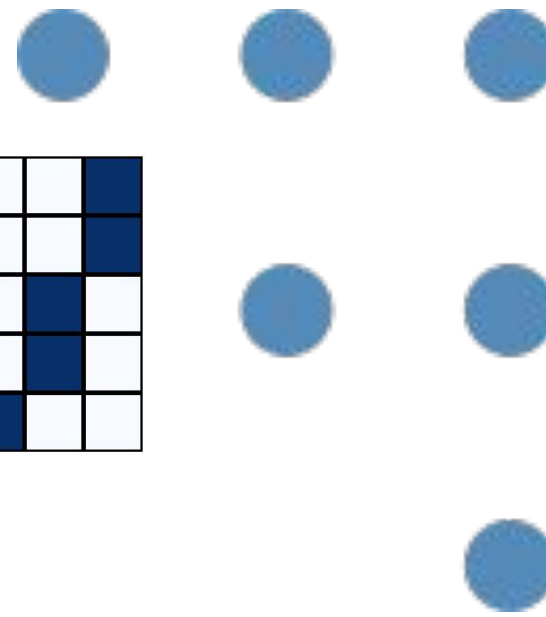
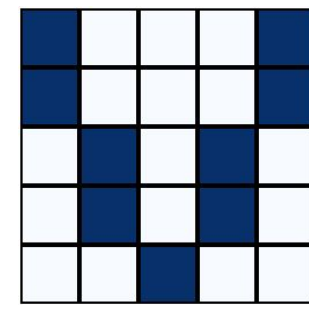
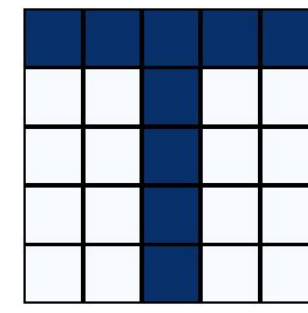
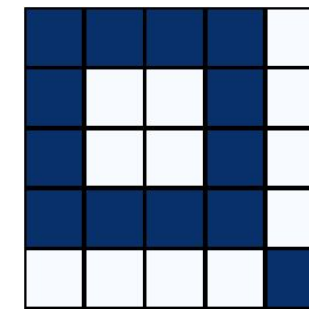
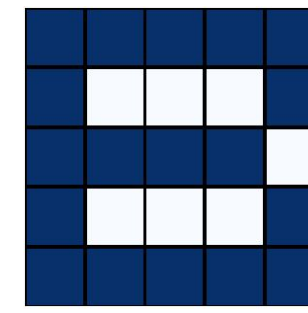


Evolución de Estados

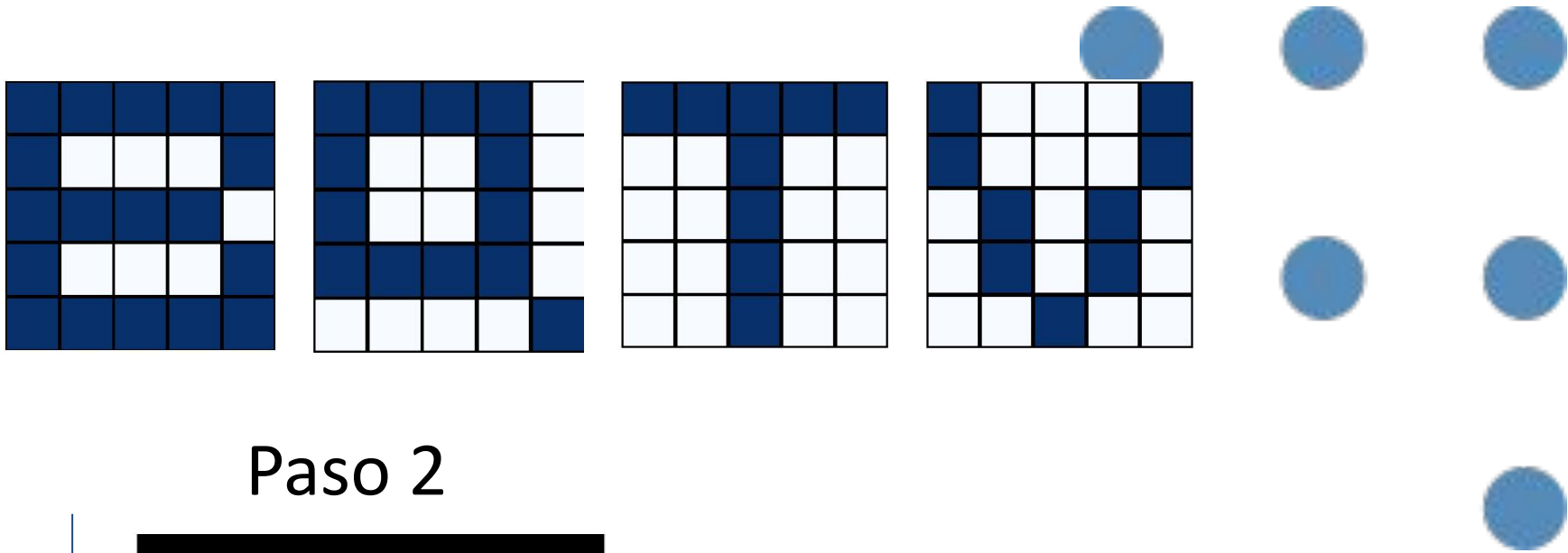
noise = 0.12



Evaluación del mejor subconjunto con ruido



Evaluación del mejor subconjunto con ruido



Letra

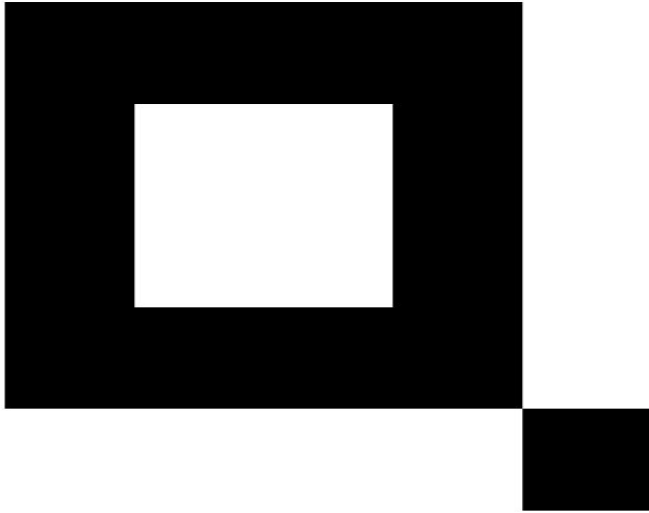
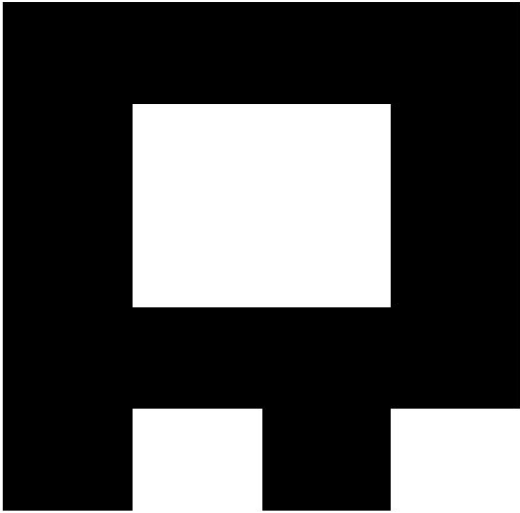
Paso 0

Paso 1

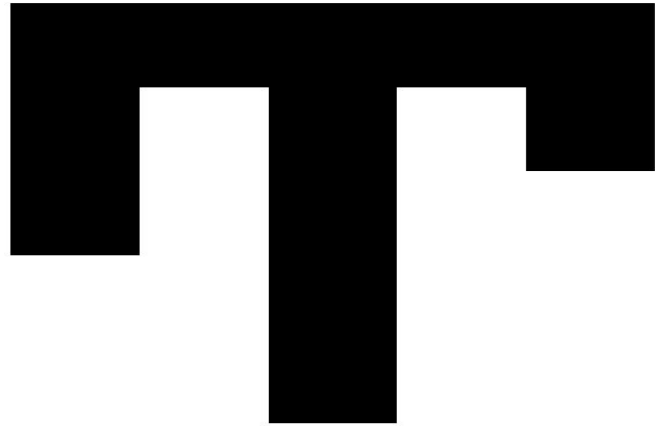
Paso 2

noise = 0.12

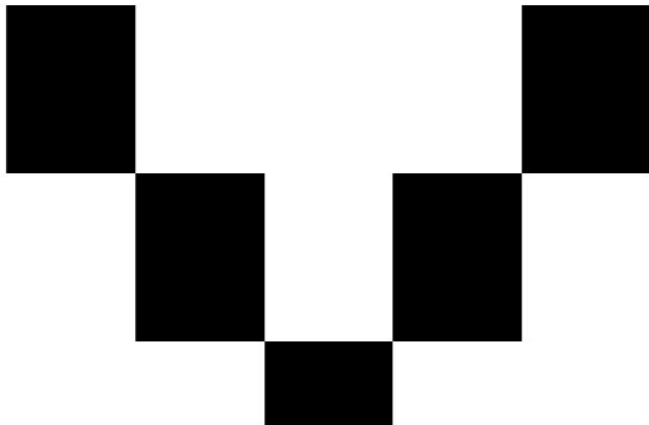
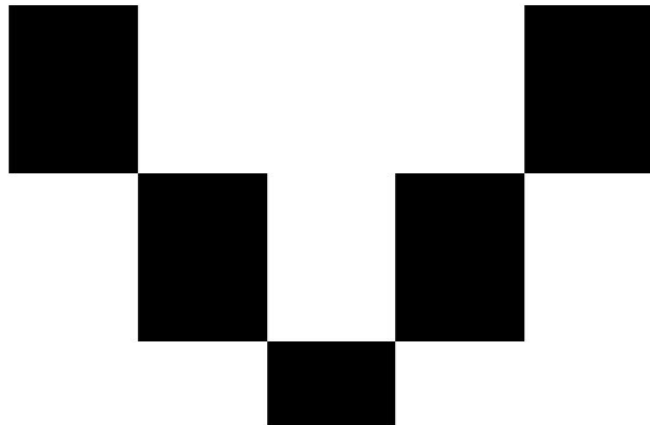
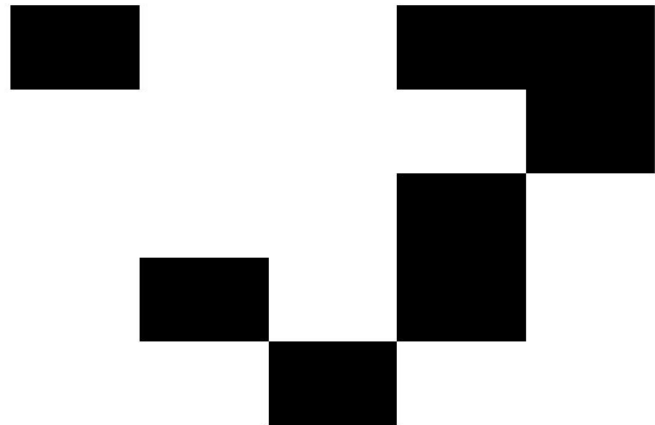
Q



T

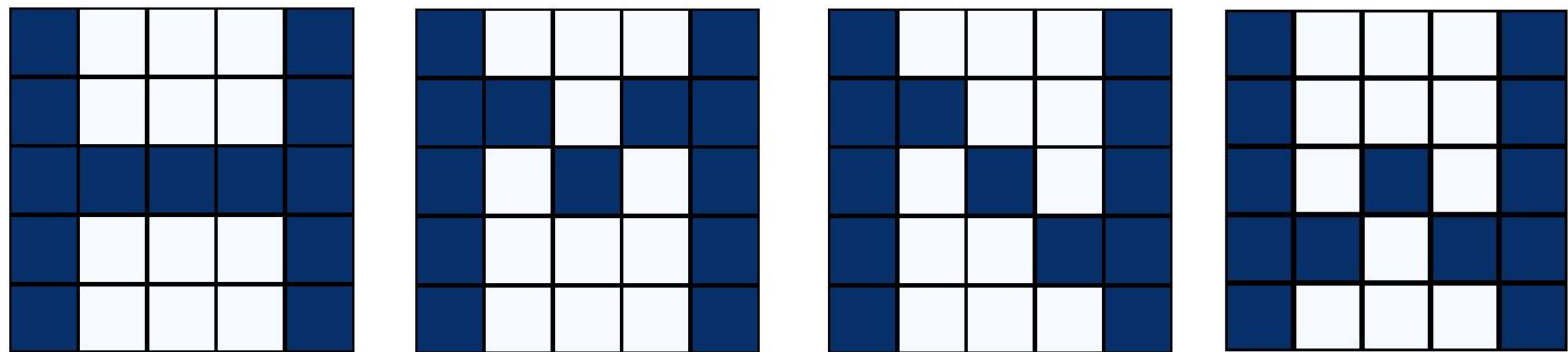


V



Evaluación del peor subconjunto con ruido

El input fue una H con ruido pero converge a una N!



noise = 0.12 (invierte 3 posiciones del patrón)

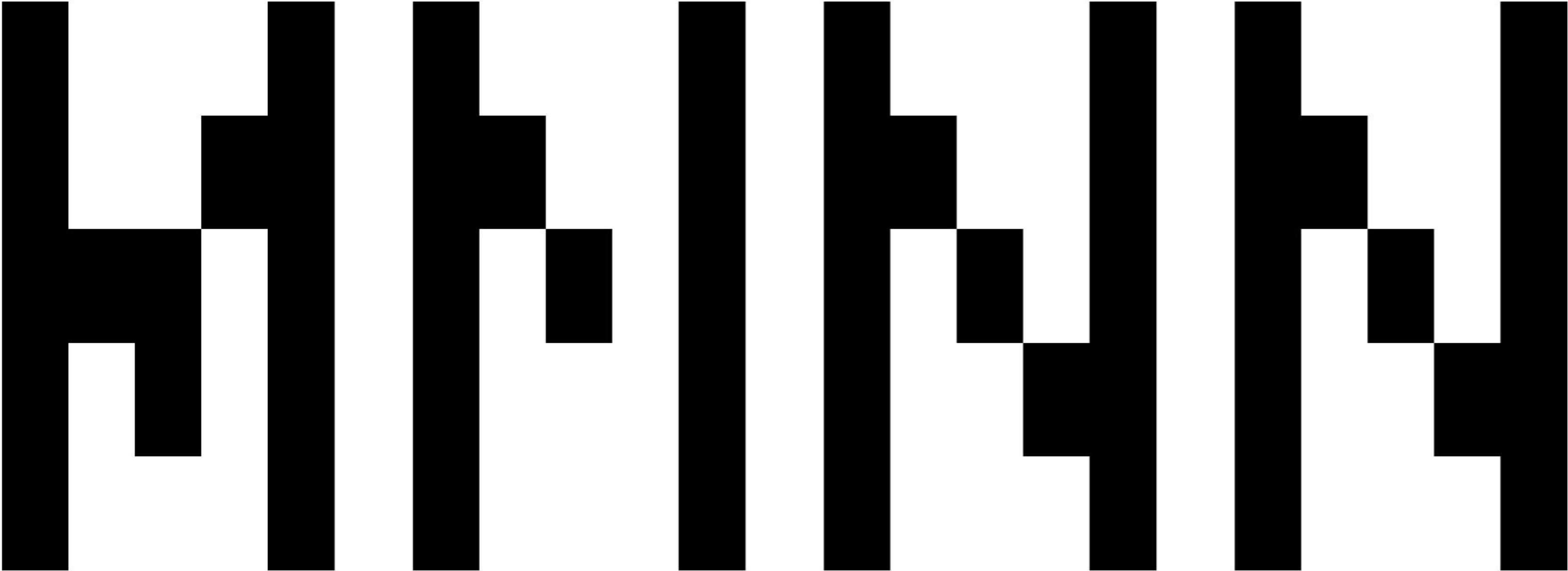
Evolución de Estados

Paso 0

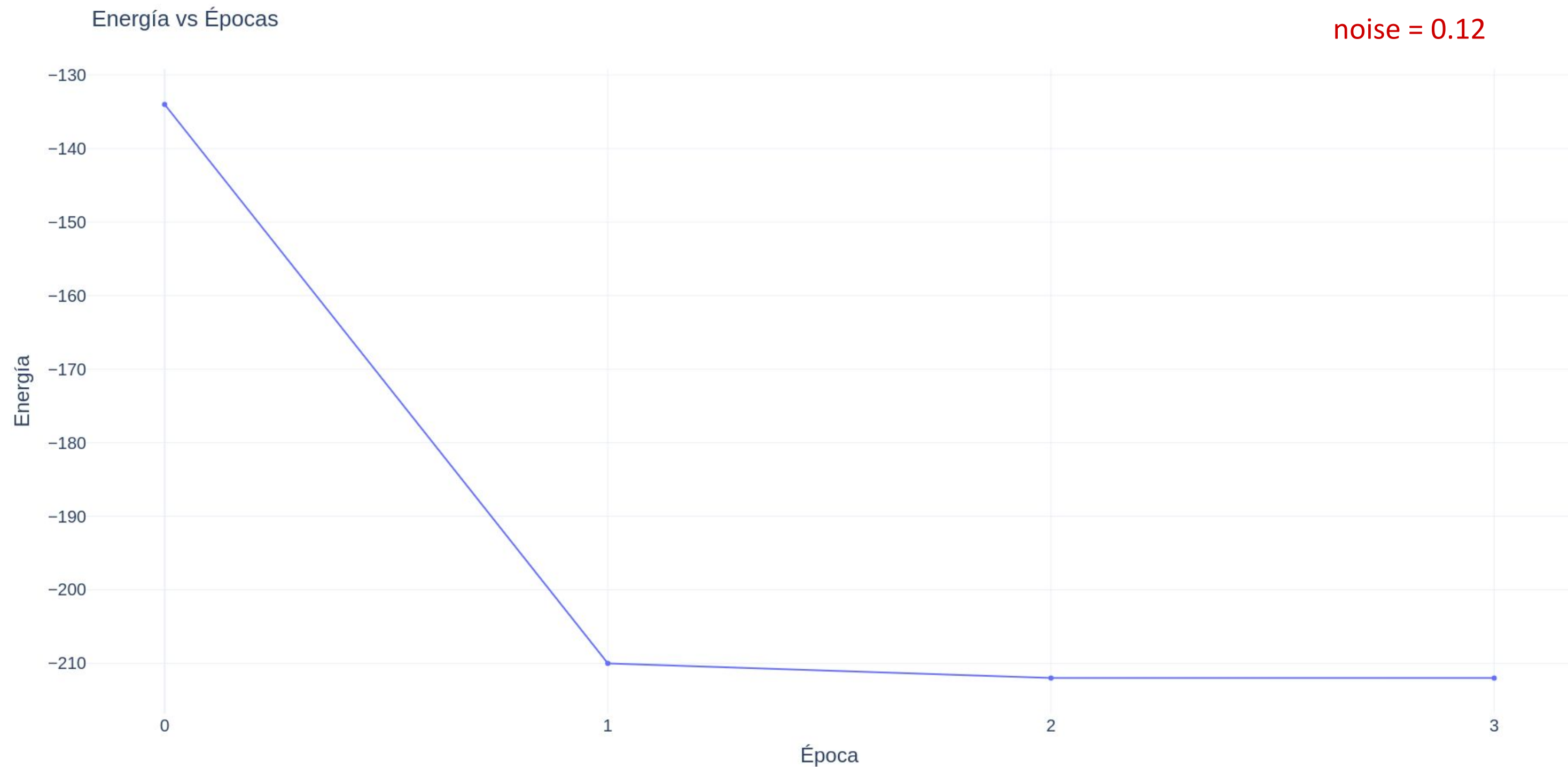
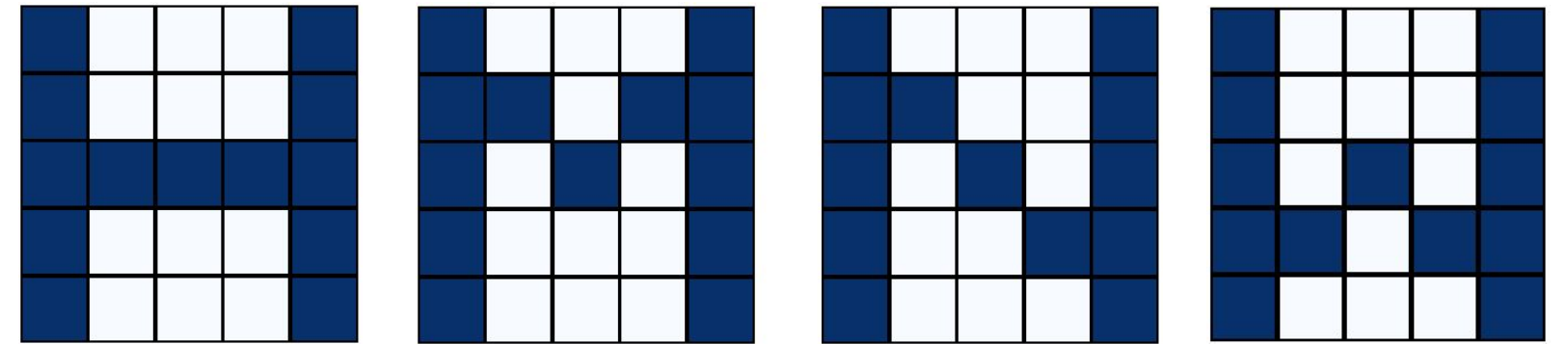
Paso 1

Paso 2

Paso 3



Evaluación del peor subconjunto con ruido



Evaluación del peor subconjunto con ruido



Letra

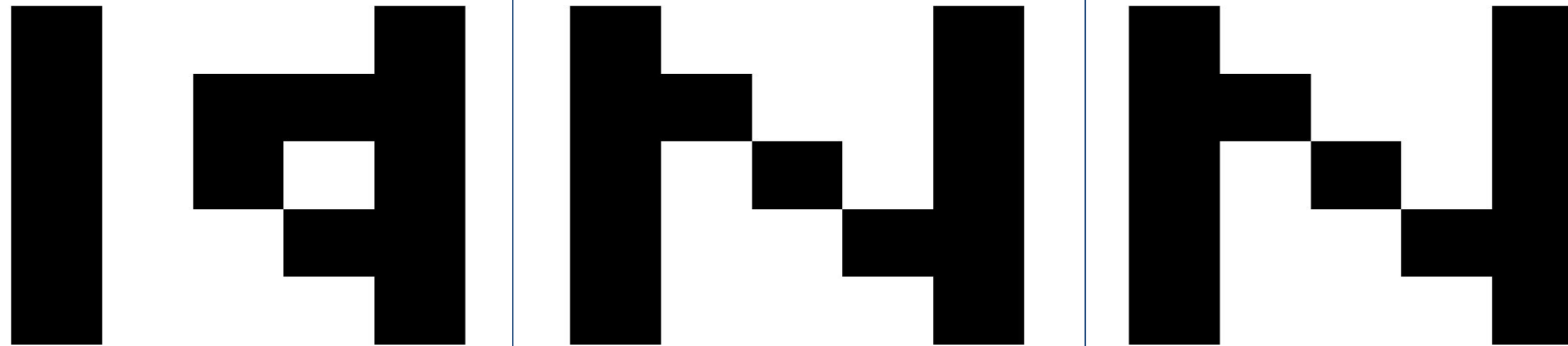
Paso 0

Paso 1

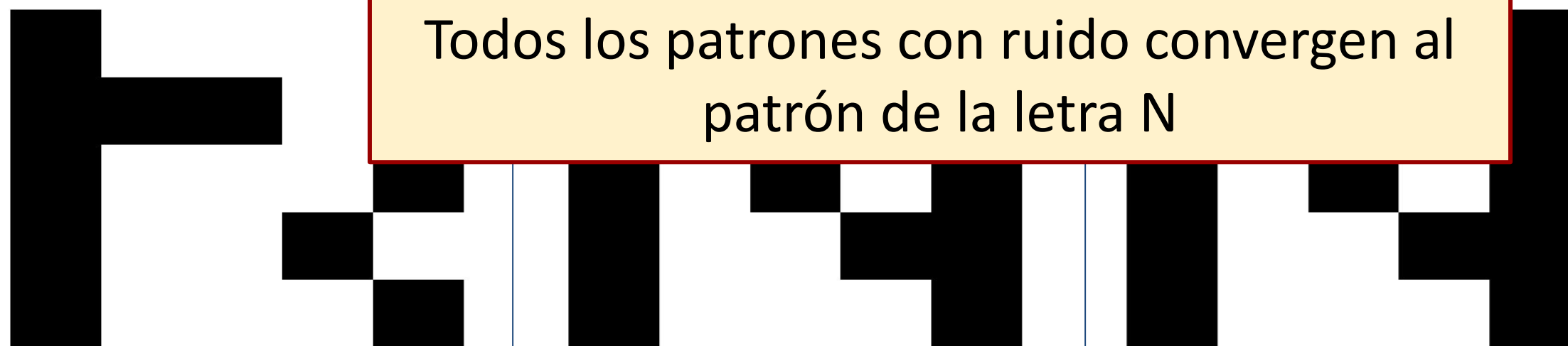
Paso 2

noise = 0.12

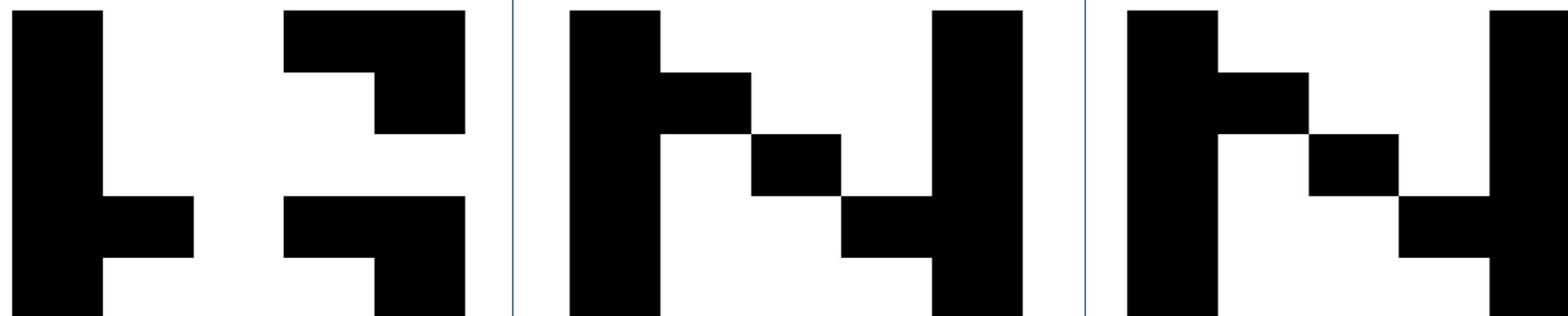
M



N

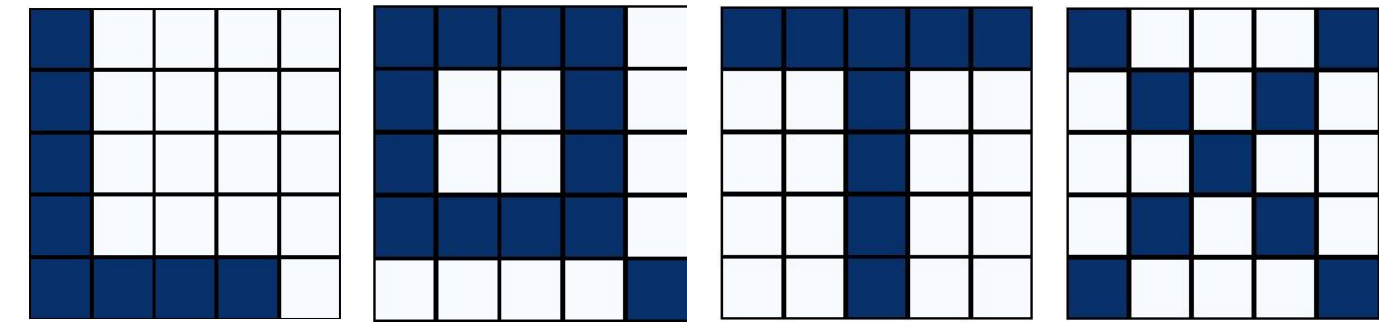


W



Todos los patrones con ruido convergen al patrón de la letra N

¿Qué ocurre si a uno de los mejores subconjuntos de patrones le aplicamos mucho ruido?



Fuimos aumentando el ruido hasta 0.28 donde ocurrió lo siguiente...

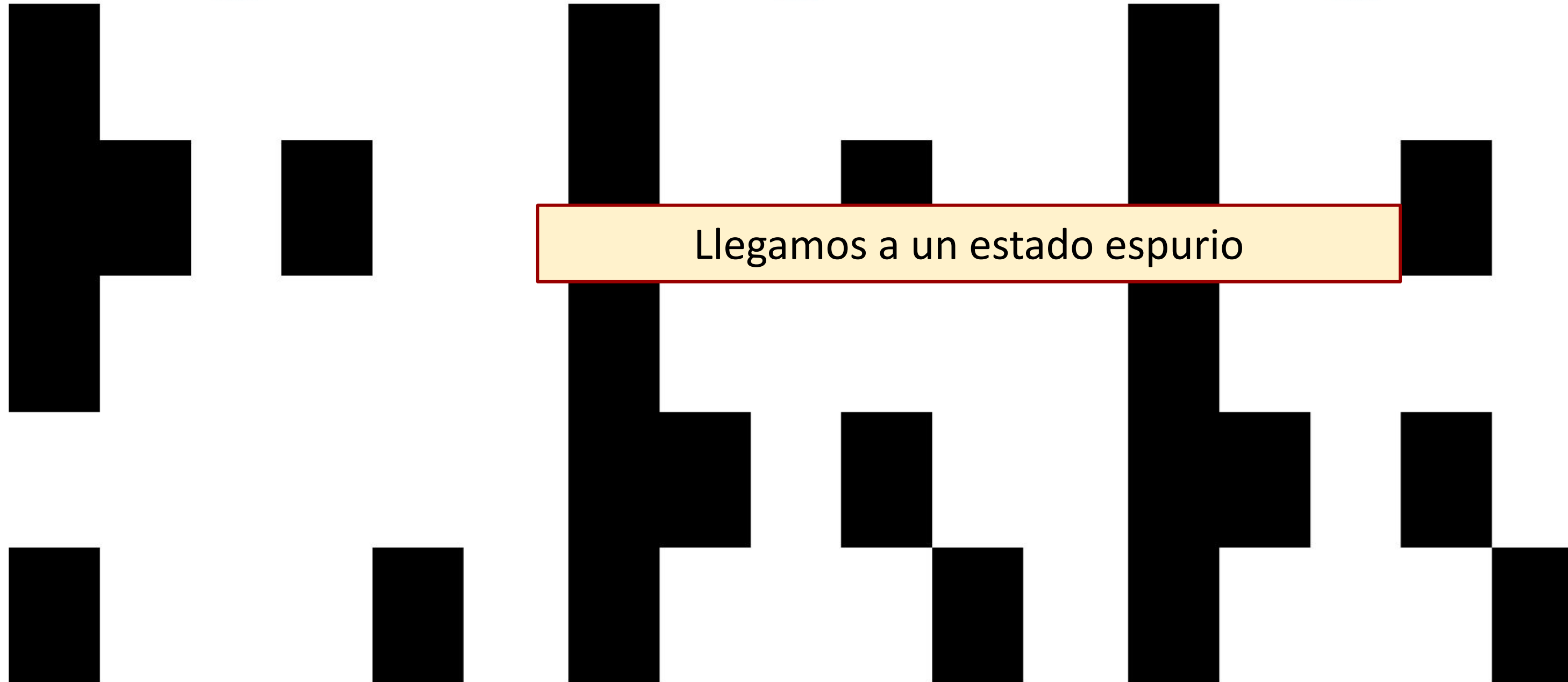
Producto interno promedio: 1

Evolución de Estados

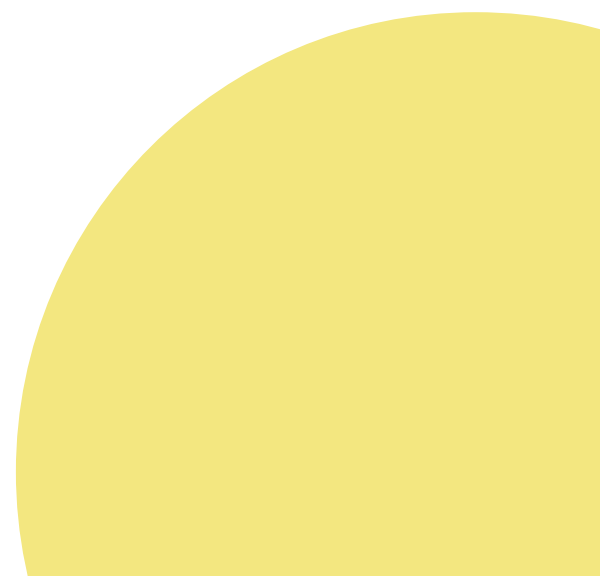
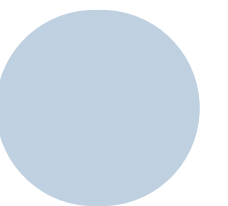
Paso 0

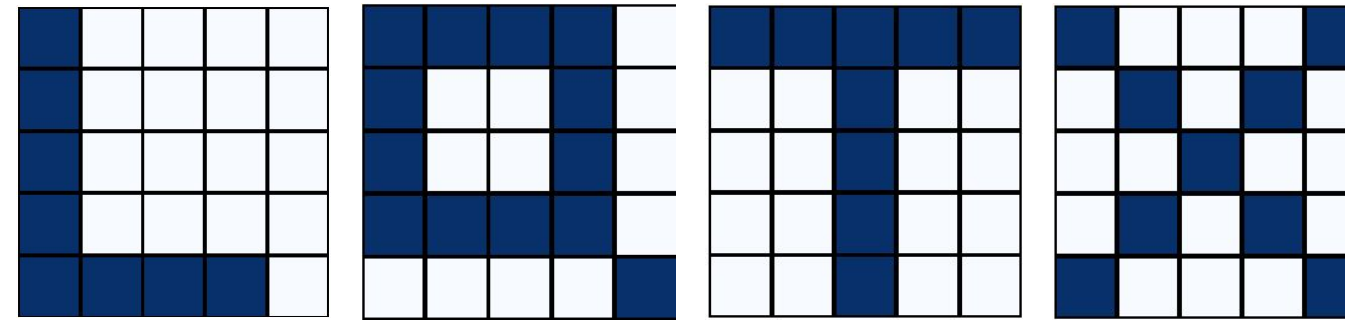
Paso 1

Paso 2



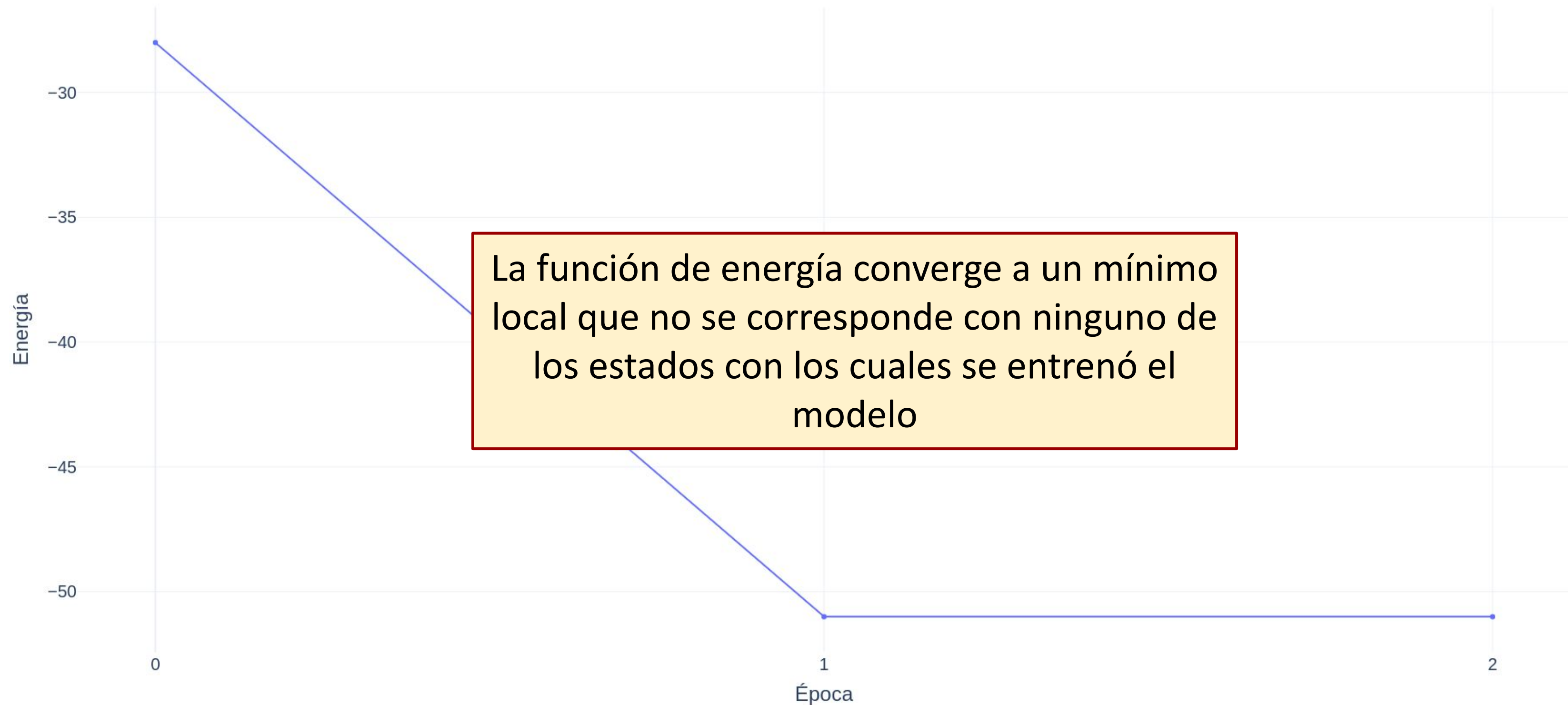
noise = 0.28 (se invierten
7 posiciones en el patrón)



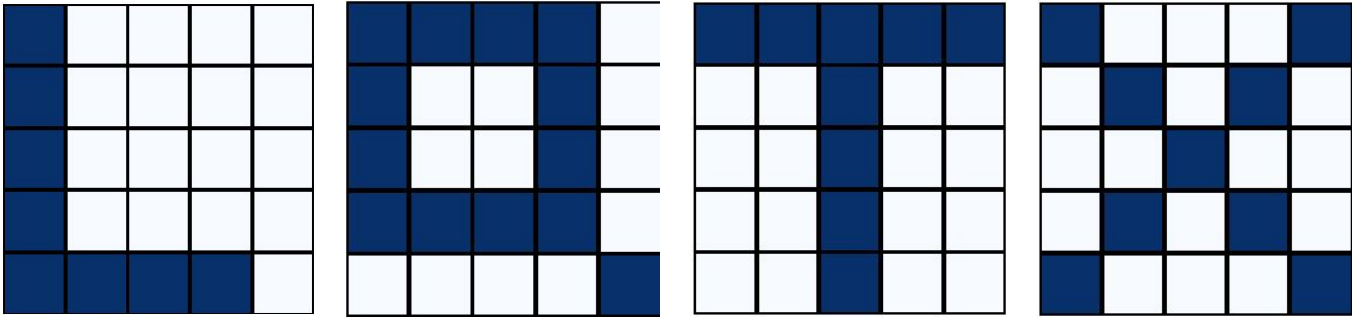


Producto interno promedio: 1

Energía vs Épocas



Mejor estado con mucho ruido



noise = 0.28

Letra

Paso 0

Paso 1

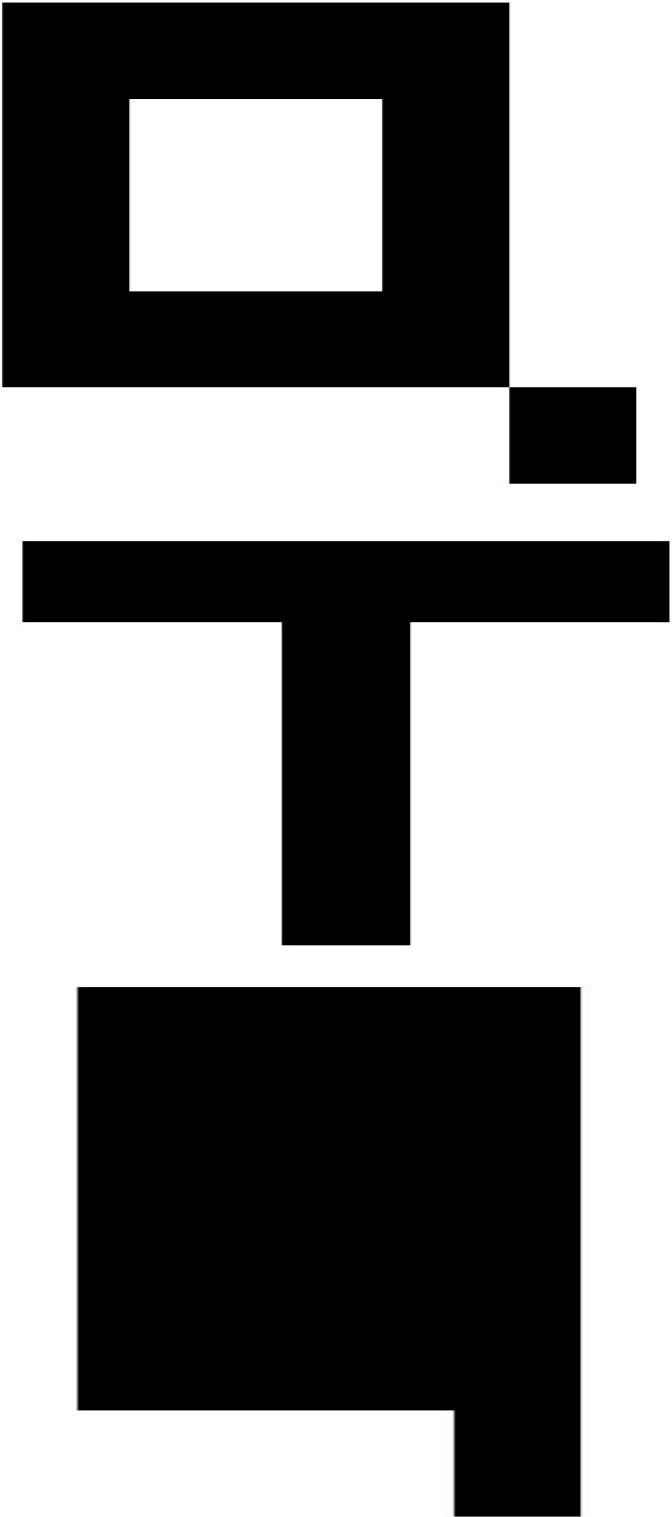
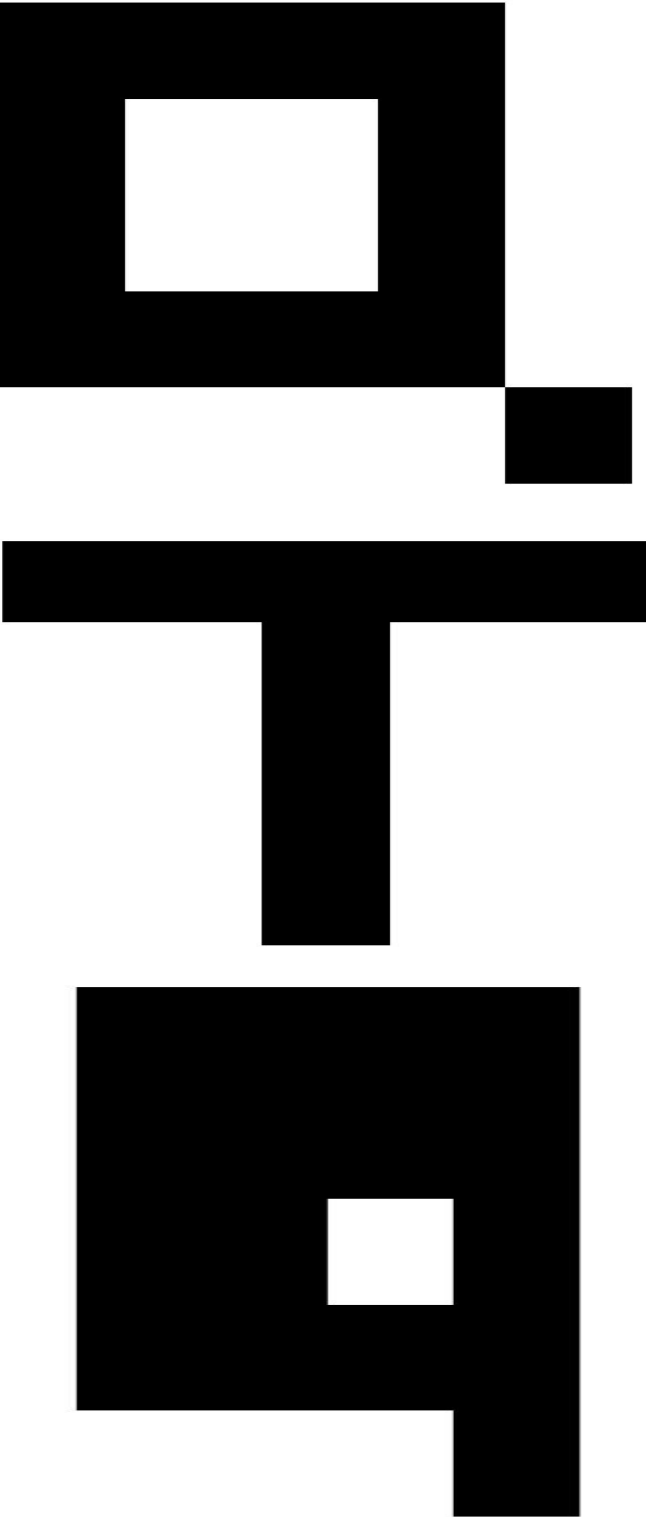
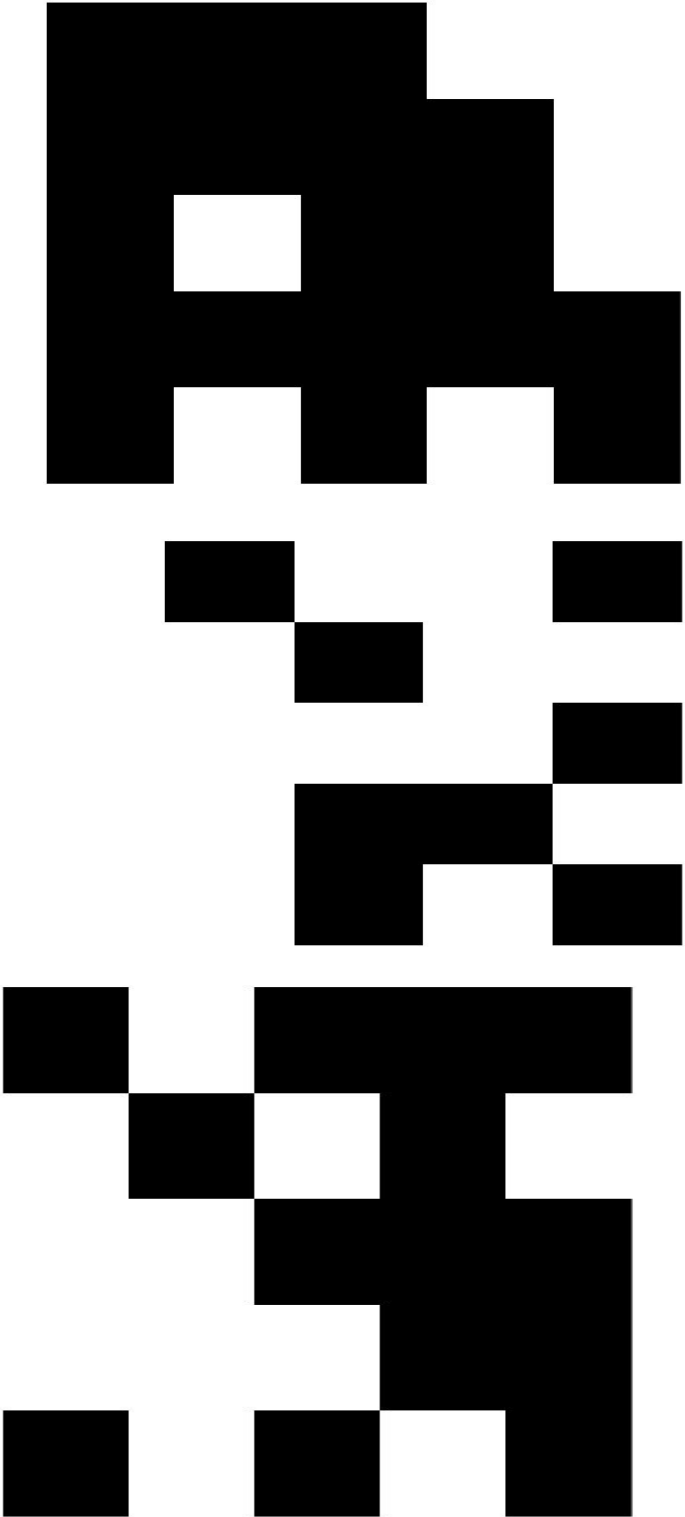
Paso 2

Paso 3

Q

T

X



Estado espurio!



Conclusiones

- El modelo no puede asociar correctamente los patrones cuando se utilizan conjuntos de patrones similares para el entrenamiento
- A mayor distinción entre los patrones del conjunto de entrenamiento, más resistente se vuelve el modelo al ruido
- El modelo no es infalible ya que puede asociar patrones ruidosos de entrada a patrones inexistentes (cae en un estado espurio)



¡MUCHAS
GRACIAS!

