APRENDIZAJE NO SUPERVISADO

Modelo de Kohonen

Sistemas de Inteligencia Artificial - 2023

TABLA DE CONTENIDOS

02.1

MODELO DE KOHONEN

REDES DE KOHONEN

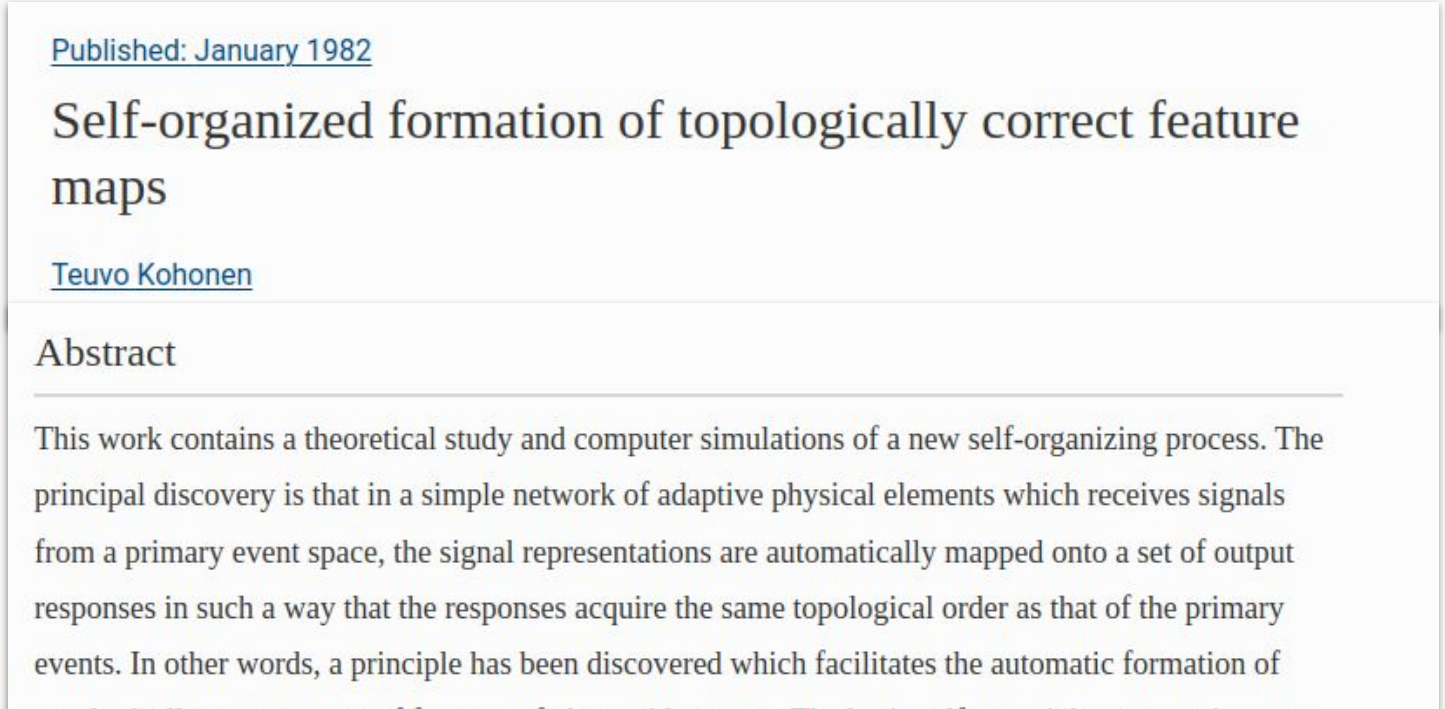
Aprendizaje No Supervisado
No existe información externa que indique si la red neuronal está operando correcta o incorrectamente.

Red de Kohonen
Durante el proceso de aprendizaje descubre por sí misma regularidades (patrones) en los datos de entrada.

SOM: Mapas Auto-Organizados

REDES DE KOHONEN

El autor es un investigador finlandés, Teuvo Kohonen, que publicó su idea por primera vez en 1982 [1] y luego siguió trabajando mucho tiempo [2].



Published: January 1982

Self-organized formation of topologically correct feature maps

Teuvo Kohonen

Abstract

This work contains a theoretical study and computer simulations of a new self-organizing process. The principal discovery is that in a simple network of adaptive physical elements which receives signals from a primary event space, the signal representations are automatically mapped onto a set of output responses in such a way that the responses acquire the same topological order as that of the primary events. In other words, a principle has been discovered which facilitates the automatic formation of

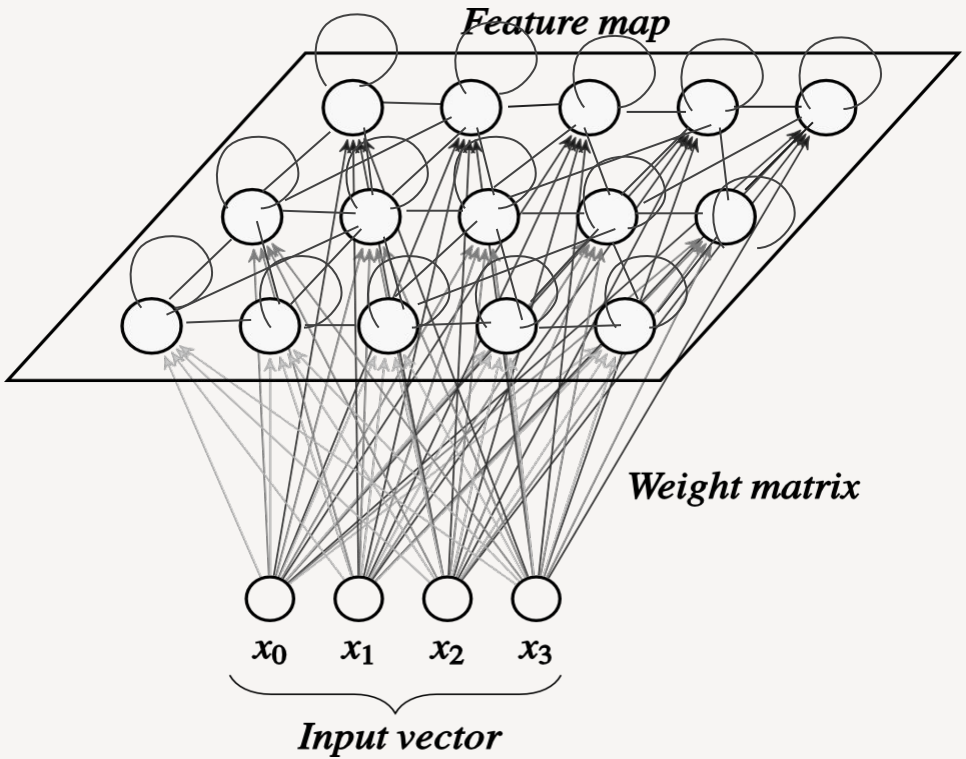
ARQUITECTURA

Las neuronas están conectadas

- con sí mismas positivamente
- con las neuronas vecinas. (R=n)

INPUT: elemento del training set

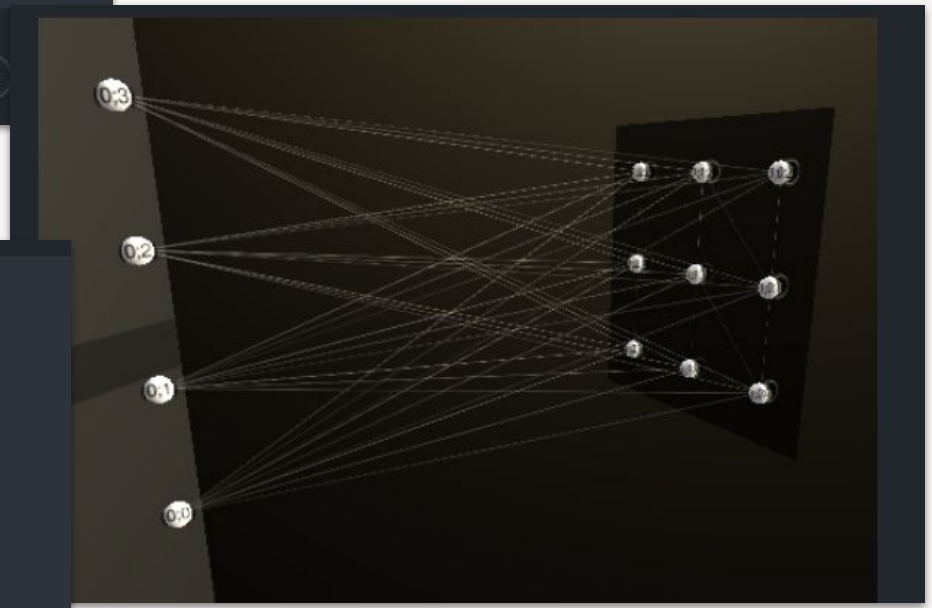
OUTPUT: grilla/mapa (M)



ARQUITECTURA

Search or jump to... / Pull requests

eugepineiro / neural-network-vr Public



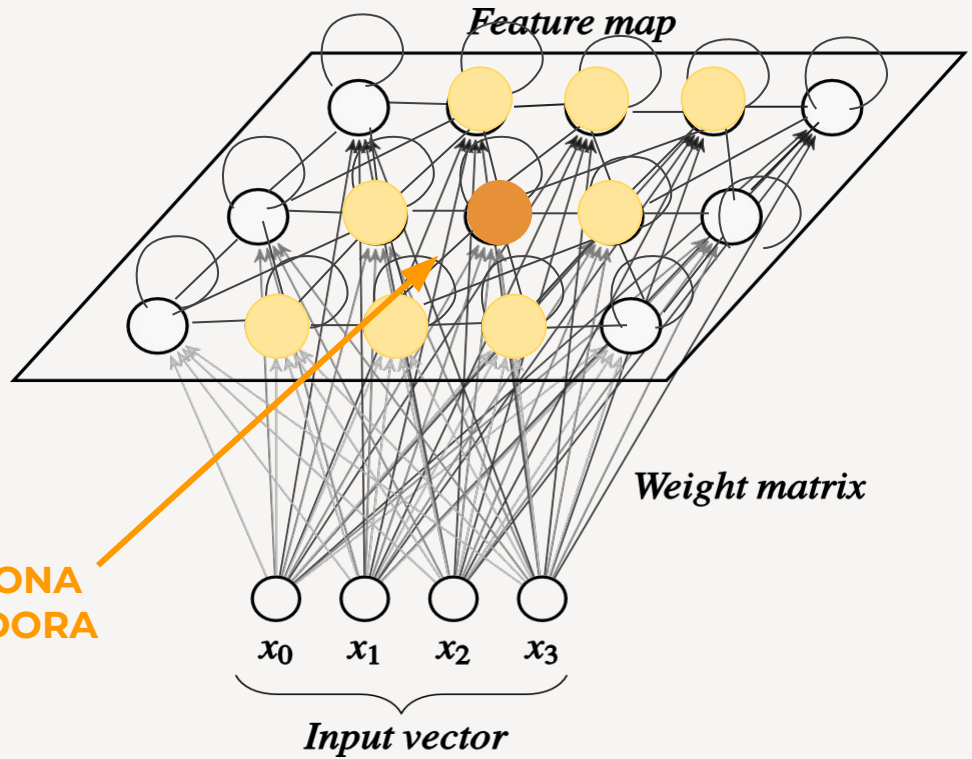

```
{
    "nn_type": "KOHONEN",
    "layers": [5, 3, 2],
    "kohonen": {
        "input_dimension": 10,
        "grid_dimension": 3
    },
    "improve_performance": false,
    "show_connections": true
}
```

APRENDIZAJE COMPETITIVO

Las neuronas compiten unas con otras

Objetivo → finalmente sólo una de las neuronas de salida se activa

Las demás son forzadas a valores de respuesta mínimos.



NEURONA GANADORA

A lo largo del tiempo (épocas), algunas unidades toman un nivel de activación mayor mientras que el nivel de las demás se anula.

NEURONA GANADORA
Dada la unidad de entrada x, la neurona que tenga vector de pesos w *"más parecido"* a x será ganadora.

Aprendizaje Competitivo

Esto implica una **clasificación**
(las entradas parecidas van hacia la misma neurona)

APRENDIZAJE COMPETITIVO

El objetivo de este aprendizaje es agrupar los datos que se introducen en la red.

El mapa nos mostrará un **agrupamiento.**

Las informaciones **similares** son clasificadas formando parte de la misma categoría o grupo y deben activar la **misma neurona** de salida.

RED DE KOHONEN

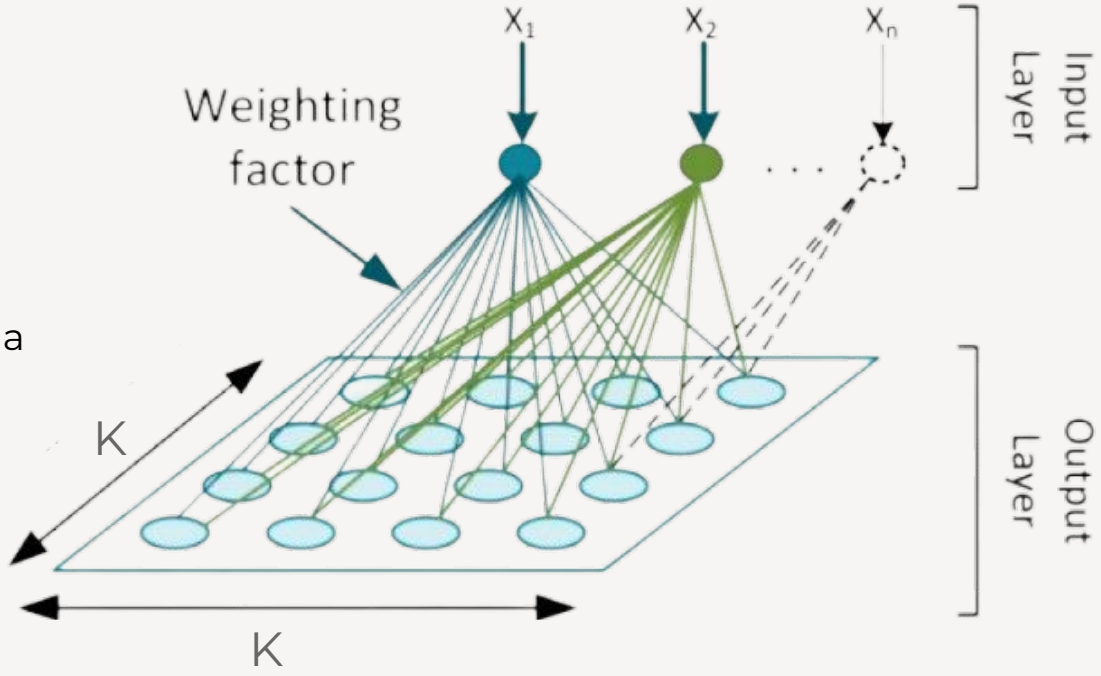
Red de una sola capa, en forma de grilla bidimensional (k × k) y en la que cada neurona está conectada a todas las componentes de un vector de entrada n-dimensional.

Entonces pasa de un espacio multidimensional a un espacio bidimensional.

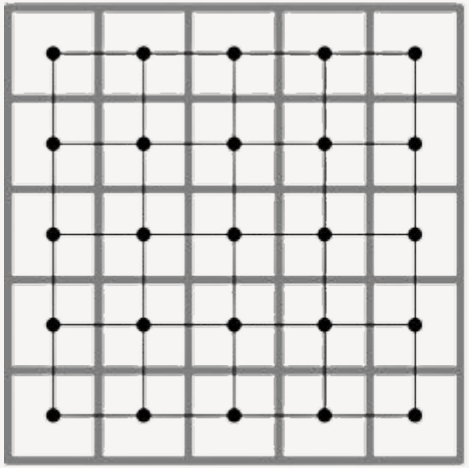
GRILLA

Dimensión KxK

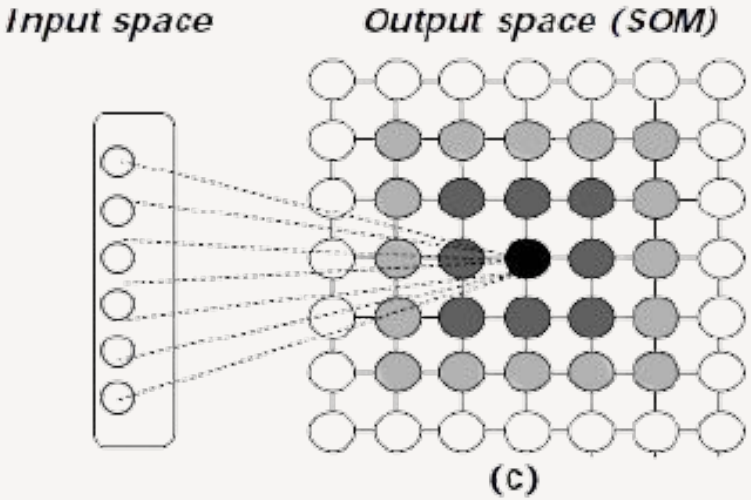
Si los datos de entrada tienen dimensión N → cada neurona de la grilla tiene N conexiones.



GRILLA RECTANGULAR

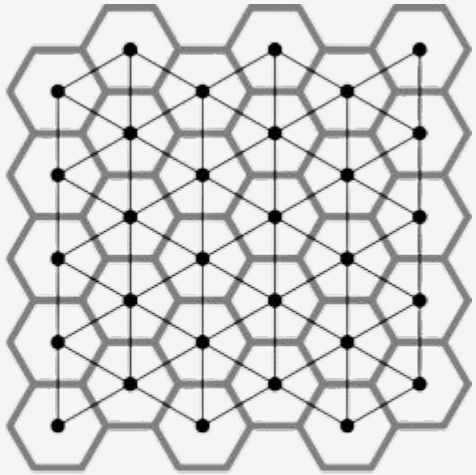


quadratic grid

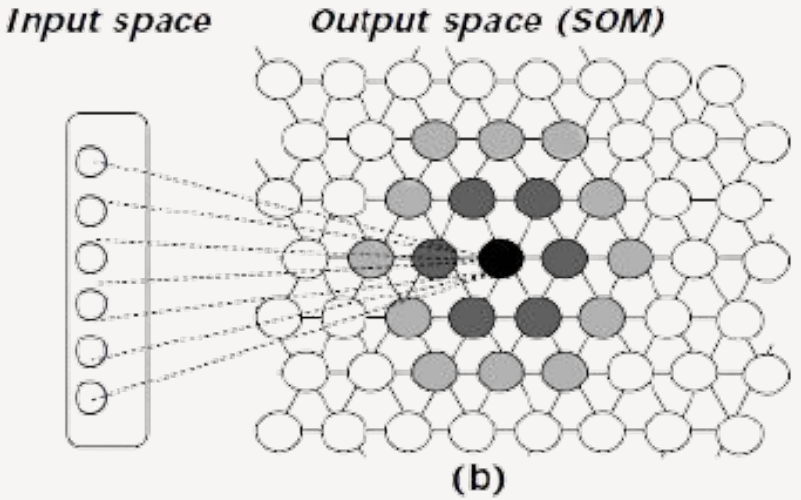


(c)

GRILLA HEXAGONAL



hexagonal grid

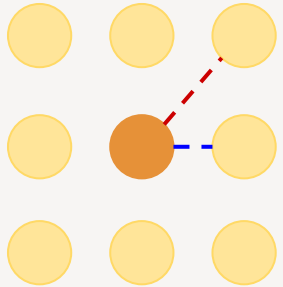


VECINDARIO

Se define un radio R, donde, para una grilla rectangular:

4-vecinos → R = 1

8-vecinos → R = sqrt(2)



Entradas similares
En cada neurona se concentran datos similares,
Neuronas vecinas contienen datos con algún grado de similitud entre sí.

CARACTERÍSTICAS

- Elegir la cantidad de neuronas de la grilla k × k.
- Cada neurona de salida $j \in \{1, \ldots, k^2\}$ tiene asociado un vector de pesos $W_j = (w_{j1}, \ldots, w_{jn})$ (el representante de esa neurona).
- W_j de cada neurona de salida tiene la misma dimensión que los datos de entrada.

02.2

ESTANDARIZACIÓN

FEATURE SCALING

Se utiliza para escalar los datos dentro un intervalo [a; b]

MIN-MAX FEATURE SCALING

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}(b - a) + a$$

Entre [0; 1] sería:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

ESTANDARIZACIÓN

Tomamos las variables del conjunto P={X_1, ..., X_p}. Cada X_i tiene n registros.
Se calculan los siguientes estadísticos unidimensionales:

MEDIA

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_i^j$$

DESVÍO ESTÁNDAR

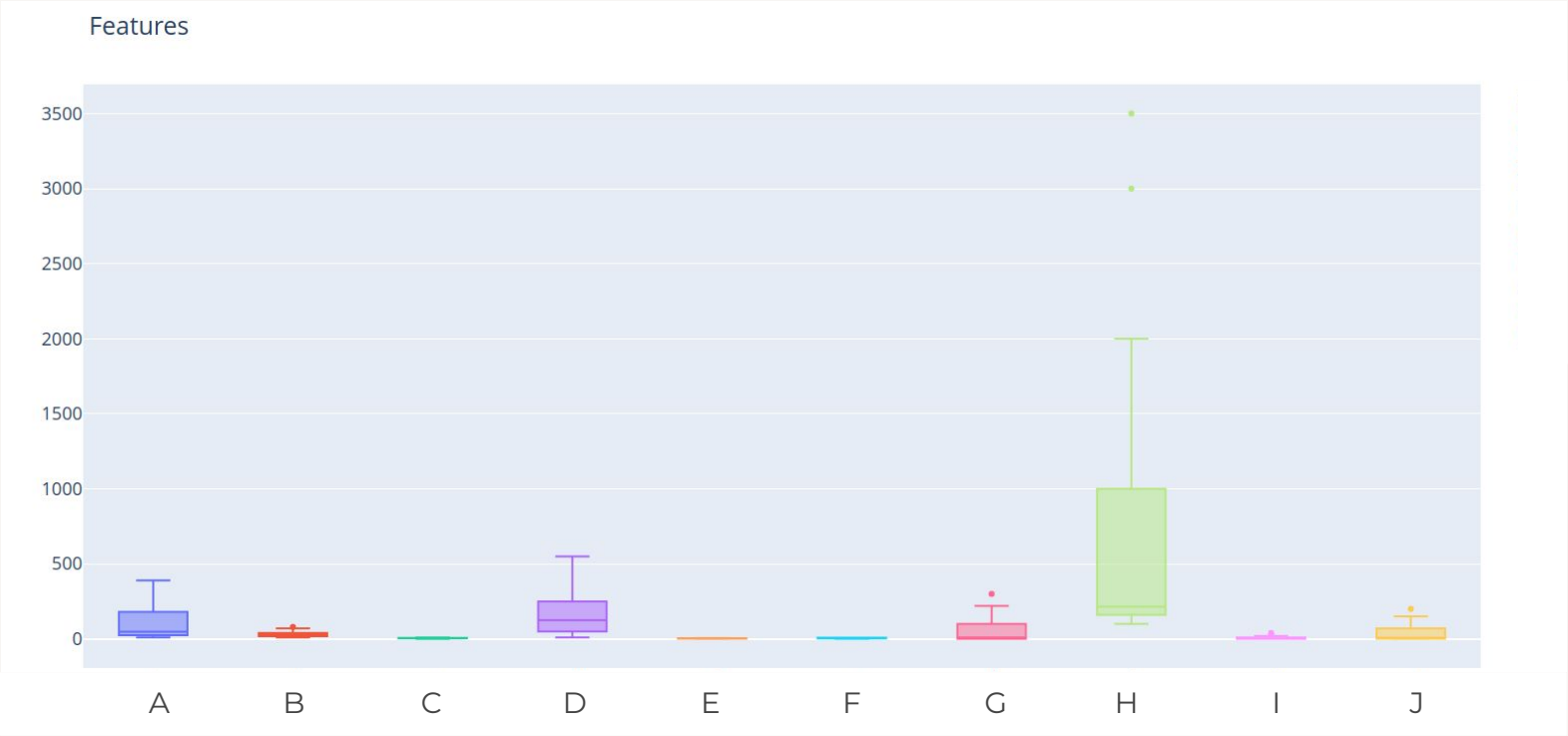
$$s_i^2 = \frac{1}{n}\sum_{j=1}^{n}(X_i^j - X_i)$$

VARIABLE ESTANDARIZADA

$$\tilde{X}_i = \frac{X_i - \bar{X}_i}{s_i}$$

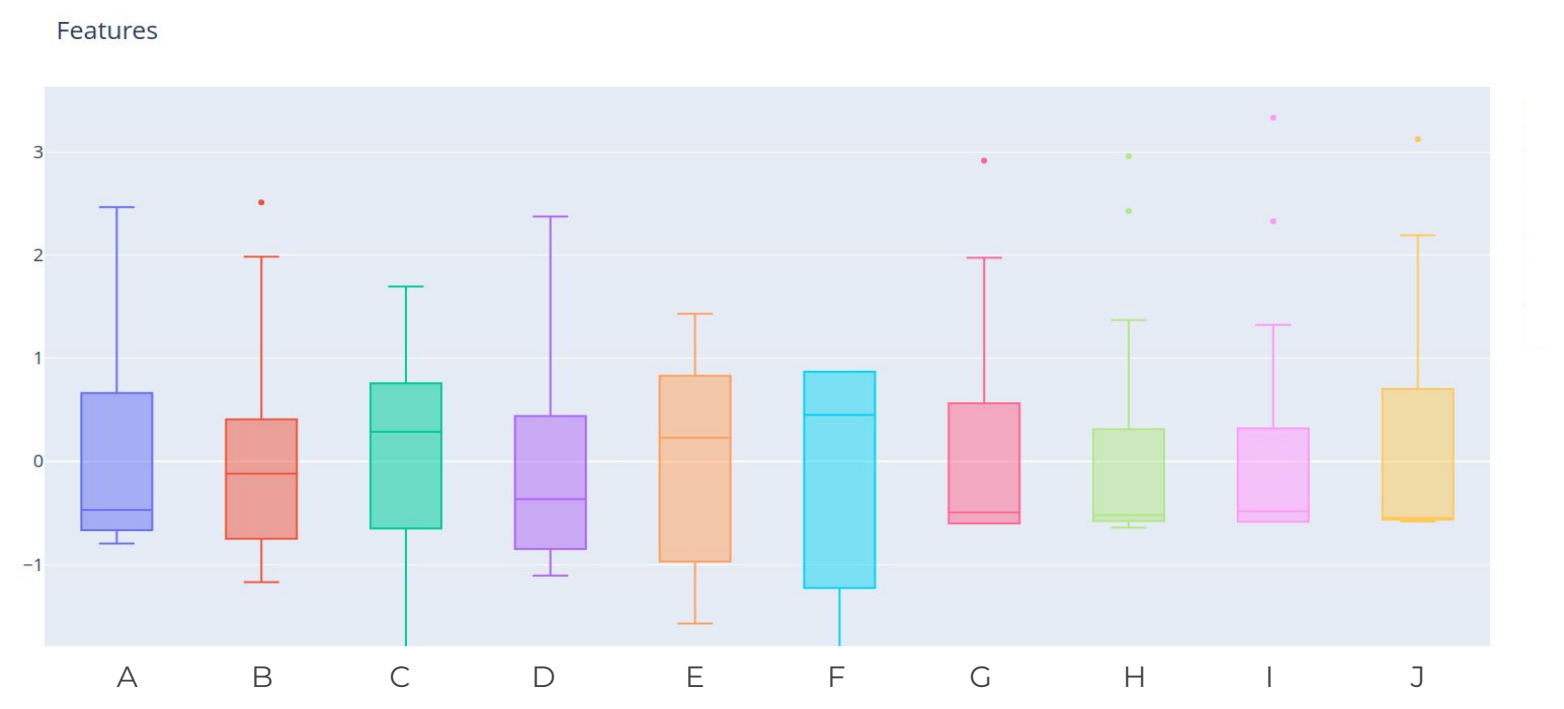
Nota: También llamada Z-Score

EJEMPLO



EJEMPLO

Si estandarizamos las variables



UNIT LENGTH SCALING

Si bien todas las anteriores son formas de normalización de datos, se le suele llamar “normalizar” o unit length scaling a dividir por la norma 2:

UNIT LENGTH SCALING

$$X' = \frac{X}{||X||}$$

02.3

ALGORITMO

ALGORITMO

INICIALIZACIÓN

1. $X^p = \{x_1^p , \ldots , x_n^p\}$, p = 1, . . . , P son los registros de entrada con dimensión n.
2. Definir la cantidad de neuronas de salida: k × k.
3. Inicializar los pesos W_j , j = 1, . . . , k^2 , cada $W_j = (w_{j1} , \ldots , w_{jn})$:
   - Con valores aleatorios con distribución uniforme.
   - Con ejemplos al azar del conjunto de entrenamiento.
4. Seleccionar un tamaño de entorno inicial con radio R(0).
5. Seleccionar la tasa de aprendizaje inicial η(0) < 1.

ALGORITMO

ITERACIÓN *i*

1. Seleccionar un registro de entrada X^p .

2. Encontrar la neurona ganadora $\hat{k}$ que tenga el vector de pesos $W_{\hat{k}}$ más cercano a X^p .
   Se define una medida de similitud d

$$W_{\hat{k}} = \underset{1 \leq j \leq N}{\arg\min} \{d(X^p - W_j)\}$$

3. Actualizar los pesos de las neuronas vecinas según la **regla de Kohonen**.

Se activa la neurona $\hat{k}$ ($n_{\hat{k}}$), que es la neurona ganadora.

ALGORITMO

REGLA DE KOHONEN (ITERACIÓN *i*, paso 3)

Está definido por el radio en la iteración, R(i), se actualiza el vecindario:

$$N_{\hat{k}}(i) = \{n / ||n - n_{\hat{k}}|| < R(i)\}$$

Donde:

- $n_{\hat{k}}$ es la neurona ganadora
- n es una neurona
- $N_{\hat{k}}(i)$ es el vecindario

R(0) es un dato de entrada
R(i) → 1 cuando i → ∞,
aunque también puede permanecer constante durante todo el proceso.

ACTUALIZACIÓN DE PESOS

Actualización de los pesos del vecindario de n_{k} utilizando la Regla de Kohonen:

- Si $j \in N_{k}(i) \rightarrow W_j^{i+1} = W_j^{i} + \eta(i) * (X^p - W_j^{i})$
- Si $j \notin N_{k}(i) \rightarrow W_j^{i+1} = W_j^{i}$

Donde $\eta(i) \rightarrow 0$. Por ejemplo $\eta(i) = 1/i$

CONVERGENCIA

Regla de Kohonen

$$W_j^{i+1} = W_j^i + \eta(i)^* (X^p - W_j^i)$$

¿Por qué converge?

$$W_{k\hat{}}^{i+1} - X^p = W_{k\hat{}}^i + \eta(i)(X^p - W_{k\hat{}}^i) - X^p$$
$$= (1-\eta(i))(W_{k\hat{}}^i - X^p)$$

Entonces

$$|| W_{k\hat{}}^{i+1} - X^p || \leq ||W_{k\hat{}}^i - X^p ||$$

Los pesos se parecen a los datos de entrada

SIMILITUD

Medidas de similitud (o funciones de propagación)

Distancia Euclídea:

$$W_{\hat{k}} = \arg\min_{1 \le j \le N} \{ \| X^p - W_j \| \}$$

Exponencial

$$W_{\hat{k}} = \arg\min_{1 \le j \le N} \{ e^{- \| X_p - W_j \|^2} \}$$

Importante: estandarizar todos los vectores

INICIALIZACIÓN

Valores iniciales de los pesos

- Se pueden inicializar con valores aleatorios
  Problema: algunas unidades quedan lejos de los valores iniciales y entonces nunca ganan.
  Se dice que son unidades muertas.

- Para evitar eso es mejor inicializar los pesos con muestras de los datos de entrada

INICIALIZACIÓN

Cantidad total de iteraciones

En función de la cantidad de neuronas de entrada (N)
Por ejemplo: 500*N.

Radio del vecindario

- R(0) puede ser el tamaño total de la grilla y va decreciendo hasta llegar a R = 1, donde solamente se actualizan las neuronas vecinas pegadas.
- Constante

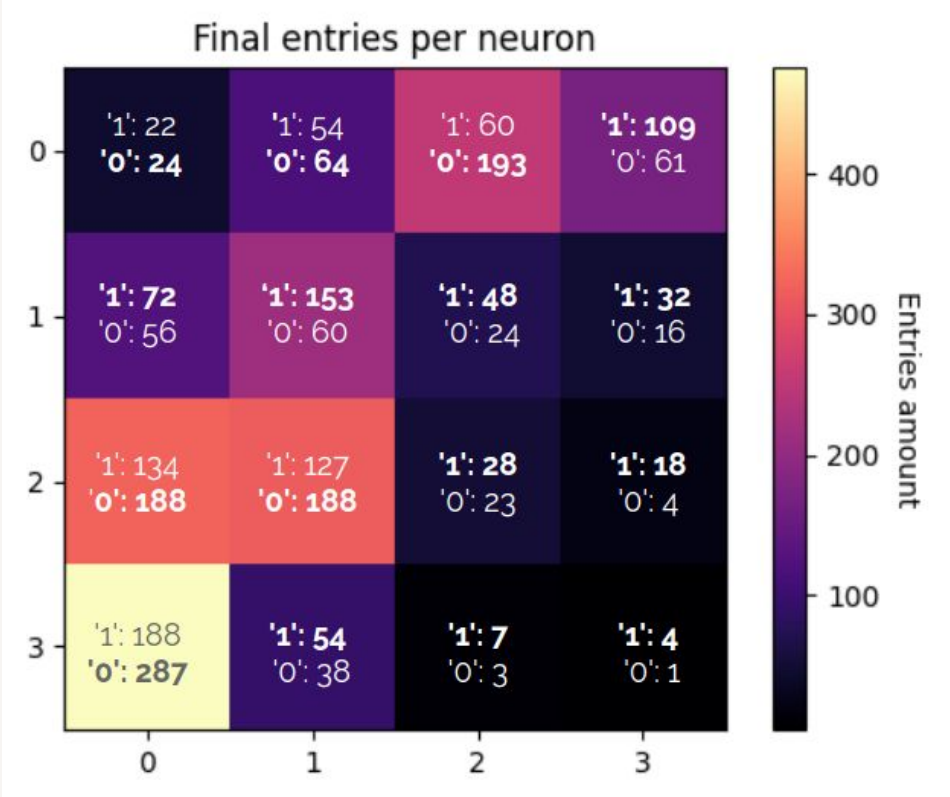
02.4

RESULTADOS

VISUALIZACIÓN DE RESULTADOS

Las neuronas de salida forman una matriz

Se puede ver en qué coordenadas se encuentra la neurona asociada a cada ejemplo de entrenamiento.



VISUALIZACIÓN DE RESULTADOS

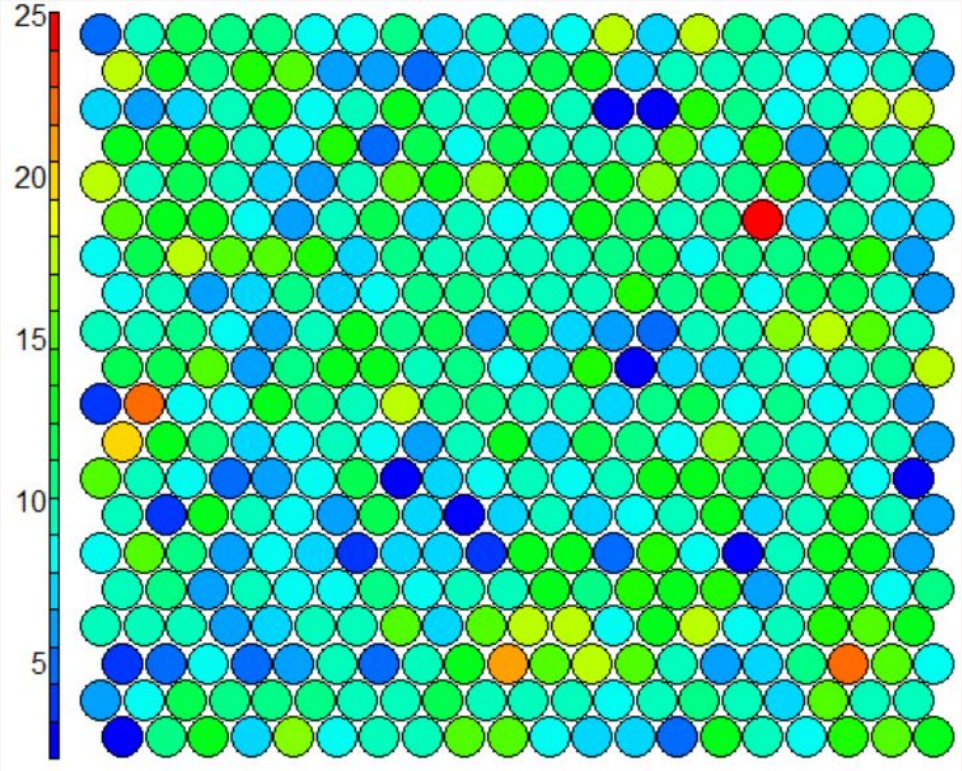
Ejemplo: Se desea hacer un censo de la población teniendo en cuenta.

- Cantidad de habitantes
- Promedio de edad
- Promedio de nivel de educación
- Número de autos registrados
- Número de personas desempleadas
- Número de personas subempleadas

¿Qué ciudades son parecidas?
Se utiliza una Red de Kohonen de 20x20

VISUALIZACIÓN DE RESULTADOS

Contar la cantidad de registros que van a cada neurona.



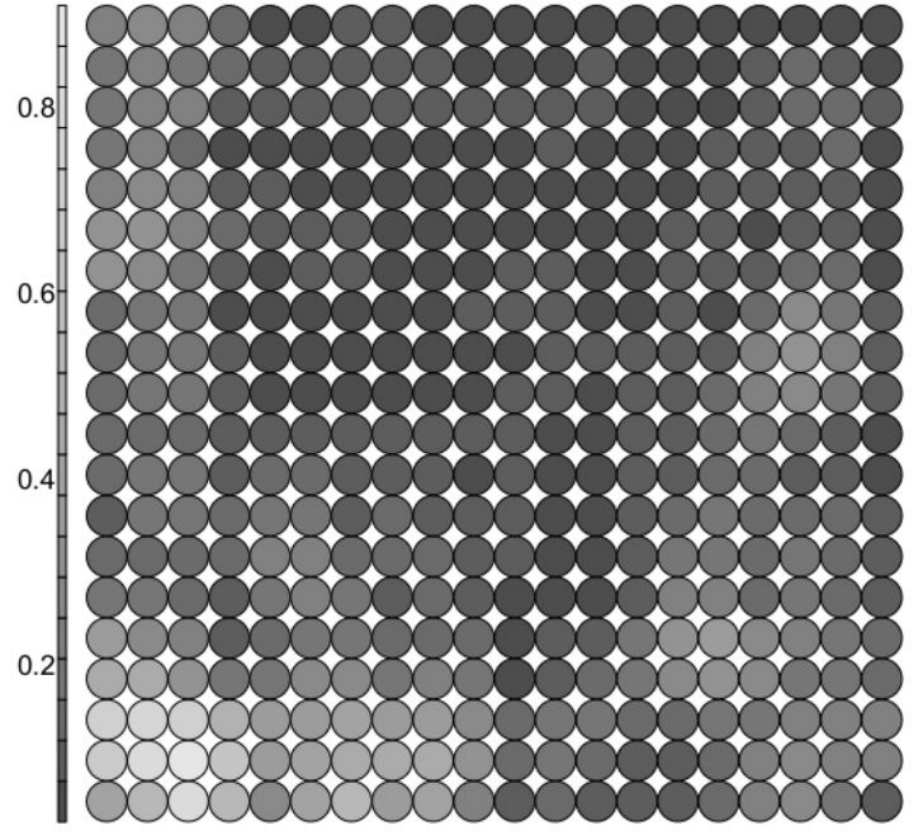
VISUALIZACIÓN DE RESULTADOS

MATRIZ U
(Unified Distance Matrix)

Para cada neurona el promedio la distancia euclídea entre:

- el vector de pesos de la neurona
- el vector de pesos de las neuronas vecinas.

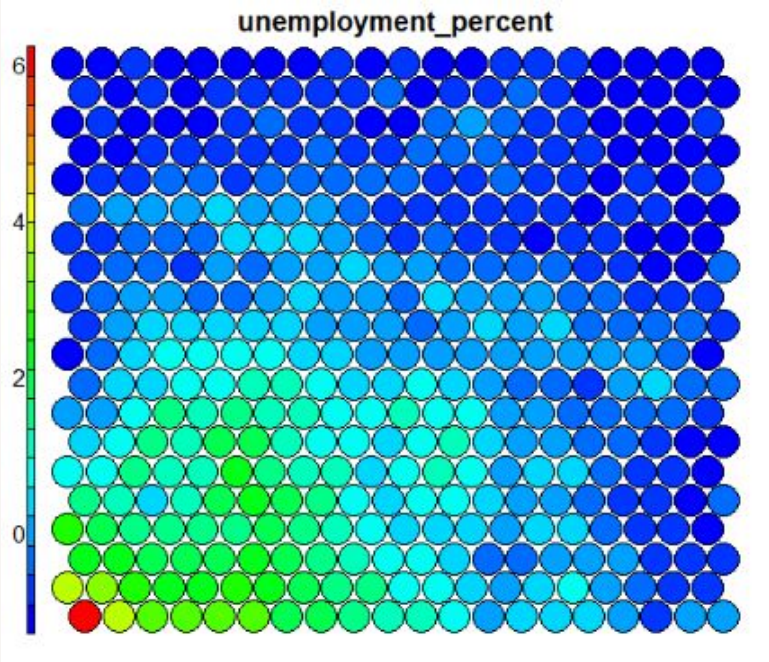
Si el método funciona entonces deberían observarse distancias pequeñas

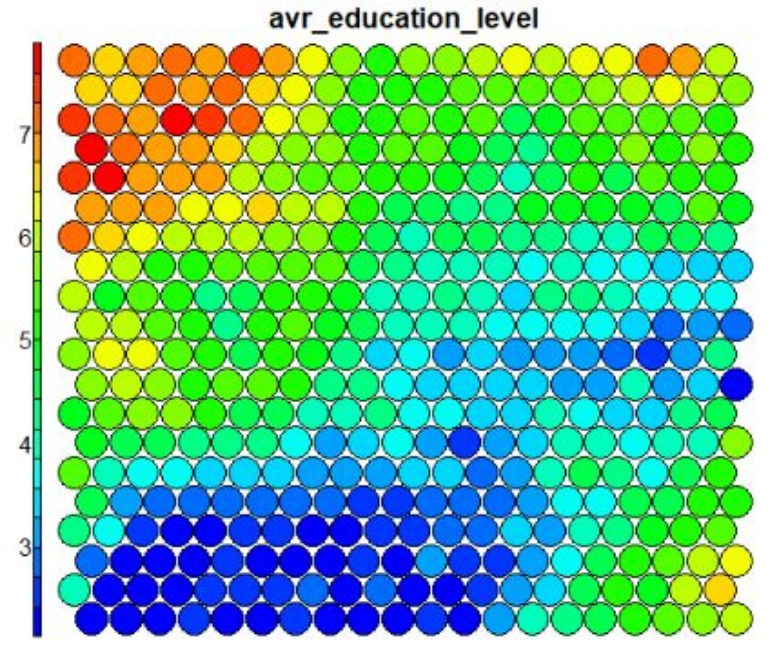


VISUALIZACIÓN DE RESULTADOS

OBSERVAR UNA SOLA VARIABLE

Se observa el valor promedio de una sola variable en cada neurona





RED DE KOHONEN

VENTAJAS

- Puede ser más rápida que el perceptrón multicapa.
- Aplicarse en casos donde el conjunto de datos no está etiquetado.
- Espacio multidimensional a bidimensional

DESVENTAJAS

- Si el conjunto de variables es muy grande puede ser difícil asociarlo con un conjunto bidimensional.
- Solo puede realizarse con variables numéricas.
- No hay un criterio demostrado para definir el tamaño de la grilla.

BIBLIOGRAFÍA

[1] T. Kohonen. *Self-organized formation of topologically correct feature maps*. Biological Cybernetics, 1(43):59–69, 1982.

[2] T. Kohonen. *The self-organizing map*. Neurocomputing, pages 1–6, 1998.