Sistemas de Inteligencia Artificial
Modelo de Kohonen

4 de mayo de 2021

Redes de Kohonen

Aprendizaje no supervisado

no existe ningún maestro externo que indique si la red neuronal está operando correcta o incorrectamente.

Durante el proceso de aprendizaje

la red de Kohonen descubre por sí misma regularidades en los datos de entrada.

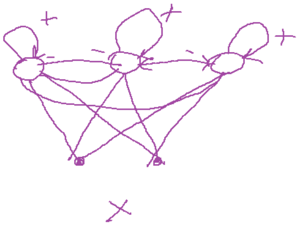
Redes de Kohonen

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].

Mapas Auto-Organizados: SOM

SOM: Self-Organized Maps

Redes de Kohonen: Introducción



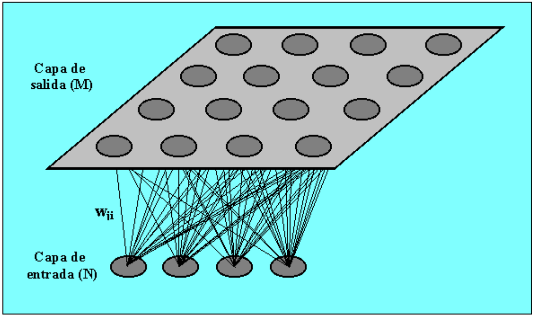
Las neuronas están conectadas con sí mismas positivamente y tiene conexiones inhibitorias con las neuronas vecinas.

Redes de Kohonen: Introducción

Este sistema

produce, a lo largo del tiempo, que alguna de las unidades tome un nivel de activación mayor mientras el nivel de las demás unidades se anula.

Estructura



La salida es una grilla (o un mapa).

Redes de Kohonen: Introducción

Neurona ganadora

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

Esto implica una clasificación

porque las entradas parecidas van hacia la misma neurona.

Aprendizaje Competitivo

Es aquel en que

las neuronas compiten unas con otras con el objetivo de que finalmente sólo una de las neuronas de salida se active y las demás sean forzadas a valores de respuesta mínimos.

Aprendizaje Competitivo

El objetivo de este aprendizaje

es agrupar los datos que se introducen en la red.

De esta forma

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

Redes de Kohonen

La red de Kohonen es

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

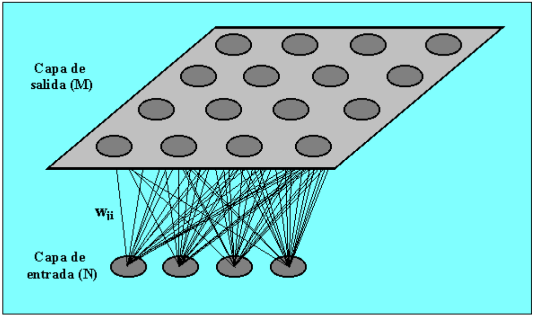
Entonces pasa de un espacio multidimensional a un espacio bidimensional.

Redes de Kohonen

Entrada

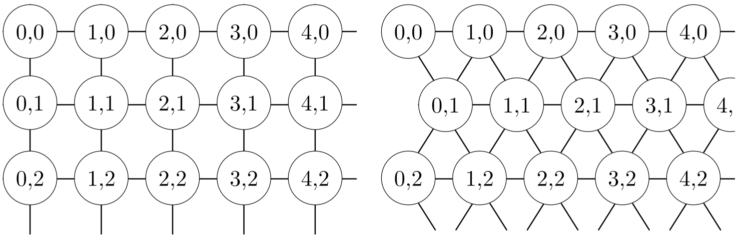
Ejemplos del conjunto de entrenamiento.

Estructura



La salida es una grilla (o un mapa).
Si los datos de entrada tienen dimensión n, entonces cada neurona de la grilla tiene n conexiones.

Forma de la grilla rectangular o hexagonal



Permite definir unidades vecinas, por ejemplo, las 4-vecinas.

Vecindario

El Radio R

- 4-vecinos $\Rightarrow R = 1$
- 8-vecinos $\Rightarrow R = \sqrt{2}$

Entradas similares...

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

Redes de Kohonen: Características

- Elegir la cantidad de neuronas de la grilla $k \times k$.
- 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).
- Los pesos W_j de cada neurona de salida, tienen la misma dimensión que los datos de entrada.

Redes de Kohonen

Algoritmo: Paso Inicial

1. $X^p = \{x_1^p, \ldots, x_n^p\}, \; p = 1, \ldots, P$ son los registros de entrada.
2. Definir la cantidad de neuronas de salida: $k \times k$.
3. Inicializar los pesos $W_j, \; j = 1, \ldots, k^2$, 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 $R(0)$.
5. Seleccionar la tasa de aprendizaje inicial $\eta(0) < 1$.

Redes de Kohonen: Algoritmo

En la iteración t

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: elegir una medida de similitud d y luego calcular:

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

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

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

Redes de Kohonen

Iteración t, paso 3

Está definido por el radio $R(t)$,

$$N_{\hat{k}}(t) = \{neu / ||neu - neu_{\hat{k}}|| < R(t)\}$$

$R(0)$ es un dato de entrada y $R(t) \to 1$ cuando $t \to \infty$, aunque también puede permanecer constante durante todo el proceso.

Redes de Kohonen: Algoritmo

Actualización de los pesos de las neuronas vecinas de $\hat{k}$

utilizando la regla de de Kohonen:

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

donde $\eta(t) \rightarrow 0$. Por ejemplo $\eta(t) = \frac{1}{t}$.

Convergencia

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

Entonces

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

Redes de Kohonen

Medidas de similitud

- Distancia Euclídea:

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

- Correlación:

$$W_{\hat{k}} = \arg \max_{1 \leq j \leq N} \{ (X^p)^t * W_j \}$$

normalizar todos los vectores.

Redes de Kohonen: Paso Inicial

Valores iniciales de los pesos

- Los pesos se pueden inicializar con valores aleatorios: Puede tener el problema de que algunas unidades queden lejos de los valores iniciales y entonces nunca ganen. Se dice que son unidades muertas.
- Para evitar eso se puede inicializar los pesos con muestras de los datos de entrada.

Redes de Kohonen

- La cantidad total de iteraciones conviene elegirla en función de la cantidad de neuronas de entrada, por ejemplo $500 * n$.
- El valor de $R(0)$ puede ser el tamaño total de la red y va decreciendo hasta llegar a $R_f = 1$, donde solamente se actualizan las neuronas vecinas pegadas.

Preprocesamiento de los datos

Estandarizar las variables del conjunto de entrenamiento y prueba

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

donde $\bar{X}_i$ es la media y s_i es la desviación estándar de X_i.

Visualización del Resultado

Observar

- Como las neuronas de salida forman una matriz, podemos ver en qué coordenadas se encuentra la neurona asociada a cada ejemplo de entrenamiento.

Ejemplo: Censo en Irlanda

Datos: Para cada ciudad se toman los datos de las siguientes variables:

- Cantidad de habitantes
- Promedio de edad
- Promedio Nivel de Educación
- Número de autos registrados
- Número de personas desempleadas
- Número de personas subempleadas

para todas las ciudades Irlandesas

Se utiliza una red en que la salida es de 20×20.

Ejemplo: Censo en Irlanda

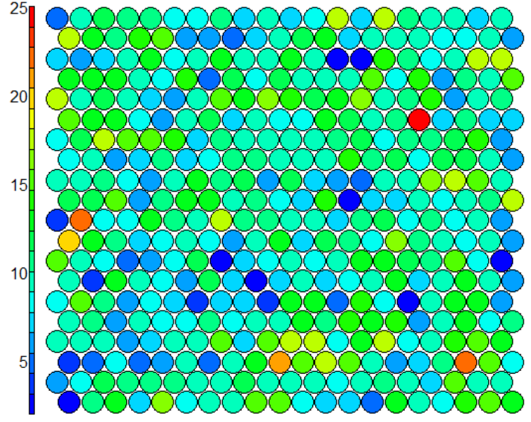
Queremos saber:

- Qué ciudades son parecidas, tomando todas las variables en cuenta.

Se utiliza una red en que la salida es de 20×20.

Visualización del Resultado

Contar la cantidad de registros que van a cada nodo



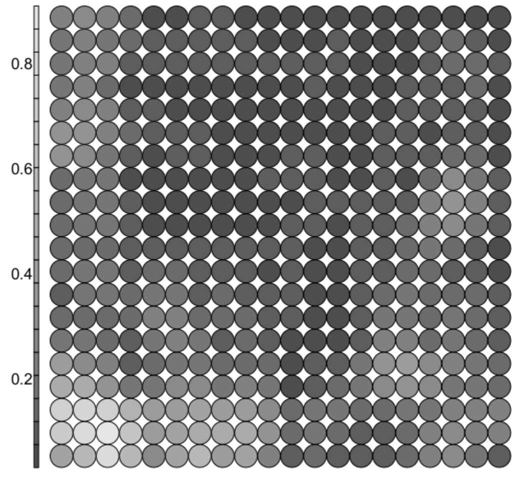
Visualización del Resultado

Matriz U

- La matriz U tiene, para cada nodo el promedio de la distancia euclídea entre el vector de pesos del nodo y el vector de pesos de los nodos vecinos.
- Si el método funciona, entonces deberían observarse todas distancias pequeñas

Visualización del Resultado

Distancias entre los pesos de neuronas vecinas



Conclusiones

Ventajas

- La red de Kohonen puede ser más rápida que el perceptrón multicapa.
- Puede aplicarse en casos donde el conjunto de datos solo tiene valores de entrada y no están etiquetados.

Desventajas

- Si el conjunto de variables es muy grande puede ser difícil asociarlo con un conjunto bidimensional.
- Hay que decidir el tamaño de la grilla desde el principio y no hay un criterio demostrado para hacerlo.

Referencias

[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.