

Abstract geometric lines in black on a white background, forming various polygons and intersecting lines, primarily located on the left side of the slide.

CLUSTERING CON REDES NEURONALES

MODELO SOM Y MUESTREO
ESTRATIFICADO CON EL ESTIMADOR
HORVITZ-THOMPSON

Cruz Martínez Ricardo

INTRODUCCIÓN

Realizar clasificación de datos siempre ha sido un tema de relevancia, sin embargo, usar algún modelo de red neuronal para aprendizaje no supervisado puede ser algo no muy ortodoxo, por lo que en este proyecto usaremos Mapas Autoorganizados (SOM), para efectuar esta tarea y veremos si los resultados son convincentes para llevar a cabo estimaciones en el ámbito de muestreo, usando estratificación. Tomando como referencia una población ficticia para comparar rendimiento, del mismo modo, lo haremos con el algoritmo k-means.

HIPÓTESIS

Dado que conocemos los datos de toda nuestra población, pensamos que el modelo SOM no será mejor que el algoritmo k-means para realizar las estimaciones, pues no es de los mejores modelos de aprendizaje no supervisado, sin embargo ambos presentan características similares aunque con diferente arquitectura.

Por otra parte, es muy probable que el modelo SOM sea mejor que realizar un muestreo aleatorio simple sin reemplazo, pues estamos hablando de un diseño muy sencillo para estimar.

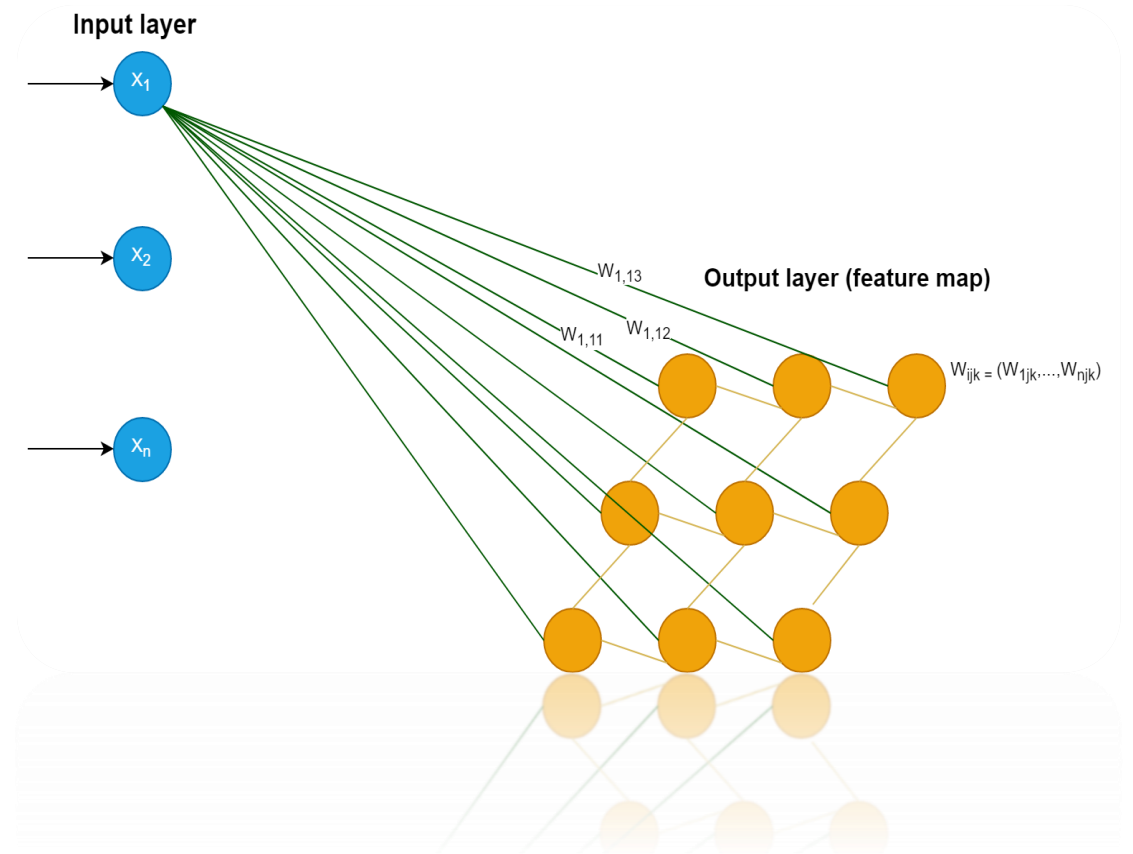


MODELO

Como se mencionó antes, ocuparemos el modelo de Mapas Autoorganizados (SOM), cuya arquitectura es algo sencilla como veremos a continuación:

- Este modelo consta de una capa de entrada de tamaño N , donde N representa el número de variables de nuestros datos y una capa de salida de tamaño M , donde M , en este caso, será el número de estratos que queremos.
- Las conexiones de ambas capas siempre son hacia adelante.
- Las neuronas de entrada están conectadas con todas y cada una de las neuronas de salida mediante pesos, por ende, tenemos que cada neurona de salida tiene asociado un vector de pesos que denotaremos W_i .
- La salida de nuestra red, será una y solo una neurona de salida la cual llamaremos neurona vencedora.

ARQUITECTURA DEL MODELO SOM



UNA DEFINICIÓN IMPORTANTE

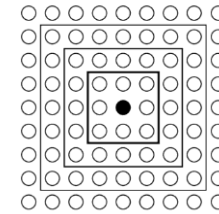
Vecindad de la neurona vencedora

Si suponemos que i^* es la neurona vencedora, las neuronas que distan d de la neurona i^* está dado por el siguiente conjunto:

$$N_{i^*}(d) = \{j : d(i^*, j) \leq d\}$$

ALGORITMO DE APRENDIZAJE (IDEA GENERAL)

- Para este proyecto, usaremos una topología rectangular.



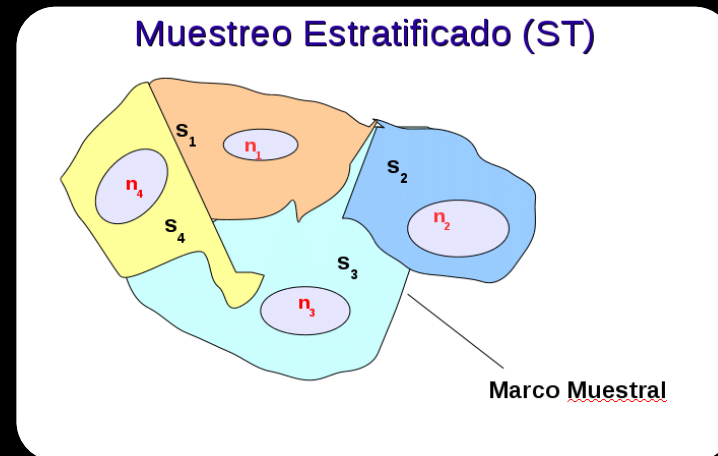
- Inicializar la distancia d y la matriz de pesos W con valores aleatorios pequeños.
- Para cada dato de entrada se selecciona el nodo j más próximo en términos de similitud.
- Actualizamos los pesos de la neurona vencedora de la siguiente manera

$$W_i(k+1) = W_i(k) + \alpha(k)(p_i - W_i(k)) \text{ para } i \in N_{i^*}(d).$$

- Con α la tasa de aprendizaje.
- Repetimos este proceso a partir del paso 3.

RESULTADOS

Dadas las pruebas realizadas, se obtuvieron resultados similares usando el modelo SOM al muestreo aleatorio simple sin reemplazo, por otra parte, los resultados obtenidos usando el algoritmo k-means resultó ser mejor que los modelos antes mencionados, por lo que nuestra hipótesis resultó ser cierta, al menos con el dataset que usamos.





CONCLUSIONES

Las redes neuronales supervisadas han cobrado gran relevancia hoy día por su enorme capacidad, sin embargo, los mapas autoorganizados, que es un modelo de red neuronal no supervisada, resultó no ser la mejor opción para nuestra tarea, aunque los resultados obtenidos no fueron del todo malos, pues se obtuvieron estimaciones muy similares al muestreo aleatorio simple con tamaños de muestra incluso menores pero el algoritmo k-means fue un tanto mejor para la construcción de nuestros estratos.