

Computación Neuronal y Evolutiva

(Práctica 3)



UNIVERSIDAD DE BURGOS

Autores:

Mario Ubierna San Mamés

Jorge Navarro González



Índice de Contenido

Aspectos relevantes de la generación de datos y programación del script.....	3
Resultados obtenidos	4
Caso 1 ([4 4],100,3,'hextop','linkdist')	5
Caso 2 ([4 4],100,3,'hextop','boxdist')	6
Caso 3 ([4 4],100,3,'gridtop','linkdist')	7
Caso 4 ([4 4],100,3,'gridtop','boxdist')	8
Caso 5 ([4 4],100,6,'hextop','linkdist')	9
Caso 6 ([4 4],100,6,'hextop','boxdist')	10
Caso 7 ([4 4],100,6,'gridtop','linkdist')	11
Caso 8 ([4 4],100,6,'gridtop','boxdist')	12
Caso 9 ([8 8],100,3,'hextop','linkdist')	13
Caso 10 ([8 8],100,3,'hextop','boxdist')	14
Caso 11 ([8 8],100,3,'gridtop','linkdist')	15
Caso 12 ([8 8],100,3,'gridtop','boxdist')	16
Caso 13 ([8 8],100,6,'hextop','linkdist')	17
Caso 14 ([8 8],100,6,'hextop','boxdist')	18
Caso 15 ([8 8],100,6,'gridtop','linkdist')	19
Caso 16 ([8 8],100,6,'gridtop','boxdist')	20
Conclusiones Obtenidas	21

Aspectos relevantes de la generación de datos y programación del script

```
function v = genclu(x,c,n,d)
% Funcion para generar clusters.
%
% genclu(X,C,N,D)
% X - matriz con los rangos de cada dimensi?n.
% C - n?mero de clusters a generar.
% N - n?mero de datos en cada cluster.
% D - Desviaci?n est?ndar de los clusters (valro por defecto: 1).
%
% Devuelve la matriz de datos generados seg?n los valores indicados para
% los par?metros.

if nargin < 3, error(message('Deben proporcionarse al menos 3 argumentos')), end
% Si no se establece un valor para la desviaci?n est?ndar, se fija el valor
% por defecto.
if nargin == 3, d = 1; end

[r,q] = size(x);
minv = min(x')';
maxv = max(x')';
v = rand(r,c) .* ((maxv-minv) * ones(1,c)) + (minv * ones(1,c));
t = c*n;
v = repmat(v,1,n) + randn(r,t)*d;
```

Esta es la primera función que hemos usado para sacar el conjunto de datos, en ella tenemos que especificar la matriz de rangos de cada dimensión, en nuestro caso va a ser de -2 a 2, número de clusters que son 4 en nuestro caso, número de datos por cada cluster que son 50 y la desviación estándar que la hemos puesto a 0.075.

```
function SOM(x,dimensiones,steps,vecindario,topologia,distancia)
%
%
net = selforgmap(dimensiones,steps,vecindario,topologia,distancia);
net = configure(net,x);
net.trainParam.epochs = 300;
net = train(net,x);
```

Esta es la función donde le pasamos el conjunto de datos, todos los parámetros los cuales vamos a cambiar, y por último también dentro de esta función se entrena.

```
function x = valores3D()
%
%
rangos = [-4 4; -4 4; -4 4];
clusters = 4;
puntos = 100;
desvs = 0.075;
x = genclu(rangos,clusters,puntos,desvs);

figure
plot3(x(1,:),x(2,:),x(3,:),'.g')
grid on;
title('Datos de entrada');
```

Esta función la hemos creado simplemente para ir generando los conjuntos de datos para elegir el que más nos gustara.



Resultados obtenidos

A continuación, se van a especificar todos los casos que hemos hecho en nuestro trabajo, como pone en el informe hemos elegido dos posibilidades de los 4 parámetros a modificar, por tanto son 16 posibles casos que habría que probar.

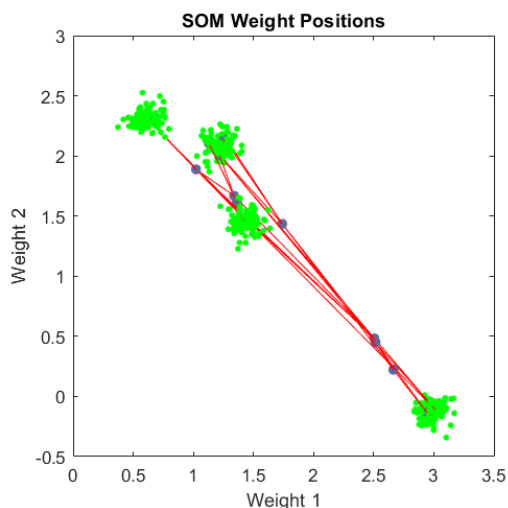
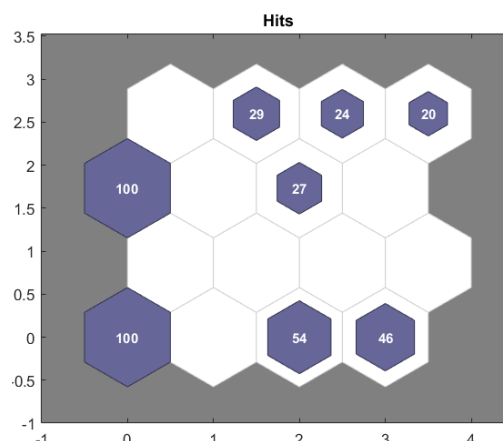
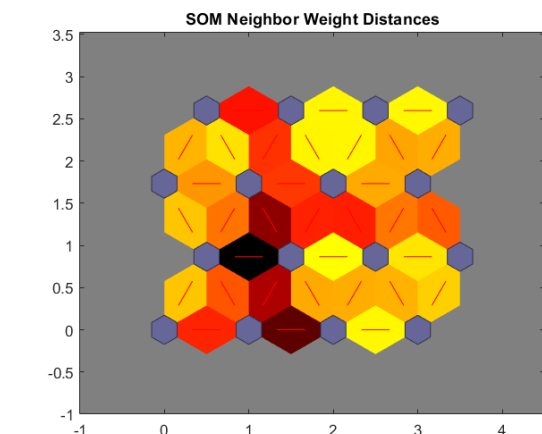
Este trabajo consta de 16 casos, el desarrollo va a seguir el siguiente orden:

- Caso 1 ([4 4],100,3,'hextop','linkdist').
- Caso 2 ([4 4],100,3,'hextop','boxdist').
- Caso 3 ([4 4],100,3,'gridtop','linkdist').
- Caso 4 ([4 4],100,3,'gridtop','boxdist').
- Caso 5 ([4 4],100,6,'hextop','linkdist').
- Caso 6 ([4 4],100,6,'hextop','boxdist').
- Caso 7 ([4 4],100,6,'gridtop','linkdist').
- Caso 8 ([4 4],100,6,'gridtop','boxdist').
- Caso 9 ([8 8],100,3,'hextop','linkdist').
- Caso 10 ([8 8],100,3,'hextop','boxdist').
- Caso 11 ([8 8],100,3,'gridtop','linkdist').
- Caso 12 ([8 8],100,3,'gridtop','boxdist').
- Caso 13 ([8 8],100,6,'hextop','linkdist').
- Caso 14 ([8 8],100,6,'hextop','boxdist').
- Caso 15 ([8 8],100,6,'gridtop','linkdist').
- Caso 16 ([8 8],100,6,'gridtop','boxdist').

Añadir también que en este caso no va a ser tan fácil como en otras prácticas elegir cuál de los posibles casos va a ser mejor, debido a que no tenemos un valor del error con el cual medir cuál de los casos va a ser el mejor. Por tanto, vamos a mirar una aproximación de cuál de los siguientes casos vamos a seleccionar como mejor.

Hemos elegido para probar dos valores distintos para cada uno de los parámetros de dimensiones de la rejilla, tamaño del vecindario, función de topología y función de distancia como se indica en la práctica.

Caso 1 ([4 4],100,3,'hextop','linkdist')

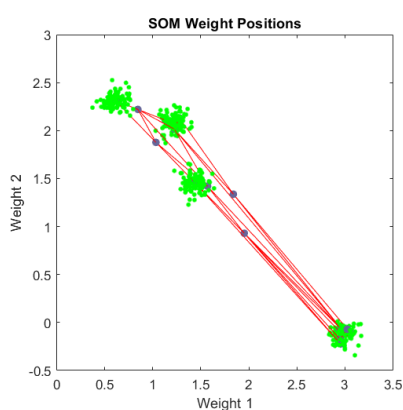
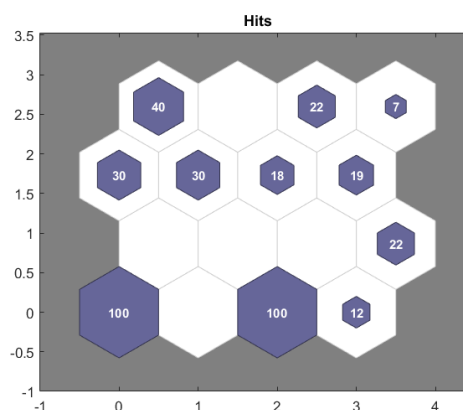
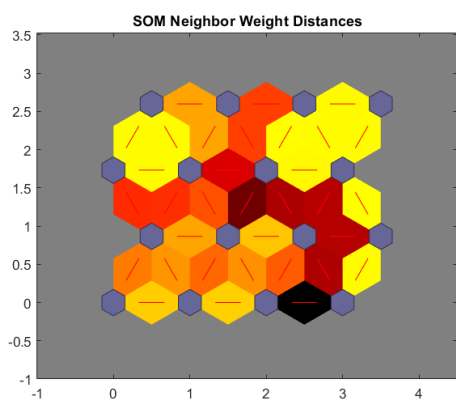


Como se puede ver en estas imágenes, podemos ver que se diferencian perfectamente los 4 clusters que hemos especificado en nuestro conjunto de datos.

Vamos a poseer dos neuronas las cuales van a tener el valor máximo de 100 en dos de los clusters y luego vamos a tener otros dos clusters que poseen más de 1 neurona.

En la primera imagen podemos ver como los grupos de neuronas que pertenecen al mismo cluster van a tener unos enlaces en amarillo, lo cual significa que las neuronas están muy cerca unas de otras en ese lugar. También vamos a ver en la segunda imagen neuronas que no tienen nada, es decir en blanco, esto se explica bien con la tercera imagen, ya que esas neuronas son las que se corresponden con las que no están dentro de ningún cluster, es decir, que no manejan nada de información.

Caso 2 ([4 4],100,3,'hextop','boxdist')

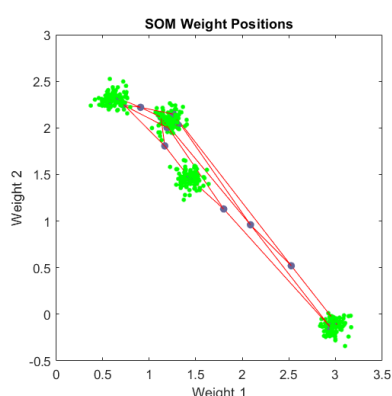
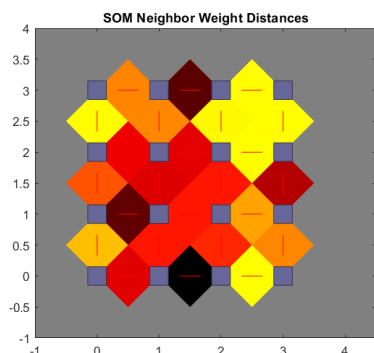


En estas gráficas hemos cambiado la función de distancia respecto a la anterior, como se puede ver es más complicado quizá ver los diferentes clusters si sólo vemos la segunda imagen.

Pero si nos fijamos también en la primera imagen, podemos ver los enlaces amarillos aparecer en las zonas donde se concentran los clusters.

También podemos fijarnos en las neuronas en blanco que como dijimos anteriormente y es siempre lo mismo en todos los casos, se corresponden con neuronas que no tienen información en ellas, es decir, que por ellas no pasa información.

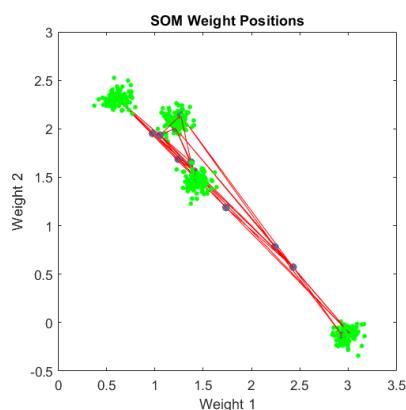
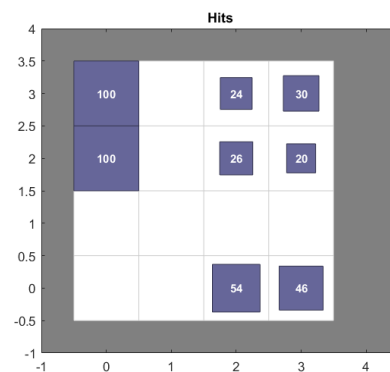
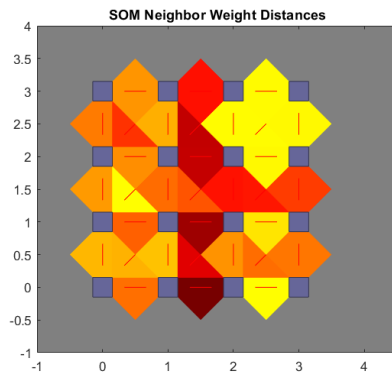
Caso 3 ([4 4],100,3,'gridtop','linkdist')



En este ejemplo no se aprecia tan fácilmente uno de los clusters debido a que hay un dato de una de las neuronas que como se puede apreciar en la segunda foto, se corresponde con el cluster de arriba a la derecha, pero si nos fijamos en la imagen de las distancias, hay una distancia no tan cercana de esa neurona con el resto.

También podemos comentar como en todas las anteriores que las neuronas que están en blanco en la segunda imagen se corresponden con las neuronas que en la tercera imagen están entre los clusters, sin nada de información pasando por ellas.

Caso 4 ([4 4],100,3,'gridtop','boxdist')



Como se puede observar en este ejemplo se ve bastante definidos todos los clusters que existen en esta red.

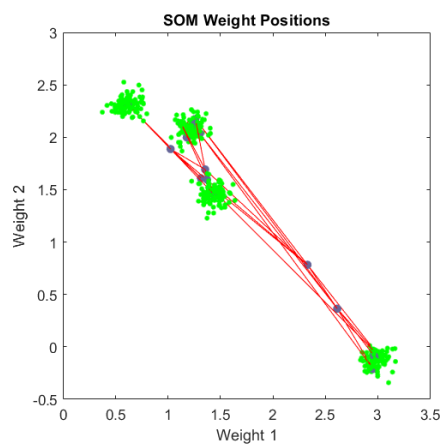
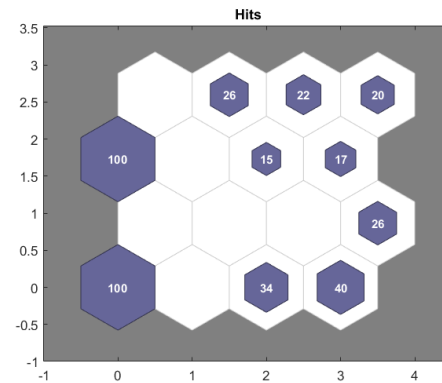
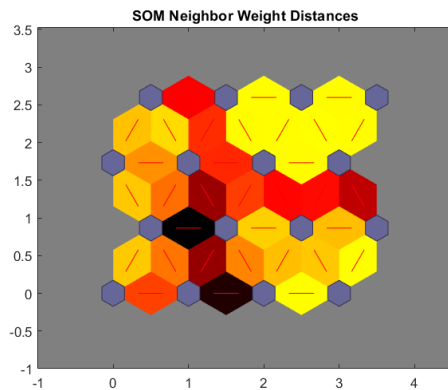
Podemos ver en la segunda imagen como puede parecer que dos clusters arriba a la izquierda están muy juntos y podrían distinguirse con dificultad, pero sabemos que todas las neuronas de un cluster tienen que sumar 100, además si nos vamos a la

gráfica de las distancias podemos ver que en realidad no están tan juntas entre sí.

También podemos ver que las neuronas que son dos o más en los distintos clusters que tenemos, en la gráfica de distancias están bastante juntas, con enlaces amarillos, por tanto, podemos ver con facilidad que pertenecen al mismo cluster juntando las dos gráficas.

Se puede decir que las neuronas de la segunda imagen que tenemos en blanco son las neuronas que en la tercera imagen están sin pertenecer a ningún cluster.

Caso 5 ([4 4],100,6,'hextop','linkdist')



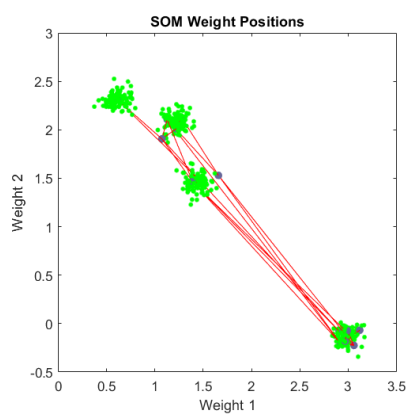
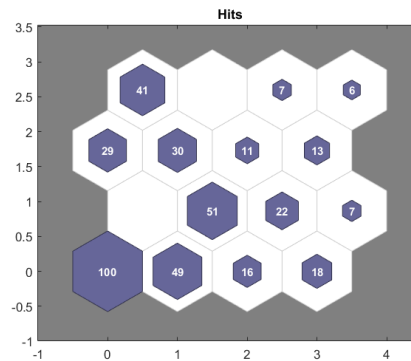
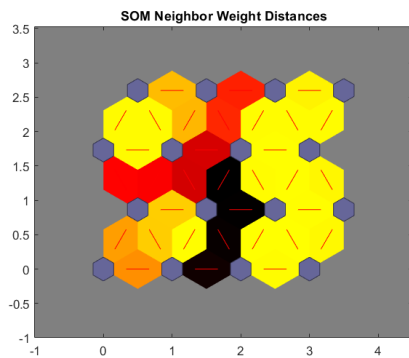
Como se puede ver en este ejemplo se puede ver con relativa claridad cual son cada uno de los clusters, si nos fijamos en cualquiera de las imágenes vemos como están separados unos de otros.

Además, fijándonos en la primera imagen podemos ver que todas las neuronas pertenecientes al mismo cluster se encuentran con enlaces amarillos, que son los que indican la

mayor cercanía entre neuronas.

Podemos ver también como en todos los casos, que las neuronas en blanco o neuronas que en la gráfica de distancias están muy separadas de las demás son las que no pasa nada de información por ellas.

Caso 6 ([4 4],100,6,'hextop','boxdist')



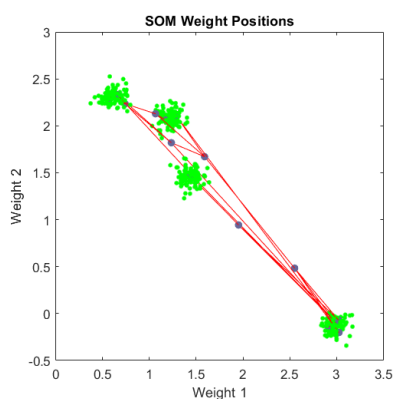
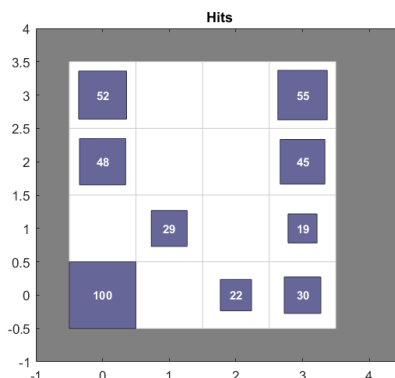
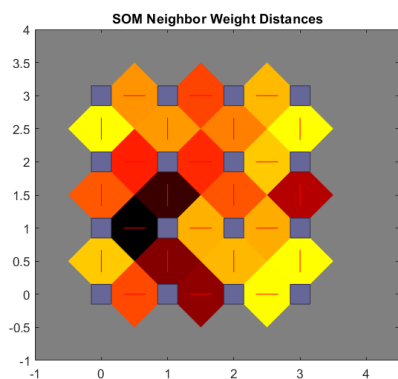
Estas gráficas tampoco se ven demasiado bien la diferencia entre clusters, debido a que, si vemos la segunda imagen, podemos ver datos muy juntos en los cuales no se ve con claridad con cual va cada dato, también si nos fijamos en el gráfico de distancias podemos intuir más o menos como van a ir las neuronas y su distribución, pero no es algo que se vea con exactitud.

Como decimos siempre, las neuronas en blanco en la segunda gráfica se corresponden con las neuronas en las cuales no hay paso de información.

Añadir también que tampoco es del todo óptimo que las neuronas dentro de los clusters estén demasiado dispersas, como es en este caso, por ejemplo.



Caso 7 ([4 4],100,6,'gridtop','linkdist')



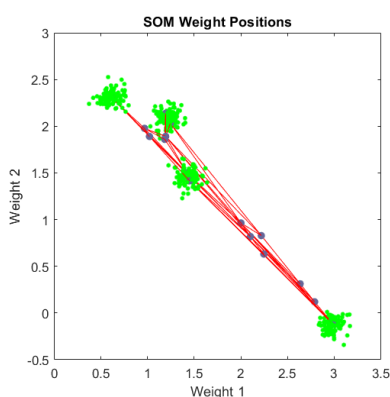
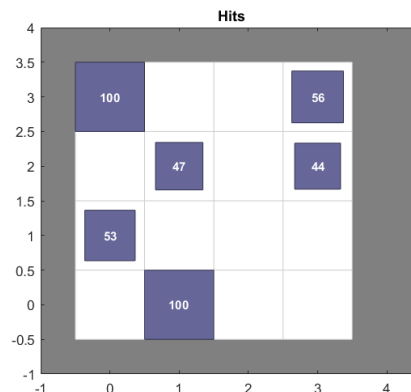
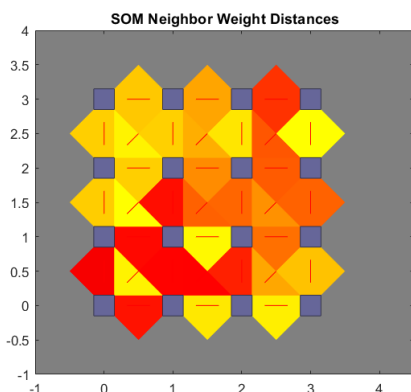
En este caso si mezclamos la primera gráfica con la segunda gráfica podemos ver que se ve relativamente bien la diferencia entre los distintos clusters.

Queda una neurona un poco apartada de uno de ellos, pero más o menos sí que se puede intuir.

Aun así, no es de las que mejor se puede ver por lo tanto no va a ser una de las seleccionadas como mejor.

Como hemos dicho ya en todos los casos, cuando en la segunda gráfica no hay nada, es decir, se encuentra en blanco quiere decir que no hay ninguna información pasando por dicha neurona.

Caso 8 ([4 4],100,6,'gridtop','boxdist')

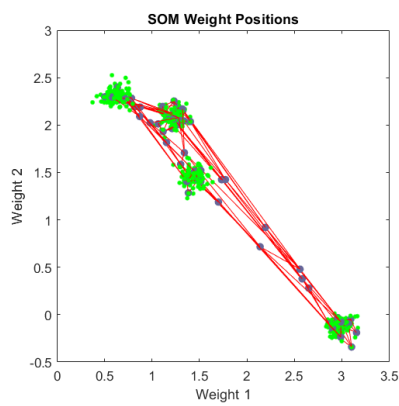
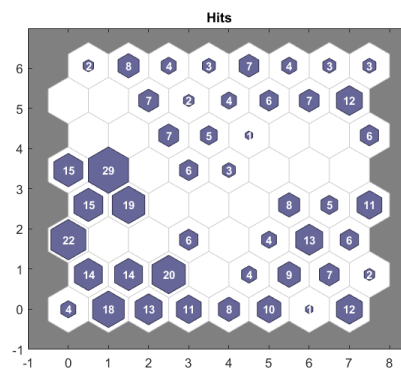
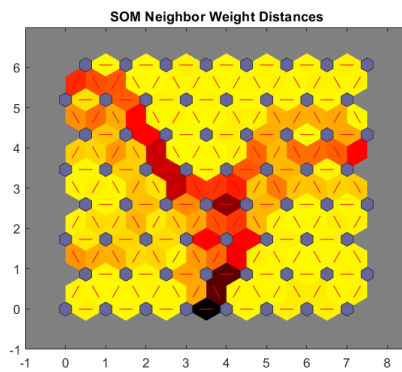


En este caso se ve bastante claro cómo va la distribución de los distintos clusters, si miramos la gráfica segunda, pero como tenemos la primera distribuida de esta forma en la cual todas las neuronas están relativamente cerca unas de otras, no podríamos viendo esa imagen intuir claramente cómo va a ser su distribución.

A su favor podemos decir que los clusters no están las neuronas muy distribuidas, sino que hay menos neuronas en cada cluster con más peso por cada neurona. Por tanto, esto va a ser un aspecto positivo de esta red.

Los espacios de neuronas en blanco se corresponden con las neuronas por las cuales no pasa nada de información.

Caso 9 ([8 8],100,3,'hextop','linkdist')

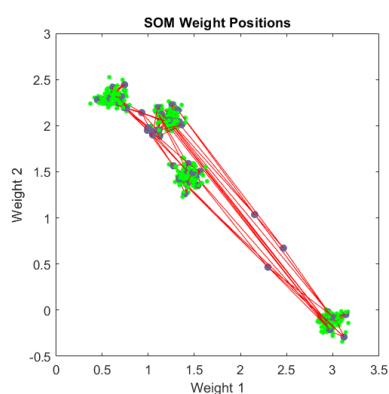
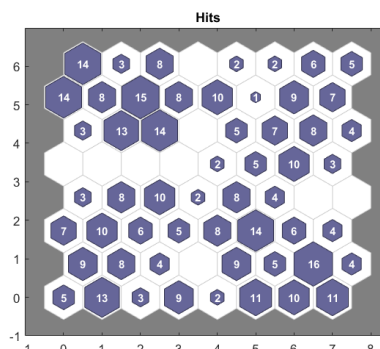
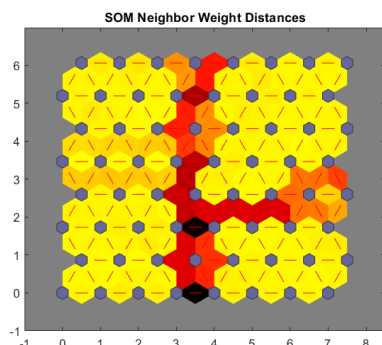


En este caso se puede intuir un poco como va la distribución de los clusters, pero se puede ver que hay dos de ellos con muchas neuronas dentro que están muy juntos, por tanto, hay una ligera dificultad para poder saber a cuál se corresponde cada neurona.

Este es uno de los motivos por los cuales este caso no es uno de los más óptimos a elegir entre los mejores de todos los casos.

En este caso y todos los demás las neuronas que aparecen en blanco van a ser las neuronas por las cuales no pasa nada de información a través de ellas.

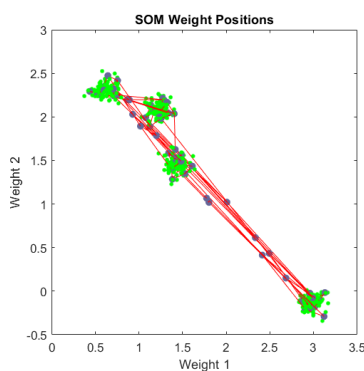
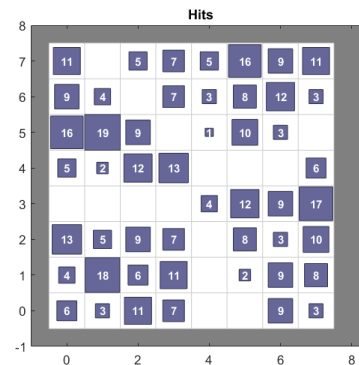
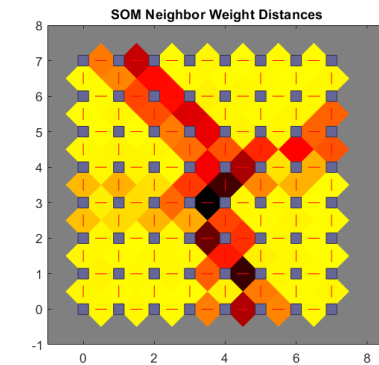
Caso 10 ([8 8],100,3,'hextop','boxdist')



Como podemos ver en estas imágenes es uno de los peores casos que hemos visto en este trabajo ya que podemos ver que hay dos clusters que no están separados del todo entre sí, por tanto, puede costar un poco más su visualización y además hay muchas neuronas repartidas por cada cluster, lo cual no es lo más óptimo.

Hay muy pocas neuronas que queden en blanco en la segunda imagen pero podemos ver que hay algunas las cuales se corresponden con las neuronas que están sueltas sin pertenecer a ningún cluster, que se ve eso en la tercera gráfica.

Caso 11 ([8 8],100,3,'gridtop','linkdist')



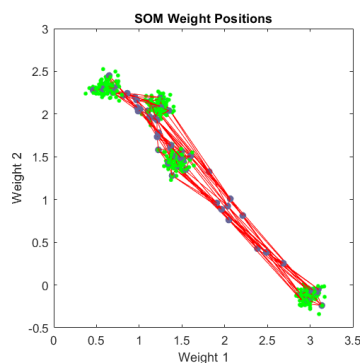
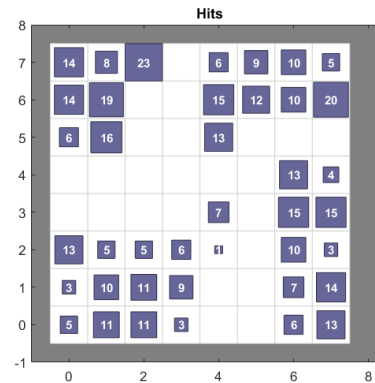
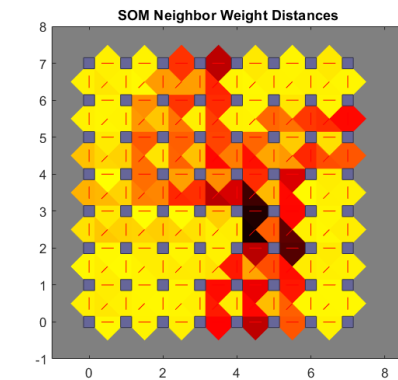
Este caso tampoco es lo mejor de todo porque podemos ver que hay mucha dispersión de neuronas, en las cuales no podemos distinguir exactamente cuál pertenece a cada cluster.

Aun así, si nos fijamos bien en los dos primeros gráficos podemos ver un poco mejor a cuál pertenece cada neurona, pero sí que es verdad que hay dos cluster separados entre ellos por

unos enlaces casi amarillos, lo cual significa que están más cerca, y puede dar lugar a dificultades para distinguir los distintos clusters.

Las neuronas en blanco en la segunda gráfica son las correspondientes a aquellas neuronas que no son de ningún cluster.

Caso 12 ([8 8],100,3,'gridtop','boxdist')

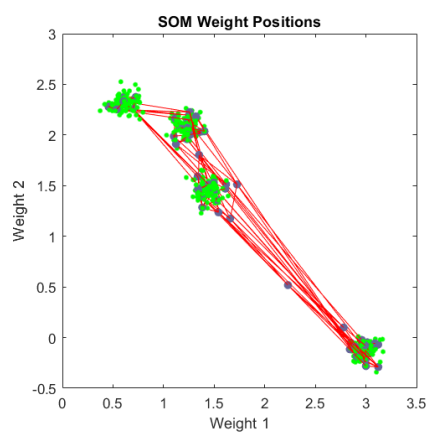
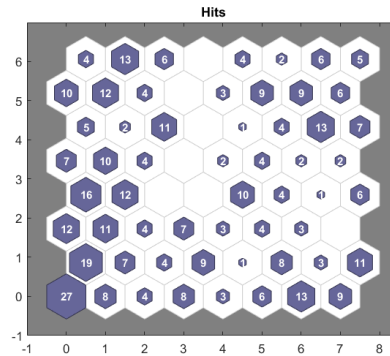
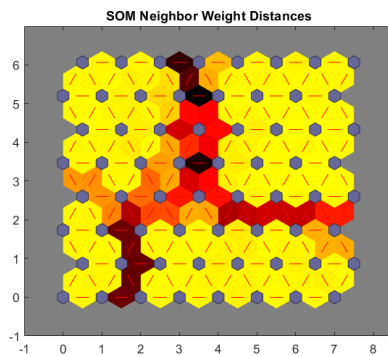


Otro ejemplo como el anterior de un caso que no se ve muy bien a priori viendo la gráfica de distancias cómo iría la representación de los distintos clusters.

Si nos fijamos ya en las dos primeras podemos adivinar con un poco más de exactitud a cuál se corresponde cada neurona, con cada cluster.

Como decimos siempre todos los huecos en blanco que hay en la segunda gráfica van a corresponderse con los sitios donde no hay neuronas dentro de un cluster.

Caso 13 ([8 8],100,6,'hextop','linkdist')

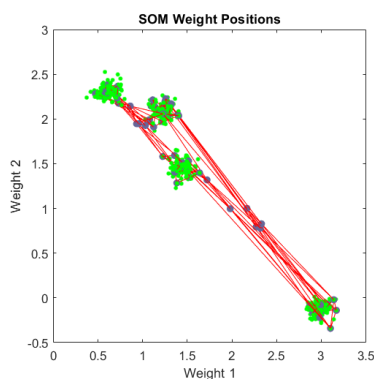
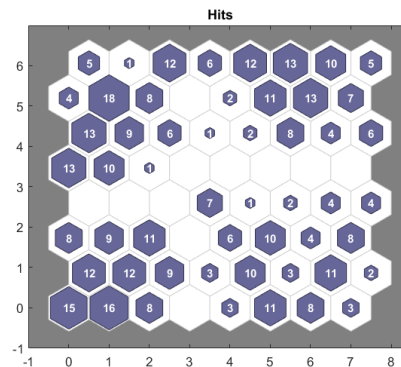
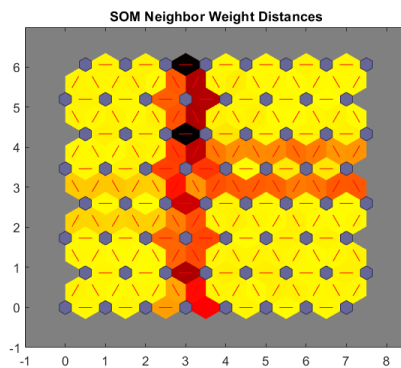


Para poder distinguir los cuatro clusters en este caso sí que se puede saber perfectamente cómo va la distribución de los clusters solo con ver la primera gráfica, por lo tanto, la segunda nos serviría de indicativo para ver luego qué neuronas tienen cada valor.

Los huecos en blanco, que hay pocos, en la segunda gráfica, se corresponden con las neuronas que no pertenecen a ningún

cluster.

Caso 14 ([8 8],100,6,'hextop','boxdist')

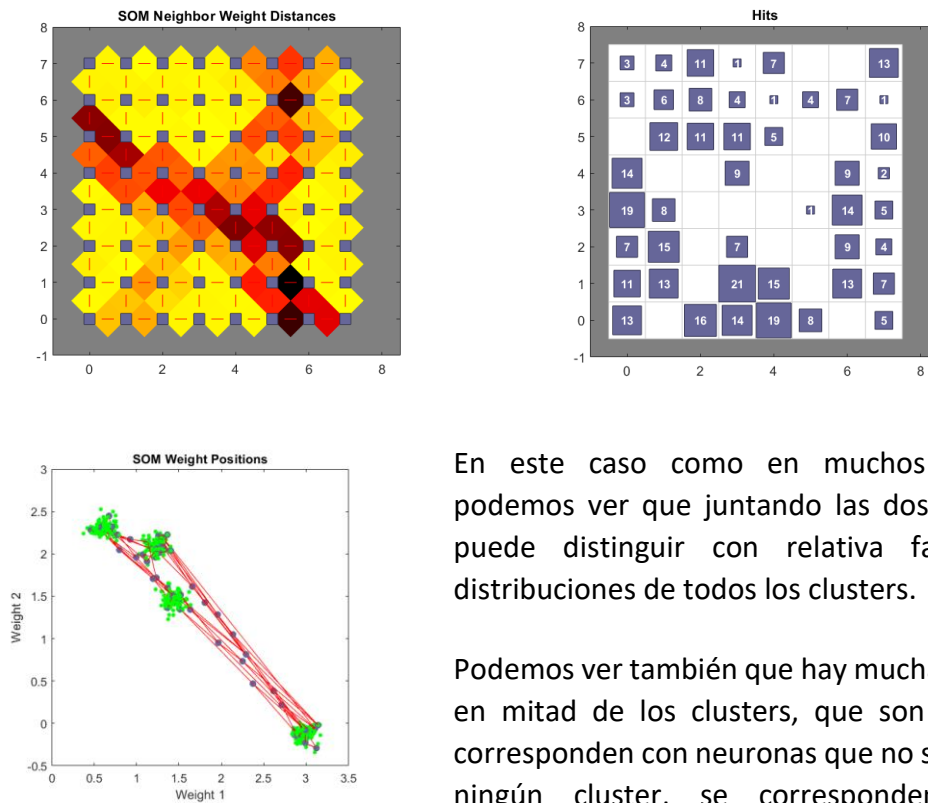


A pesar de que hay dos clusters que no se distingue muy bien con la gráfica de distancias hasta qué punto llega cada uno, si juntamos los dos primeros gráficos, podemos ver bastante bien la representación de los cuatro clusters.

Además, las neuronas dentro de cada cluster tampoco está el peso repartido entre un gran número de neuronas, por tanto, puede ser un ejemplo bastante óptimo.

Todas las neuronas en blanco en la segunda gráfica son las cuales no están dentro de ningún cluster.

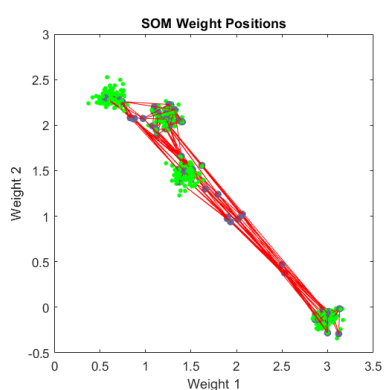
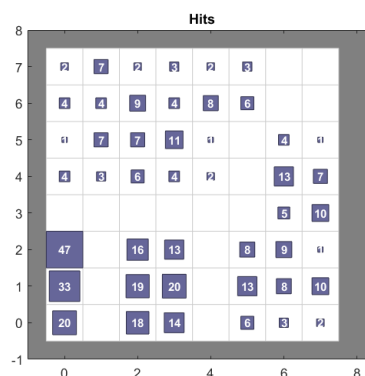
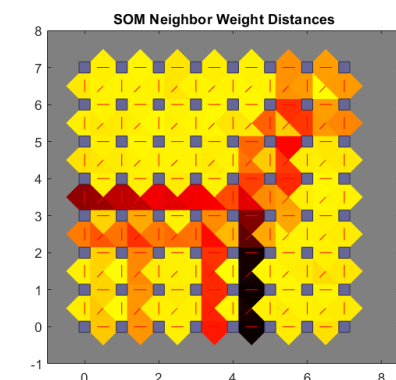
Caso 15 ([8 8],100,6,'gridtop','linkdist')



En este caso como en muchos anteriores podemos ver que juntando las dos gráficas se puede distinguir con relativa facilidad las distribuciones de todos los clusters.

Podemos ver también que hay muchas neuronas en mitad de los clusters, que son las que se corresponden con neuronas que no son parte de ningún cluster, se corresponden con las neuronas que en la segunda gráfica están en blanco.

Caso 16 ([8 8],100,6,'gridtop','boxdist')



En esta gráfica podemos ver como hay algunas neuronas bien distinguidas dentro de sus clusters pero vemos que hay otros con un gran número de neuronas y muy poco peso en cada cluster.

Hay muchos rasgos comunes con el resto de los casos asique tampoco podemos diferenciar de una forma muy clara cuál será la mejor.

Las neuronas en blanco son todas las neuronas que no pertenecen a ningún cluster.



Conclusiones Obtenidas

Como se puede ver de forma general, luego ya lo veremos en las conclusiones de forma más particular, no podemos elegir a priori, cuál va a ser lo mejor que podemos seleccionar.

Pero aun así podemos mencionar que lo mejor va a ser elegir una red en la que se distingan con claridad los clusters que tenemos y que tampoco esté muy dispersas las neuronas con sus correspondientes pesos, es mejor que esté todo más concentrado.

Una vez obtenidos todos los datos ya, y, teniendo en cuenta los gráficos que hemos sacado podemos hacernos una idea o intentar valorar cual puede ser la mejor o las mejores opciones que tener en cuenta y cuáles no van a ser tan buenas.

Vamos a decir que lo mejor para poder seleccionar una de los casos o varios casos va a ser que se puedan distinguir bien los clusters que vamos a tener, en los casos en los que quizá hay mayor dificultad para saber qué neuronas pertenecen a cada cluster, van a ser desechados, debido a que no son los más óptimos.

También debemos añadir que los casos en los cuales hay muchas neuronas en un cluster y por tanto se divide mucho el peso entre todas las neuronas, tampoco van a ser del todo óptimos, de esta forma vamos a rechazar también algunos otros casos.

Dicho esto, hemos considerado como casos favorables los siguientes:

- Caso 1.
- Caso 4.
- Caso 13.
- Caso 16.

Los dos primeros casos son con una rejilla bidimensional más pequeña, con esta rejilla más pequeña es más fácil localizar los distintos clusters, pero hemos considerado que en estos dos ejemplos es donde mejor se ve.

Después vamos a tener los casos 13 y 16 ya que nos parecen también en los que mejor se ve de manera individual y colectiva en los gráficos que neuronas pertenecen a cada cluster. También hemos elegido dichos casos porque también son posiblemente de los que no hay un gran número de neuronas dividiendo demasiado el peso en cada cluster.

Aun así repetimos que no se puede escoger con exactitud uno de los casos considerándolo el mejor, sino que es un poco subjetivo y hemos elegido los que nosotros hemos creído más correctos.