

Aprendizaje Automático

PRUEBA DE EVALUACIÓN A DISTANCIA

Mapas Auto-Organizados de Kohonen

Alumno: Diego Díaz

Contenidos

Introducción.....	1
Apartado 4.	2
4.1. Esquema de la red de neuronas.....	2
4.2. Representación en el eje de coordenadas de los ejemplos.....	2
4.3 Representación en el eje de coordenadas de las neuronas.	3
4.4. Conclusiones de la distribución de neuronas.	3
Apartado 5.	3
5.1. Representación en el eje de coordenadas de los ejemplos.....	3
5.2. Representación en el eje de coordenadas de las neuronas.	4
5.3. Conclusiones de la distribución de neuronas.	4
Apartado 6.	4
6.1. Representación en el eje de coordenadas de las neuronas.	4
6.2. Representación en el eje de coordenadas de los ejemplos.....	5
6.3. Conclusiones de la distribución de neuronas.	5
Apartado 7.	6
7.1. Matriz de distancias.	6
Apartado 8.	6
8.1. Ventajas.	7
8.2. Inconvenientes.....	7
8.3. Comparación entre mapas auto-organizados y k-medias.	8
Conclusiones.	8

Introducción.

En este documento se encuentran las respuestas a la segunda prueba de evaluación a distancia (PED 2) de la asignatura *Aprendizaje Automático* de la UNED del curso 2015-2016.

El enunciado de las preguntas se encuentra en el documento *Practica_MapasAutoOrganizados.pdf*. Dicho enunciado no se incluye en este documento.

Apartado 4.

El fichero que se ha utilizado es "Distrib-Uniforme-2D.arff".

4.1. Esquema de la red de neuronas.

En la **Figura 1** se muestra un esquema de arquitectura del mapa de Kohonen. Se representan las dos entradas E1,E2 y las nueve neuronas de la capa de salida (neuronas de la capa de competición numeradas). Se etiqueta el valor de los pesos μ_{14} y μ_{24} que corresponde con el valor "value" en la tabla de resultados de acuerdo al modelo obtenido, es decir, los pesos de las conexiones que van desde las dos entradas a la neurona $n=4$.

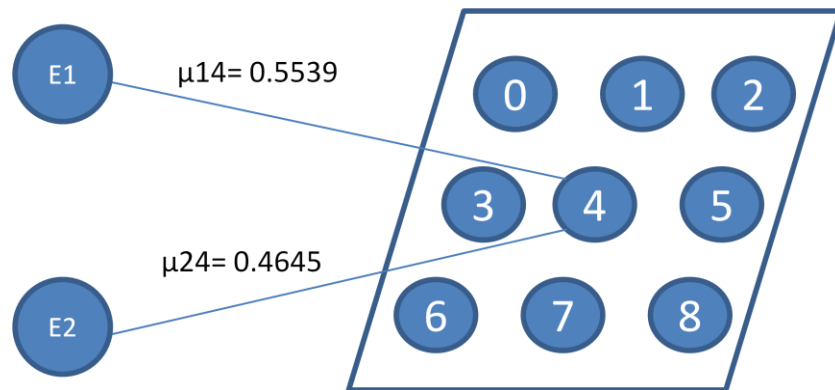


Figura 1. Esquema de representación del modelo obtenido mostrando los pesos de las conexiones con la neurona 4.

4.2. Representación en el eje de coordenadas de los ejemplos.

Se representa en la **Figura 2** todos los ejemplos del fichero en un campo de dos dimensiones (eje de coordenadas).

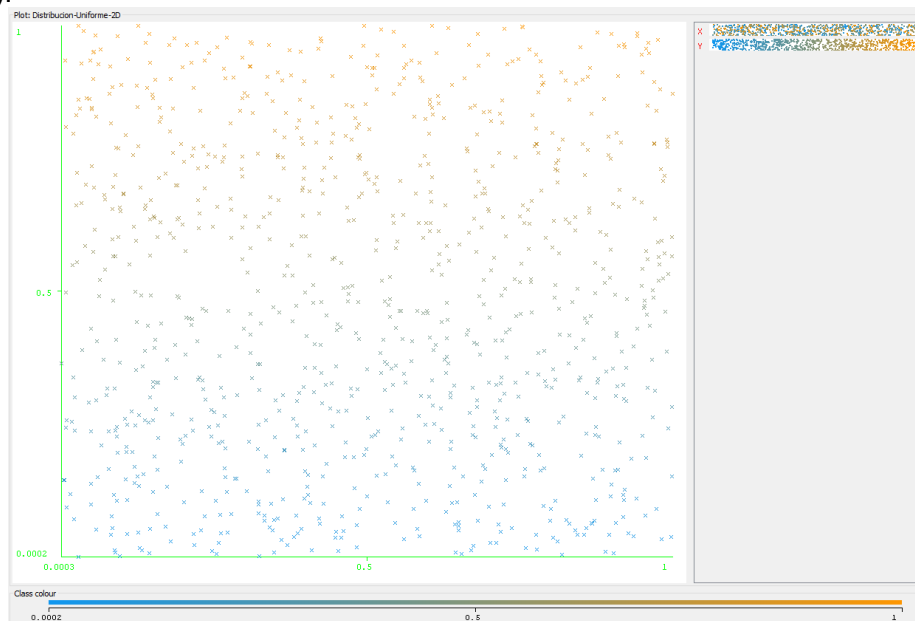


Figura 2. Representación gráfica ejemplos del fichero "Distrib-Uniforme-2D.arff".

4.3 Representación en el eje de coordenadas de las neuronas.

A continuación se representa en la **Figura 3**, las posiciones de cada neurona de acuerdo a los pesos de las conexiones que llegan a cada neurona, según el modelo obtenido en Weka.

Neurona	X	Y
0	0,80	0,17
1	0,86	0,52
2	0,82	0,84
3	0,44	0,16
4	0,55	0,46
5	0,48	0,82
6	0,14	0,19
7	0,18	0,54
8	0,16	0,84

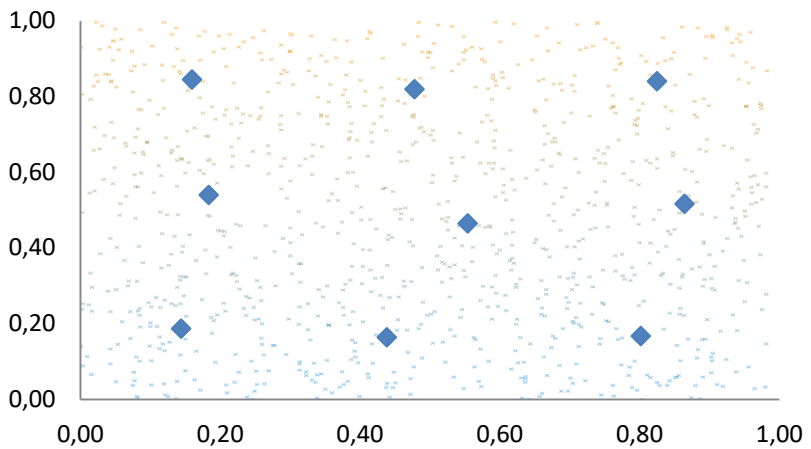


Figura 3. Distribución de las neuronas de acuerdo a los pesos de las conexiones de cada una.

4.4. Conclusiones de la distribución de neuronas.

Después de las iteraciones, la malla de neuronas tiende a aproximarse a la distribución de datos como puede observarse en la gráfica de la **Figura 3**. Es por este motivo que se puede afirmar que la distribución de las neuronas es coherente con el tipo de distribución de datos utilizado para aprender el modelo.

Apartado 5.

El fichero que se ha utilizado es "Distrib-Normal-2D.arff".

5.1. Representación en el eje de coordenadas de los ejemplos.

Se representa en la **Figura 4** todos los ejemplos del fichero en un campo de dos dimensiones (eje de coordenadas).

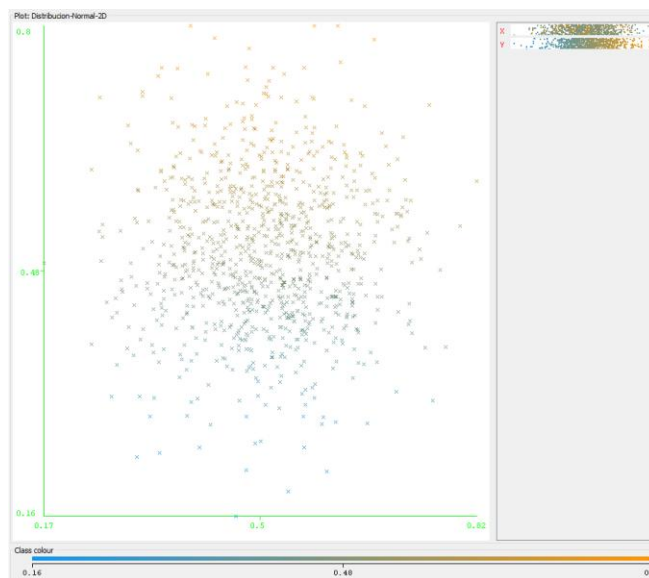


Figura 4. Representación gráfica fichero "Distrib-Normal-2D.arff".

5.2. Representación en el eje de coordenadas de las neuronas.

A continuación se representa en la **Figura 5**, las posiciones de cada neurona de acuerdo a los pesos de las conexiones que llegan a cada neurona, según el modelo obtenido en Weka.

Neurona	X	Y
0	0,3459	0,4083
1	0,4846	0,4397
2	0,519	0,3193
3	0,3916	0,5379
4	0,5183	0,5431
5	0,606	0,4291
6	0,4371	0,6662
7	0,5662	0,6447
8	0,6605	0,5507

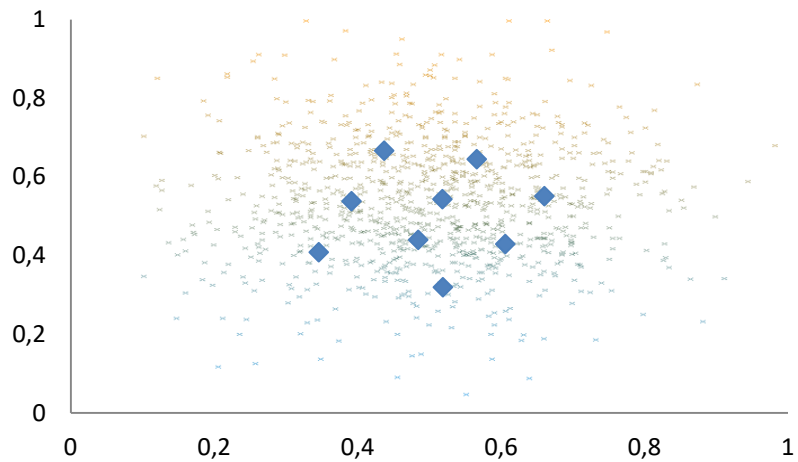


Figura 5. Distribución de las neuronas de acuerdo a los pesos de las conexiones de cada una.

5.3. Conclusiones de la distribución de neuronas.

Después de las iteraciones, la malla de neuronas tiende a aproximarse a la distribución de datos como puede observarse en la gráfica de la **Figura 5**. La zona de alta densidad ejerce un efecto de atracción en la distribución de las neuronas, ya que el peso inicial que se aplica a cada conexión de las neuronas es mayor en los primeros ejemplos, por lo que sí es coherente el resultado. Se podrían probar diferentes reordenaciones de los datos o suavizar el decremento que se aplica en los pesos de las conexiones con las neuronas para que la malla abarcara también los puntos periféricos más dispersos.

Apartado 6.

El fichero que se ha utilizado es "Datos_2D.arff".

6.1. Representación en el eje de coordenadas de las neuronas.

A continuación se representa en la **Figura 6**, las posiciones de cada neurona de acuerdo a los pesos de las conexiones que llegan a cada neurona, según el modelo obtenido en Weka.

	0	1	2	3	4	5	6	7	8	9	10	11	12
X	0,8816	0,7863	0,5625	0,8048	0,9812	1,0696	0,9901	1,0350	1,3061	1,5813	1,3580	1,2348	1,6960
Y	0,5897	0,8595	1,1095	1,1415	1,4001	0,7430	0,9618	1,1447	1,2612	2,0975	0,8069	1,0104	1,8439

	13	14	15	16	17	18	19	20	21	22	23	24
X	1,8410	1,7643	1,8625	1,8491	1,9779	1,9737	2,0410	2,1113	2,2286	2,4533	2,1135	2,3348
Y	2,0274	2,3232	1,4232	1,6655	1,8923	2,1878	2,4218	1,5893	1,8181	2,0005	2,0621	2,2961

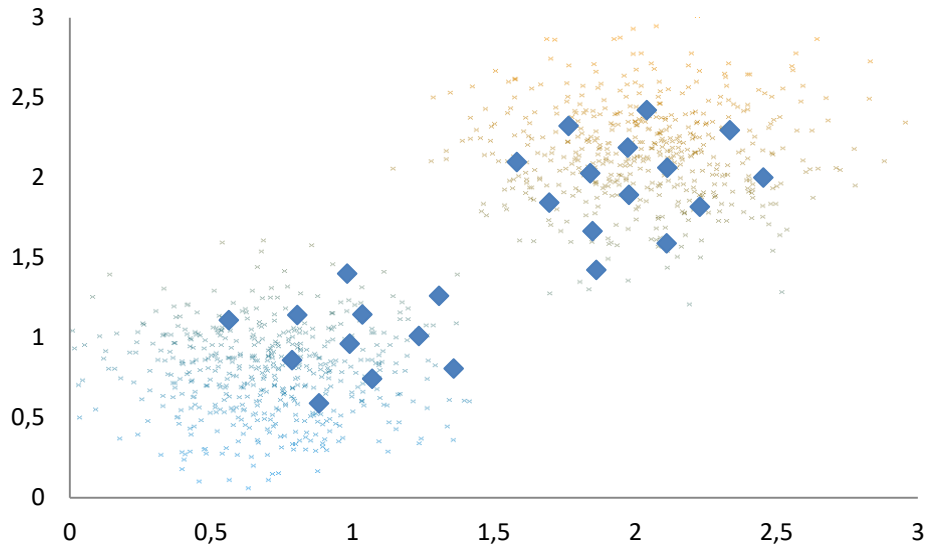


Figura 6. Distribución de las neuronas de acuerdo a los pesos de las conexiones de cada una.

6.2. Representación en el eje de coordenadas de los ejemplos.

Se representa en la **Figura 7** todos los ejemplos del fichero en un campo de dos dimensiones (eje de coordenadas).

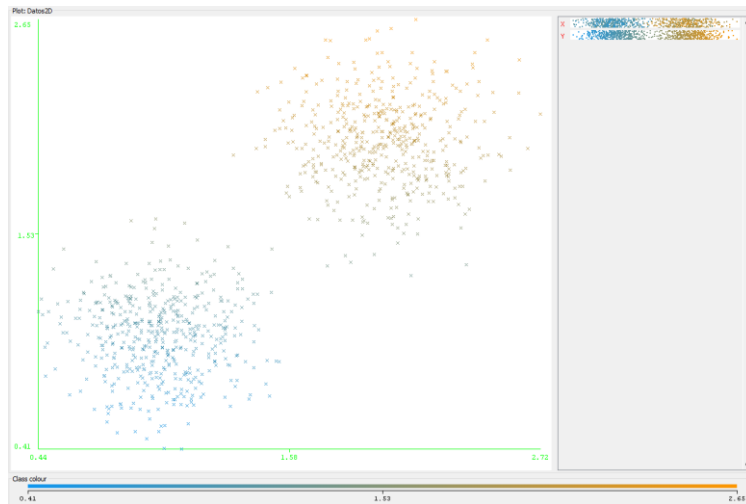


Figura 7. Representación gráfica fichero "Datos-2D.arff".

6.3. Conclusiones de la distribución de neuronas.

En la **Figura 6** se puede observar que incluso partiendo de dos núcleos bien diferenciados en el espacio de ejemplos de entrenamiento, se crean dos mallas que tienden a aproximarse a la distribución de datos. En este último experimento se puede concluir por tanto que los mapas auto-organizados se pueden utilizar para descubrir 2 agrupaciones (clusters) en un conjunto de datos.

Apartado 7.

El fichero que se ha utilizado es "Datos_4D.arff".

7.1. Matriz de distancias.

Se calcula la distancia euclídea del vector de pesos de cada neurona con el vector de pesos de sus vecinas; obtenemos una matriz de distancias entre neuronas que se observa en la **Tabla 1**.

La fórmula de la distancia euclídea:

$$dE(N_i, N_j) = \sqrt{(\mu_{1i} - \mu_{1j})^2 + (\mu_{2i} - \mu_{2j})^2 + (\mu_{3i} - \mu_{3j})^2 + (\mu_{4i} - \mu_{4j})^2}$$

N_i representa la Neurona i -ésima y $dE(N_i, N_j)$ representa la distancia euclídea del vector de pesos de la neurona N_i ($\mu_{1i}, \mu_{2i}, \mu_{3i}, \mu_{4i}$) y el vector de pesos de la neurona N_j ($\mu_{1j}, \mu_{2j}, \mu_{3j}, \mu_{4j}$).

N0	D0;1 = 0,69	N1	D1;2 = 2,05	N2	D2;3 = 0,48	N3	D3;4 = 0,47	N4
D0;5 = 0,84		D1;6 = 0,49		D2;7 = 0,53		D3;8 = 0,92		D4;9 = 0,49
N5	D5;6 = 0,5	N6	D6;7 = 2,01	N7	D7;8 = 0,49	N8	D8;9 = 0,48	N9
D5;10 = 0,62		D6;11 = 0,58		D7;12 = 0,71		D8;13 = 0,63		D9;14 = 0,63
N10	D10;11 = 0,67	N11	D11;12 = 2,06	N12	D12;13 = 0,58	N13	D13;14 = 0,55	N14
D10;15 = 2,27		D11;16 = 1,66		D12;17 = 1,78		D13;18 = 1,76		D14;19 = 1,72
N15	D15;16 = 0,55	N16	D16;17 = 1,71	N17	D17;18 = 0,61	N18	D18;19 = 0,53	N19
D15;20 = 0,46		D16;21 = 0,60		D17;22 = 0,54		D18;23 = 0,54		D19;24 = 0,62
N20	D20;21 = 0,55	N21	D21;22 = 2,05	N22	D22;23 = 0,69	N23	D23;24 = 0,63	N24

Tabla1. Matriz de distancias entre neuronas vecinas.

Se considera que las neuronas pertenecen al mismo cluster si la distancia entre ellas es menor o igual que 1, se puede observar en la **Tabla 1** que las distancias en rojo delimitan 4 zonas diferenciadas. Estas zonas representan los cuatro nuevos clusters, los cuales quedan agrupados de la siguiente manera:

Cluster 1 = Se agrupan N0, N1, N5, N6, N10, N11

Cluster 2 = Se agrupan N2, N3, N4, N7, N8, N9, N12, N13, N14

Cluster 3 = Se agrupan N15, N16, N20, N21

Cluster 4 = Se agrupan N17, N18, N19, N22, N23, N24

Apartado 8.

En este apartado se realiza una comparación de los mapas auto-organizados y k-medias. Este último es otro método de clasificación no supervisada muy similar al de las redes de Kohonen. Cada subapartado contiene una tabla que compara las similitudes entre ambos. La **Tabla 2** del primer subapartado muestra las ventajas que tienen estos dos métodos y la **Tabla 3** del segundo subapartado, los inconvenientes.¹ En el tercer subapartado, se comparan ambos métodos.

¹ Información resumida del libro "Aprendizaje Automático". Editorial Sanz y Torres

8.1. Ventajas.

Mapas auto-organizados	K-medias
El sistema incorpora el mecanismo del vecindario, el cual permite saber qué prototipos están relacionados entre sí.	Una ventaja que presenta esta técnica es el particionado automático de todo el espacio, dando lugar a lo que se conoce como el espacio de Voronoi (también basados en la distancia euclidiana).
Ruido en entrada: no se ven considerablemente afectados por errores en los ejemplos de entrenamiento, o en los datos de entrada. La red es capaz de eliminar el ruido tanto en aprendizaje, como en la clasificación de nuevos ejemplos.	Ruido en entradas: no trata el ruido explícitamente, pero al incluir cada ejemplo en la clase más parecida, se suaviza bastante el ruido.
Ruido en estructura: la pérdida de células en la capa de competición no es demasiado crítica, sobre todo si hay un número elevado de células en dicha capa. La eliminación de células en la entrada, si la célula no es demasiado crítica, su desaparición no afectará demasiado al proceso de clasificación.	Ruido en estructura: la desaparición de una clase afecta sólo a los ejemplos futuros que fueran más parecidos a esa clase, aunque siempre se podrán clasificar en la siguiente clase más parecida.
La fiabilidad de la red dependerá de la fiabilidad del conjunto de ejemplos.	Depende de dónde se coloquen los centroides iniciales, pero, en general, se comporta muy bien aun cuando los ejemplos tengan ruido.
Los mapas auto-organizados de Kohonen pueden convertirse en modelos supervisados introduciendo algunas modificaciones en el modelo. Esta variante se conoce como modelo LVQ, de su nombre en inglés <i>Learning Vector Quantization</i> .	Alguna variante de esta estrategia puede generar clases no disjuntas [Everitt, 1981], lo que supone que las clases compartan algunos ejemplos.

Tabla 2. Ventajas en mapas auto-organizados y k-medias.

8.2. Inconvenientes.

Mapas auto-organizados	K-medias
El inconveniente es que el orden en que los ejemplos son mostrados a la red es relevante, introduciendo un fuerte sesgo. Para solucionar este sesgo, después de cada presentación completa del conjunto de ejemplos, se reordenan de forma aleatoria, para que en la siguiente iteración los ejemplos se presenten en un orden diferente y aleatorio.	El aprendizaje se realiza en función de todos los ejemplos que caen en la misma clase y no se realiza para cada ejemplo de forma específica. Esto no permite adaptaciones más suaves y más apropiadas cuando los ejemplos están más dispersos.
El número de categorías se convierte en una especie de umbral que determina lo parecidos que deben ser los ejemplos para que pertenezcan a la misma categoría. Por lo tanto el número de categorías iniciales condicionará fuertemente el resultado final de la red.	Tener que proporcionar "a priori" el número de clases en las que se desea partir los ejemplos.
No existe la posibilidad de evaluar el aprendizaje realizado, en fases posteriores.	No se realiza ningún control de la tarea aprendida.
Sin redundancia sería imposible encontrar patrones o características en los datos.	

Tabla 3. Inconvenientes en mapas auto-organizados y k-medias.

8.3. Comparación entre mapas auto-organizados y k-medias.

Comparando desde un punto de vista teórico, ambos métodos serían efectivos en los experimentos de los apartados 4, 5 y 7. Sería interesante comprobar desde un punto de vista práctico que efectivamente la eliminación de entradas o de células en la capa de competición no resulta crítica o la utilización de algún algoritmo para la construcción del Diagrama de Voronói. En referencia al experimento del apartado 6, la selección de los centroides iniciales podría influir negativamente a la hora de conseguir el objetivo al menos que se disponga de información previa de la agrupación de los datos de entrenamiento.

Desde el punto de vista de cálculo, en los mapas auto-organizados, una vez determinada la célula ganadora, se modifican los pesos de las conexiones entre dicha neurona y las de entrada mediante una ecuación que varía según el valor de una tasa de aprendizaje α la cual es decreciente con el tiempo según un decremento con valor constante o logarítmico; esto implica que se da más importancia a las primeras iteraciones que son las que tienden a formar la estructura de la red y después se ajusta con ligeros desplazamientos hasta alcanzar un equilibrio. Valores de α grandes harán que la neurona, después del desplazamiento, se sitúe muy cerca del ejemplo, mientras que valores de α muy pequeños harán que la neurona apenas se desplace. Es por este motivo que el orden en los que los ejemplos son mostrados a la red es relevante, como ya se señaló anteriormente. Otro aspecto relacionado con la tasa de aprendizaje es que puede influir en el tiempo de cálculo. Según el método del cálculo de la tasa α , el tiempo de ejecución podrá verse afectado en más medida que el algoritmo de k-medias.

Conclusiones.

Se han realizado cuatro experimentos, en donde se puede observar de una manera gráfica, las diferentes formas de usar un mapa auto-organizado. Esto ha resultado interesante ya que la interpretación de datos desde un punto de vista porcentual no es siempre sencilla de realizar.

En los dos primeros experimentos, apartados cuatro y cinco respectivamente, se aprende una representación discreta de la distribución de un conjunto de datos colocados tanto en una distribución uniforme como en una distribución normalizada. Se observa cómo las neuronas de la capa de competición de un mapa auto-organizado son capaces de reacomodarse en el plano X-Y para ajustarse a la distribución del conjunto de datos de entrenamiento.

El tercer experimento, que corresponde al apartado seis, se utiliza para descubrir dos agrupaciones (clusters) en un conjunto de datos. Resulta interesante destacar que de las veinticinco neuronas de la capa de competición se han repartido casi al cincuenta por ciento (cuarenta y cuatro por ciento en un grupo y cincuenta y seis por ciento en el otro grupo) lo cual tiene sentido teniendo en cuenta de que el tamaño en número de ejemplos de entrenamiento es similar.

En el cuarto experimento, que se presenta en el apartado siete, se aprende la manera de reducir la dimensionalidad de un conjunto de ejemplos de entrenamiento. Es interesante el poder ver la agrupación de los nuevos clusters visualmente en una tabla donde se muestran las neuronas y las distancias euclidianas del vector de pesos entre una neurona y los vectores de pesos de sus vecinas.

Conclusión final: ya que los resultados obtenidos han sido coherentes, se puede afirmar que si se dispone de una cantidad de datos grande con la suficiente redundancia, el método de mapa auto-organizado cumple los objetivos planteados.