

Escuela Politécnica Superior Campus Colmenarejo
Universidad Carlos III de Madrid

Problema de clasificación Vehicle silhouettes

REDES DE NEURONAS ARTIFICIALES

Grado en Ingeniería Informática

Autor: Luis Cabrero García

Profesor: José María Valls Ferrán

Curso 4º - Gr80

Índice general

Índice general	III
Índice de figuras	V
Índice de tablas	V

1 Introducción	1
1.1 Motivación	1
1.2 Objetivos	1
1.3 Estructura de la memoria	1
2 Preparación de los datos	3
2.1 Clasificación de los datos	3
2.2 Preparación de los conjuntos de entrenamiento y test	3
2.3 Normalización de los conjuntos de entrenamiento y test	3
3 Perceptrón Multicapa	5
3.1 Perceptron Multicapa (MLP)	5
3.2 Script en RSNNS	6
3.3 Experimentación realizada	7
4 Learning Vector Quantization	15
4.1 Learning Vector Quantization (LVQ)	15
4.2 Experimentación realizada	15
5 Mapas autoorganizados de Kohonen	27
5.1 Kohonen (SOM)	27
5.2 Experimentación realizada	27
6 Conclusiones	31
Bibliografía	33

Índice de figuras

3.1	Red Perceptrón multicapa [2]	5
3.2	Esquema MLP [2]	6
3.3	MLP $\gamma = 0,1$, neuronas = 100 y ciclos = 10000 pareja 1	7
3.4	MLP $\gamma = 0,1$, neuronas = 100 y ciclos = 10000 pareja 2	7
3.5	MLP $\gamma = 0,1$, neuronas = 100 y ciclos = 10000 pareja 3	8
3.6	MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 1000 pareja 1	8
3.7	MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 1000 pareja 2	8
3.8	MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 10000 pareja 3	9
3.9	MLP $\gamma = 0,01$, neuronas = 10,10,5 y ciclos = 1000 pareja 1	9
3.10	MLP $\gamma = 0,01$, neuronas = 10,10,5 y ciclos = 1000 pareja 2	10
3.11	MLP $\gamma = 0,01$, neuronas = 10,10,5 y ciclos = 1000 pareja 3	10
3.12	MLP $\gamma = 0,1$, neuronas = 100,50,25 y ciclos = 1000 pareja 1	10
3.13	MLP $\gamma = 0,1$, neuronas = 100,50,25 y ciclos = 1000 pareja 2	11
3.14	MLP $\gamma = 0,1$, neuronas = 100,50,25 y ciclos = 1000 pareja 3	11
3.15	MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000 pareja 1	11
3.16	MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000 pareja 2	12
3.17	MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000 pareja 3	12
3.18	MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000 pareja 1	12
3.19	MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000 pareja 2	13
3.20	MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000 pareja 3	13
4.1	LVQ 20prototipos pareja 1	16
4.2	LVQ 20prototipos pareja 2	16
4.3	LVQ 20prototipos pareja 3	17
4.4	LVQ 50prototipos pareja 1	18
4.5	LVQ 50prototipos pareja 2	18
4.6	LVQ 50prototipos pareja 3	19
4.7	LVQ 100prototipos pareja 1	19
4.8	LVQ 100prototipos pareja 2	20
4.9	LVQ 100prototipos pareja 3	20
4.10	LVQ 180prototipos pareja 1	21
4.11	LVQ 180prototipos pareja 2	22
4.12	LVQ 180prototipos pareja 3	22
4.13	LVQ 196prototipos pareja 1	23
4.14	LVQ 196prototipos pareja 2	24
4.15	LVQ 196prototipos pareja 3	25
5.1	SOM $x = 15y = 15$ pareja 1	29
5.2	SOM $x = 15y = 15$ pareja 2	30
5.3	SOM $x = 15y = 15$ pareja 3	30

Índice de tablas

3.1	MLP $\gamma = 0,1$, <i>neuronas</i> = 100 y <i>ciclos</i> = 10000	7
3.2	MLP $\gamma = 0,1$, <i>neuronas</i> = 100,50 y <i>ciclos</i> = 1000	8
3.3	MLP $\gamma = 0,01$, <i>neuronas</i> = 10,10,5 y <i>ciclos</i> = 1000	9
3.4	MLP $\gamma = 0,1$, <i>neuronas</i> = 100,50,25 y <i>ciclos</i> = 1000	10
3.5	MLP $\gamma = 0,1$, <i>neuronas</i> = 100,50,25,12 y <i>ciclos</i> = 1000	11
3.6	MLP $\gamma = 0,8$, <i>neuronas</i> = 100 y <i>ciclos</i> = 1000	12
4.1	LVQ 20 <i>prototipos</i>	15
4.2	LVQ 50 <i>prototipos</i>	17
4.3	LVQ 100 <i>prototipos</i>	19
4.4	LVQ 180 <i>prototipos</i>	21
4.5	LVQ 196 <i>prototipos</i>	23
5.1	SOM $x = 1y = 10$	28
5.2	SOM $x = 10y = 1$	28
5.3	SOM $x = 10y = 10$	28
5.4	SOM $x = 15y = 15$	29

CAPÍTULO 1

Introducción

La presente práctica trata de abordar mediante el punto de vista de las redes de neuronas supervisadas y no supervisadas la resolución de un problema real: la clasificación de determinadas siluetas de vehículos en cuatro tipos dados. Estos datos han sido tomados del KEEL[1].

1.1 Motivación

La utilidad de los modelos de clasificación que vamos a aplicar de forma práctica es reseñable: nos permiten determinar la categoría a la que pertenece un vehículo determinado a partir de sus características, resultando de gran interés y aplicable a problemas reales.

1.2 Objetivos

Los objetivos de la práctica son: procesar y preparar los datos del KEEL[1] para validación cruzada estratificada, clasificar los datos utilizando el Perceptrón Multicapa, clasificar los datos utilizando LVQ, y por último, utilizando mapas autoorganizados de Kohonen. Posteriormente se hará una comparación entre los tres modelos para ver en que casos unos modelos pueden clasificar mejor que otros.

1.3 Estructura de la memoria

La memoria dispone de tres índices, uno para consultar cada una de las secciones y otros dos para figuras y tablas. En la memoria se puede encontrar tres apartados generales para el modelo MLP, LVQ y mapas de Kohonen. Dentro de los respectivos apartados de cada modelo, se explica la base teórica de cada uno de ellos, se explican los distintos experimentos realizados y se comparan los resultados en función de los parámetros escogidos. En las conclusiones se explican las diferencias entre los modelos y se comparan los resultados obtenidos en la experimentación. Al final de la memoria se indican las referencias bibliográficas consultadas para llevar a cabo el trabajo.

CAPÍTULO 2

Preparación de los datos

2.1 Clasificación de los datos

En este capítulo se va a explicar la preparación de los datos que se ha llevado a cabo para poder posteriormente aplicar los diferentes modelos.

Lo primero que se hace con los datos es separarlos en cuatro hojas, cada una de ellas correspondiente con cada una de las clases existentes: bus, van, saab y opel. Posteriormente lo que se hace es formar tres conjuntos de datos: P1, P2 y P3, que serán otras tres hojas de datos. En cada uno de los tres conjuntos de datos se introducen una tercera parte de los datos de cada una de las clases.

2.2 Preparación de los conjuntos de entrenamiento y test

Una vez tenemos los datos originales repartidos en los tres conjuntos de datos, lo que debemos hacer es distribuir estos conjuntos para formar 3 parejas de conjuntos de entrenamiento y test. El proceso para formar estos ficheros es ir copiando los datos de los conjuntos P1, P2 y P3, en distintas combinaciones para poder formar los conjuntos.

2.3 Normalización de los conjuntos de entrenamiento y test

Cuando se dispone de los conjuntos de entrenamiento y test lo que se debe hacer es normalizar los datos en función de los valores máximo y mínimo del conjunto original. Para ello se aplica la siguiente fórmula:

$$Valor_{normalizado} = \frac{V_{original} - Valor_{min}}{Valor_{max} - Valor_{min}} \quad (2.1)$$

Una vez hemos normalizado todos los datos de **todos** los conjuntos de entrenamiento y test, debemos desordenar los patrones de entrenamiento para presentárselos a nuestros modelos. Esto lo conseguimos añadiendo una columna aleatoria y ordenando posteriormente por esa columna.

CAPÍTULO 3

Perceptrón Multicapa

En este capítulo se va a explicar en primer lugar el modelo Perceptrón Multicapa, que sirve para solucionar problemas de clasificación, es decir, estableciendo una correspondencia entre un conjunto de datos y un conjunto de clases determinadas. Para poder afrontar el problema necesitamos conocer el número de clases. Si el número de clases es desconocido entonces es necesario utilizar redes de neuronas no supervisadas, que serán abordadas en los próximos capítulos.

3.1 Perceptron Multicapa (MLP)

Es una red de neuronas artificiales formada por múltiples capas (3.1) que resuelve problemas que no son linealmente separables. Además, se ha demostrado que es un aproximador universal, es decir, cualquier función continua en el espacio \mathbb{R}^n puede aproximarse con este modelo.

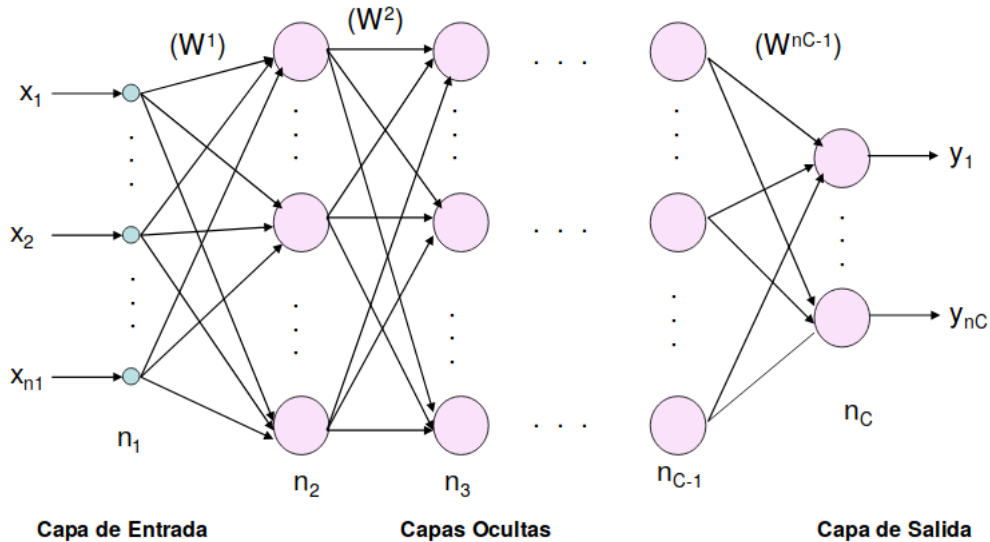


Figura 3.1: Red Perceptrón multicapa [2]

La arquitectura está distribuida en tres tipos de neuronas: neuronas de entrada, que únicamente reciben las entradas y las propagan a la siguiente capa, las neuronas ocultas, que procesan de manera no lineal las entradas, y las neuronas de salida, que devuelven las salidas al exterior.

Teniendo dos neuronas i y j :

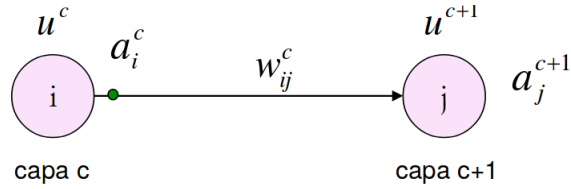


Figura 3.2: Esquema MLP [2]

Los pesos y el umbral (3.1) se expresan en forma matricial, e influyen en la activación de la neurona perteneciente a la capa siguiente. Las neuronas de entrada simplemente tienen su activación equivalente a la entrada que reciben, sin embargo, tanto las neuronas de la capa de salida (3.3) como las neuronas de las capas ocultas (3.2) necesitan computar los pesos y sus entradas sumándoles el umbral.

$$W^c = (w_{ij}^c) = \begin{pmatrix} w_{11}^c & w_{12}^c & \dots & w_{1n_{c+1}}^c \\ w_{21}^c & w_{22}^c & \dots & w_{2n_{c+1}}^c \\ \dots & \dots & \dots & \dots \\ w_{n_c1}^c & w_{n_c2}^c & \dots & w_{n_cn_{c+1}}^c \end{pmatrix}, U^c = (u_i^c) = \begin{pmatrix} u_1^c \\ u_2^c \\ \dots \\ u_{n_c}^c \end{pmatrix} \quad (3.1)$$

$$a_i^c = f\left(\sum_{j=1}^{n_{c-1}} w_{ji}^{c-1} * a_j^{c-1} + u_i^c\right) \quad (3.2)$$

$$y_i = a_i^c = f\left(\sum_{j=1}^{n_{c-1}} w_{ji}^{c-1} * a_j^{c-1} + u_i^c\right) \quad (3.3)$$

Siendo f la función de activación. Las funciones más utilizadas son la función sigmoideal ($f_1(x) = \frac{1}{1+e^{-x}}$) y la tangente hiperbólica ($f_2(x) = \frac{1-e^{-x}}{1+e^{-x}}$). El aprendizaje de la red se realiza de forma semejante que en Adaline [2].

3.2 Script en RSNNS

Para realizar la experimentación se proporciona un script que realiza los siguientes pasos:

- Se cargan dos ficheros de entrenamiento y test que forman una pareja. Posteriormente se cargan los ficheros correspondientes a las otras dos parejas, teniendo de esta manera la validación cruzada.
- Se codifican las cuatro clases (bus, van, saab y opel) en 4 números.
- Se seleccionan los parámetros: razón de aprendizaje, topología de la red y ciclos.
- Se ejecuta el aprendizaje y se selecciona el modelo.
- Se muestra en un gráfico la evolución del error (SSE) según avanzan los ciclos.
- Se generan matrices de confusión que muestran los valores predichos frente a los reales.
- Se halla el porcentaje de acierto.
- Se expresa en una tabla los errores por ciclo.

- Se calcula el error final MSE.
- Se guardan los resultados en ficheros.

3.3 Experimentación realizada

En cuanto a la experimentación realizada se ha prestado especial atención en conseguir una capacidad de generalización adecuada. Esto es: no nos interesa conseguir que la red sea capaz de estimar valores del conjunto de entrenamiento pero que cuando entren datos de un conjunto nuevo, no sea capaz de generalizarlos. Esta situación es conocida como **sobreaprendizaje**. Para cada uno de los experimentos realizados, será de especial interés observar la matriz de confusión para ver las predicciones que hace el modelo frente a los valores reales.

Experimentaciones

$\gamma = 0,1$, $neuronas = 100$ y $ciclos = 10000$

Con una tasa de 0.1, 100 neuronas en una capa oculta y 10000 ciclos tenemos los siguientes porcentajes de acierto:

Tabla 3.1: MLP $\gamma = 0,1$, $neuronas = 100$ y $ciclos = 10000$

Pareja	Porcentaje de acierto entrenamiento	Porcentaje de acierto test
1	0.913357400722022	0.215753424657534
2	0.760907504363002	0.754578754578755
3	0.783595113438045	0.825622775800712
Media	0.81928667333	0.59865165167

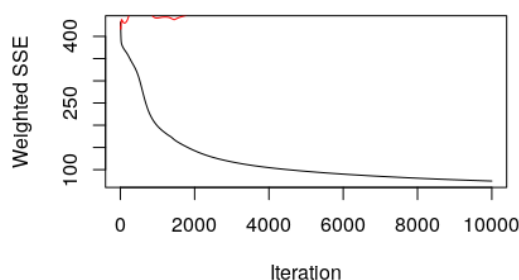


Figura 3.3: MLP $\gamma = 0,1$, $neuronas = 100$ y $ciclos = 10000$ pareja 1

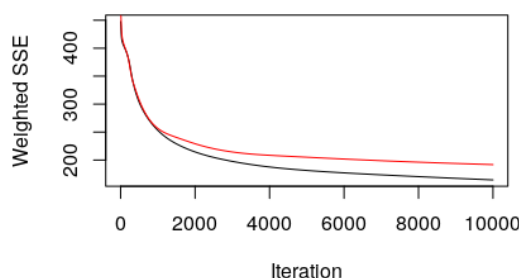


Figura 3.4: MLP $\gamma = 0,1$, $neuronas = 100$ y $ciclos = 10000$ pareja 2

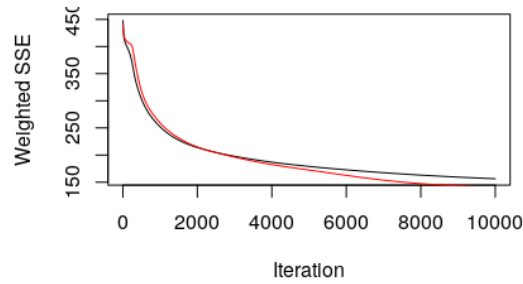


Figura 3.5: MLP $\gamma = 0,1$, neuronas = 100 y ciclos = 10000 pareja 3

Como se puede apreciar en las figuras, los resultados son bastante diferentes, la pareja 2 y 3 se aproxima bastante bien pero la primera pareja presenta resultados muy diferentes. La media del porcentaje de acierto es bastante bueno para entrenamiento pero no muy buena para test. Observando las matrices de confusión, la predicción no es muy buena para varias clases.

$\gamma = 0,1$, neuronas = 100,50 y ciclos = 1000

Con una tasa de 0.1, 100 neuronas y 50 neuronas distribuidas en dos capas ocultas y 10000 ciclos tenemos los siguientes porcentajes de acierto:

Tabla 3.2: MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 1000

Pareja	Porcentaje de acierto entrenamiento	Porcentaje de acierto test
1	0.667870036101083	0.469178082191781
2	0.614310645724258	0.600732600732601
3	0.633507853403141	0.619217081850534
Media	0.63856284666	0.56304258825

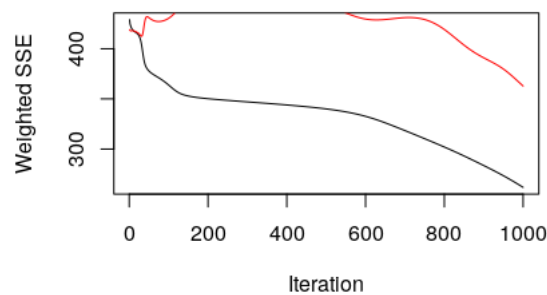


Figura 3.6: MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 1000 pareja 1

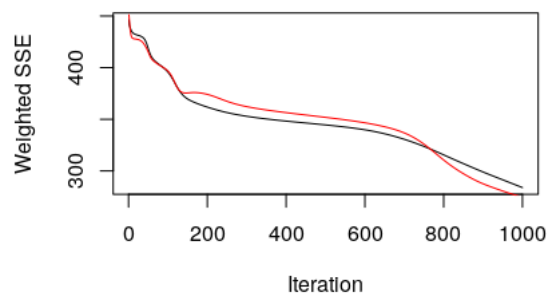


Figura 3.7: MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 1000 pareja 2

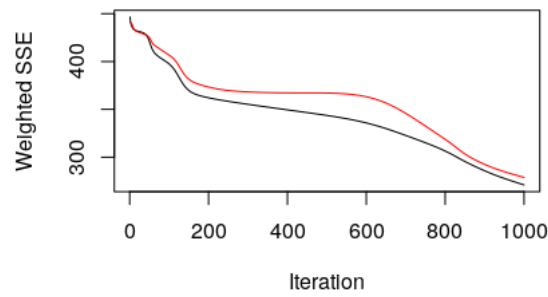


Figura 3.8: MLP $\gamma = 0,1$, neuronas = 100,50 y ciclos = 10000 pareja 3

En comparación con el primer experimento realizado, la diferencia entre el porcentaje de acierto de entrenamiento y test ha disminuido. Aun así, la primera pareja de datos sigue presentando un comportamiento diferente al de las otras dos parejas. Según lo observado en la matriz de confusión se producen menos fallos a la hora de predecir, por tanto considero que es un mejor experimento que el anterior. A pesar de que la media de acierto en el caso de entrenamiento es menor, en caso de test es mayor, se sacrifica acierto en entrenamiento para saber generalizar mejor conjuntos nuevos.

$\gamma = 0,01$, neuronas = 10,10,5 y ciclos = 1000

Con una tasa de 0.01, 10 neuronas, 10 neuronas y 5 neuronas distribuidas en tres capas ocultas y 10000 ciclos tenemos los siguientes porcentajes de acierto:

Tabla 3.3: MLP $\gamma = 0,01$, neuronas = 10,10,5 y ciclos = 1000

Pareja	Porcentaje de acierto entrenamiento	Porcentaje de acierto test
1	0.263537906137184	0.246575342465753
2	0.253054101221641	0.267399267399267
3	0.253054101221641	0.259786476868327
Media	0.25654870286	0.25792036224

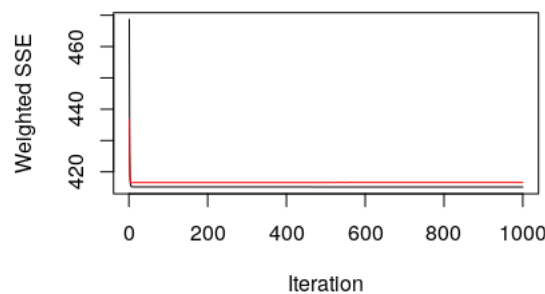


Figura 3.9: MLP $\gamma = 0,01$, neuronas = 10,10,5 y ciclos = 1000 pareja 1

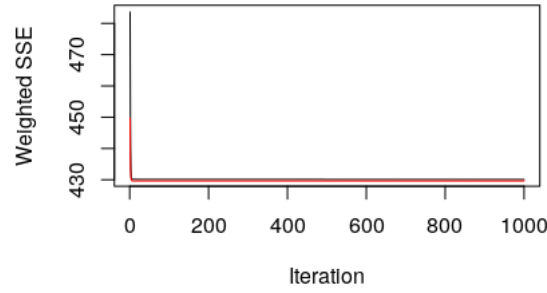


Figura 3.10: MLP $\gamma = 0,01$, neuronas = 10, 10,5 y ciclos = 1000 pareja 2

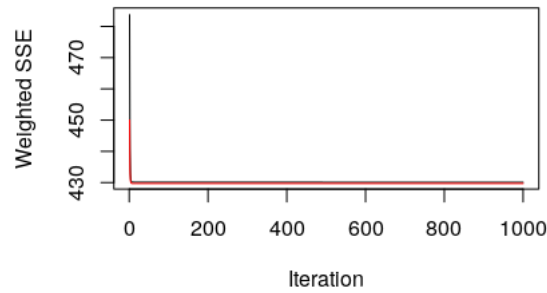


Figura 3.11: MLP $\gamma = 0,01$, neuronas = 10, 10,5 y ciclos = 1000 pareja 3

Como vemos, el resultado en este experimento es que las tres parejas tienen un comportamiento muy similar, el porcentaje de acierto tanto en test como en entrenamiento es muy parecido entre ellas, aunque es demasiado bajo como para considerarlo un buen experimento.

$\gamma = 0,1$, neuronas = 100, 50, 25 y ciclos = 1000

Con una tasa de 0.1, 100 neuronas, 50 neuronas y 25 neuronas distribuidas en tres capas ocultas y 10000 ciclos tenemos los siguientes porcentajes de acierto:

Tabla 3.4: MLP $\gamma = 0,1$, neuronas = 100, 50, 25 y ciclos = 1000

Pareja	Porcentaje de acierto entrenamiento	Porcentaje de acierto test
1	0.465703971119134	0.414383561643836
2	0.56195462478185	0.582417582417582
3	0.363001745200698	0.298932384341637
Media	0.46355344703	0.43191117613

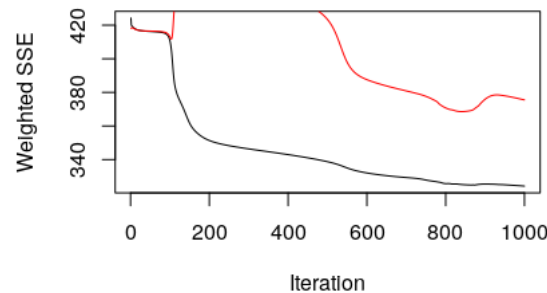


Figura 3.12: MLP $\gamma = 0,1$, neuronas = 100, 50, 25 y ciclos = 1000 pareja 1

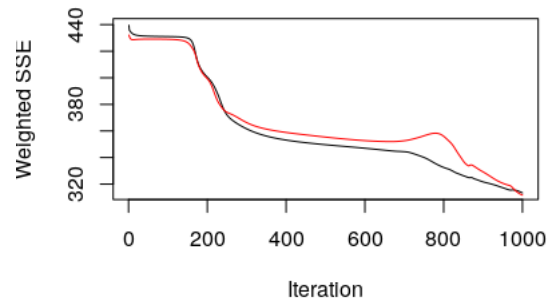


Figura 3.13: MLP $\gamma = 0,1$, neuronas = 100,50,25 y ciclos = 1000 pareja 2

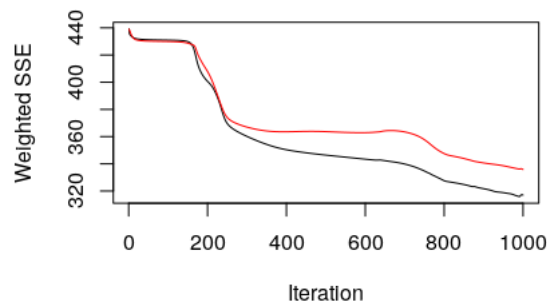


Figura 3.14: MLP $\gamma = 0,1$, neuronas = 100,50,25 y ciclos = 1000 pareja 3

El resultado observado es, de nuevo, que para la primera pareja de datos de entrenamiento y test el error del conjunto de test es más elevado que el de entrenamiento. El porcentaje de acierto es moderado pero estable relativamente.

$\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000

Con una tasa de 0.1, 100 neuronas, 50 neuronas y 25 neuronas distribuidas en tres capas ocultas y 10000 ciclos tenemos los siguientes porcentajes de acierto:

Tabla 3.5: MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000

Pareja	Porcentaje de acierto entrenamiento	Porcentaje de acierto test
1	0.263537906137184	0.246575342465753
2	0.253054101221641	0.267399267399267
3	0.253054101221641	0.259786476868327
Media	0.25654870286	0.25792036224

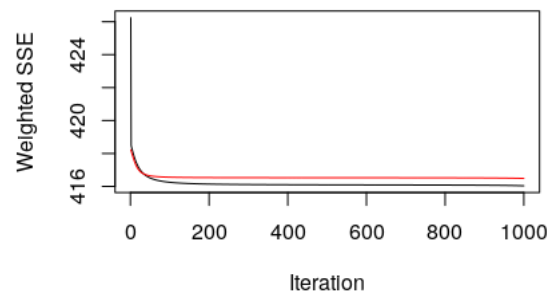


Figura 3.15: MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000 pareja 1

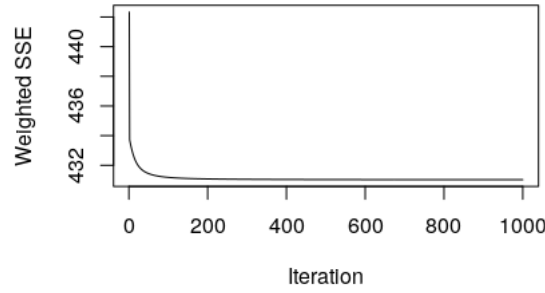


Figura 3.16: MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000 pareja 2

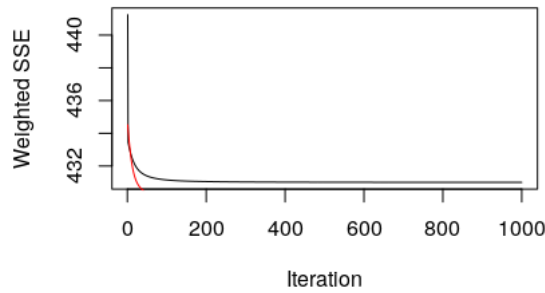


Figura 3.17: MLP $\gamma = 0,1$, neuronas = 100,50,25,12 y ciclos = 1000 pareja 3

Como vemos, el presente experimento presenta tasas de acierto bajas pero estables. Observando la matriz de confusión, se clasifican todos los patrones como uno de ellos, y por tanto el porcentaje de acierto es aproximadamente $1/4$. Este experimento no es apropiado.

$\gamma = 0,8$, neuronas = 100 y ciclos = 1000

Con una tasa de 0.8, 100 neuronas en una capa oculta y 10000 ciclos tenemos los siguientes porcentajes de acierto:

Tabla 3.6: MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000

Pareja	Porcentaje de acierto entrenamiento	Porcentaje de acierto test
1	0.846570397111913	0.424657534246575
2	0.478184991273996	0.47985347985348
3	0.696335078534031	0.651245551601424
Media	0.6736968223	0.5185855219

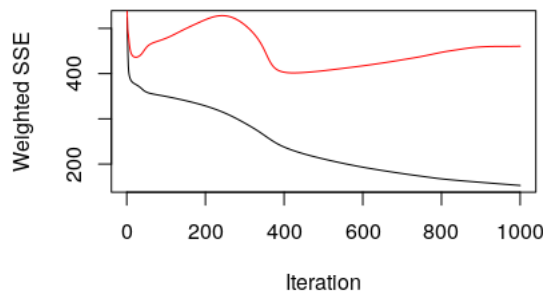


Figura 3.18: MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000 pareja 1

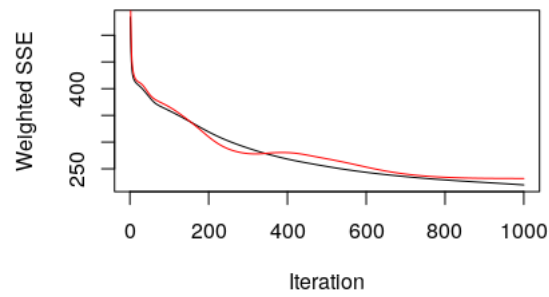


Figura 3.19: MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000 pareja 2

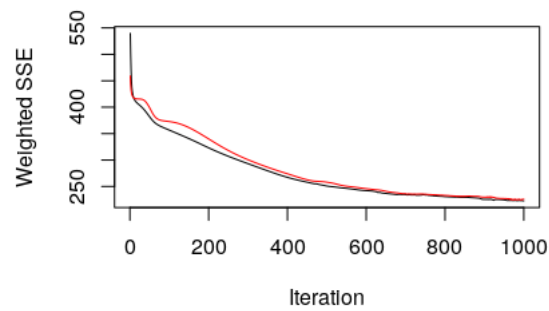


Figura 3.20: MLP $\gamma = 0,8$, neuronas = 100 y ciclos = 1000 pareja 3

Observando la matriz de confusión vemos que las predicciones se aproximan bastante bien a los valores reales, y además los porcentajes de aciertos son buenos. A pesar de todo el segundo experimento sigue ofreciendo mejores resultados.

CAPÍTULO 4

Learning Vector Quantization

En este capítulo se va a explicar en primer lugar el algoritmo LVQ, una versión supervisada del método de Kohonen en la que se conoce de antemano el número de clases, clasificando por vecindad.

4.1 Learning Vector Quantization (LVQ)

Es un algoritmo supervisado de clasificación basado en prototipos. Inicialmente los prototipos se distribuyen aleatoriamente por el espacio de entrada, y una vez que la red se entrena, se obtiene la solución al problema de clasificación que consiste en la colocación final de los prototipos.

El entrenamiento de la red se lleva a cabo de tal forma que inicialmente la distribución aleatoria de los prototipos es alterada por la introducción de los ejemplos de entrenamiento.

Para realizar la clasificación de un patrón del conjunto de test, se introduce el ejemplo, se calcula su distancia a todos los prototipos y por último se etiqueta con la clase del prototipo más cercano.

4.2 Experimentación realizada

En cuanto a la experimentación realizada se ha probado con distintos números de prototipos, escogiendo números proporcionales al número de clases existentes. Se atenderá al porcentaje de acierto realizando la media entre las tres hojas de la validación cruzada y se escogerá el mejor experimento.

Experimentaciones

20prototipos

Escogiendo 20 prototipos el resultado que obtenemos es el siguiente:

Tabla 4.1: LVQ 20*prototipos*

Pareja	Porcentaje de acierto
1	0.2397
2	0.2466
3	0.2432
Media	0.24316666666

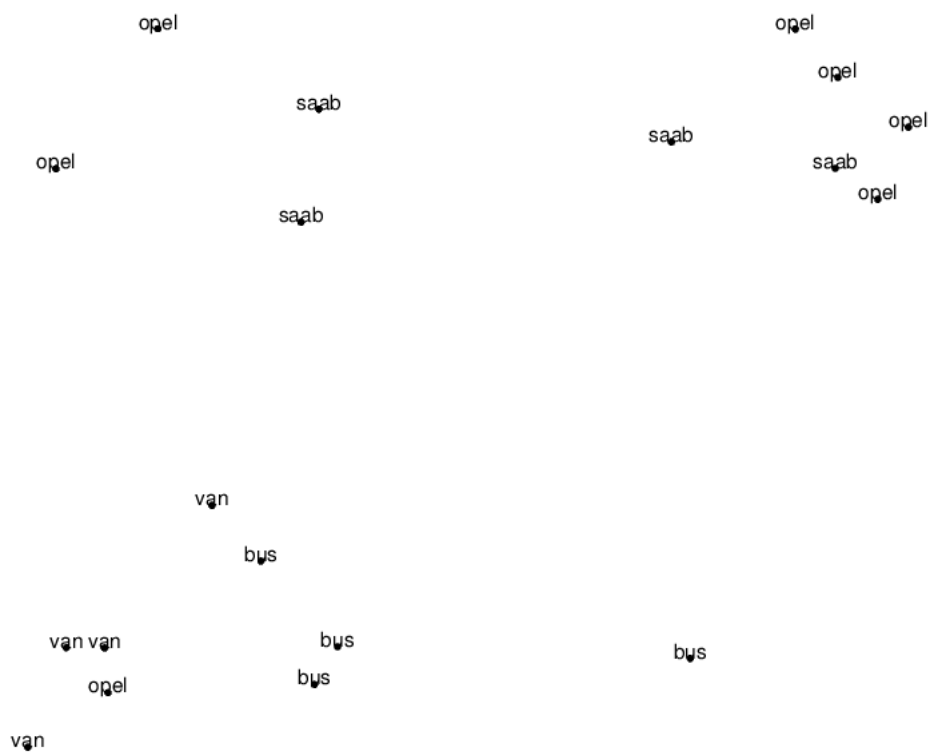


Figura 4.1: LVQ 20 *prototipos* pareja 1

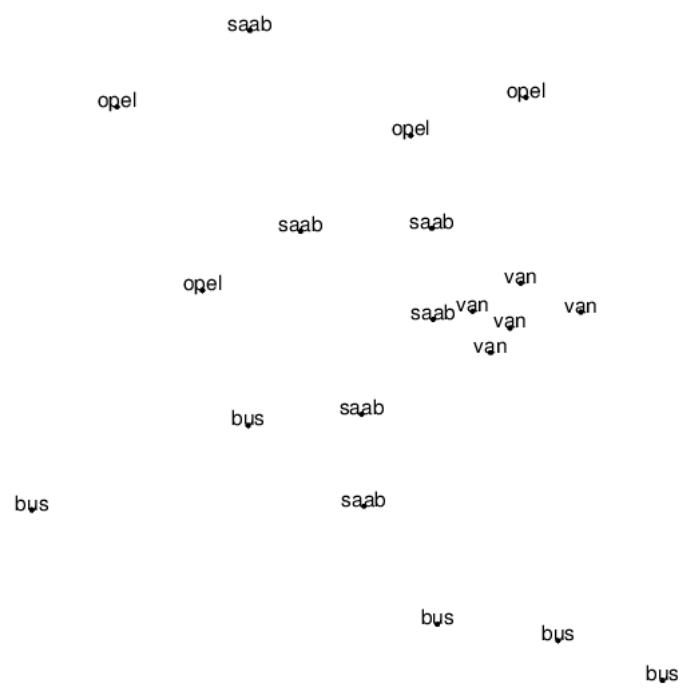


Figura 4.2: LVQ 20 *prototipos* pareja 2

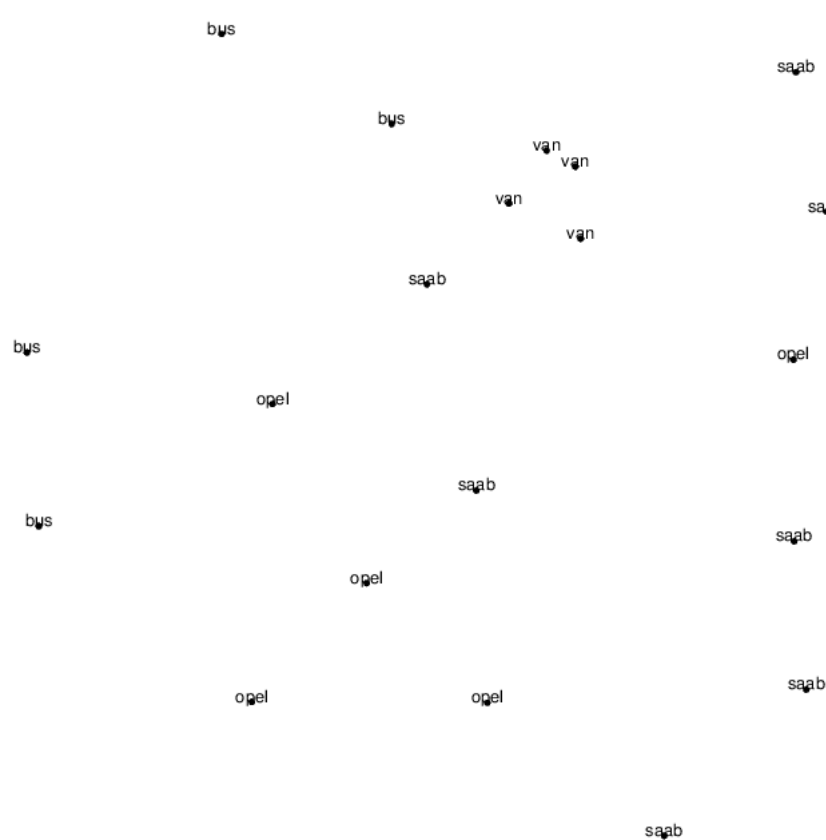


Figura 4.3: LVQ 20*prototipos* pareja 3

Como se puede apreciar en las figuras, al haber solo 20 prototipos, hay pocos prototipos por clase y la distancia entre ellos es bastante grande, el porcentaje de acierto está en torno al 24 %, que es un porcentaje de acierto en test bastante bajo. No se considera un buen modelo para clasificar correctamente.

50*prototipos*

Escogiendo 50 prototipos el resultado que obtenemos es el siguiente:

Tabla 4.2: LVQ 50*prototipos*

Pareja	Porcentaje de acierto
1	0.3938
2	0.6747
3	0.6575
Media	0.5753333333

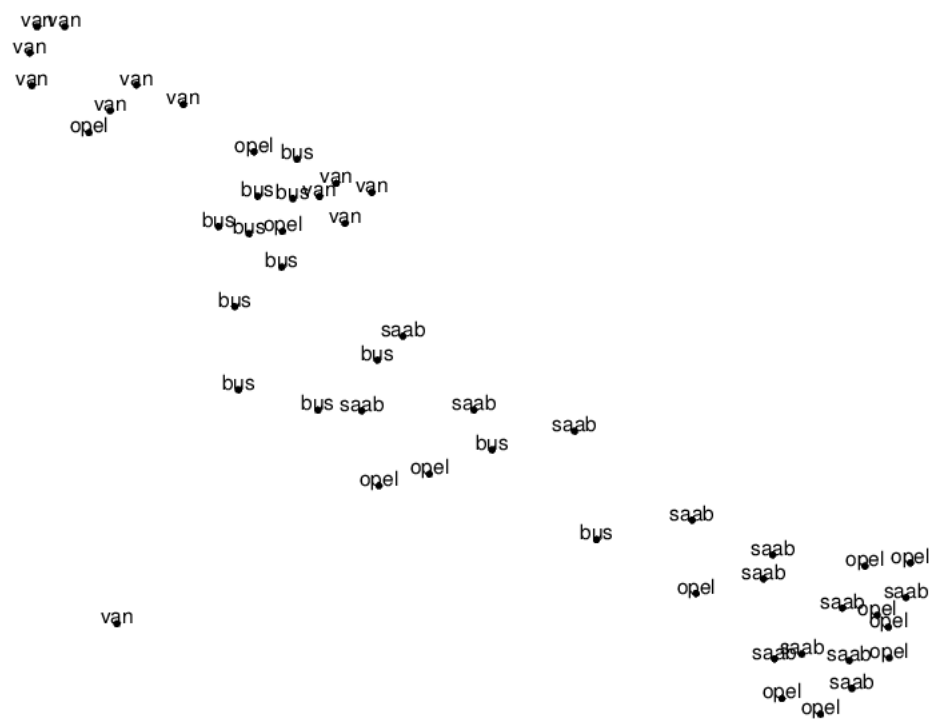


Figura 4.4: LVQ 50 *prototipos* pareja 1

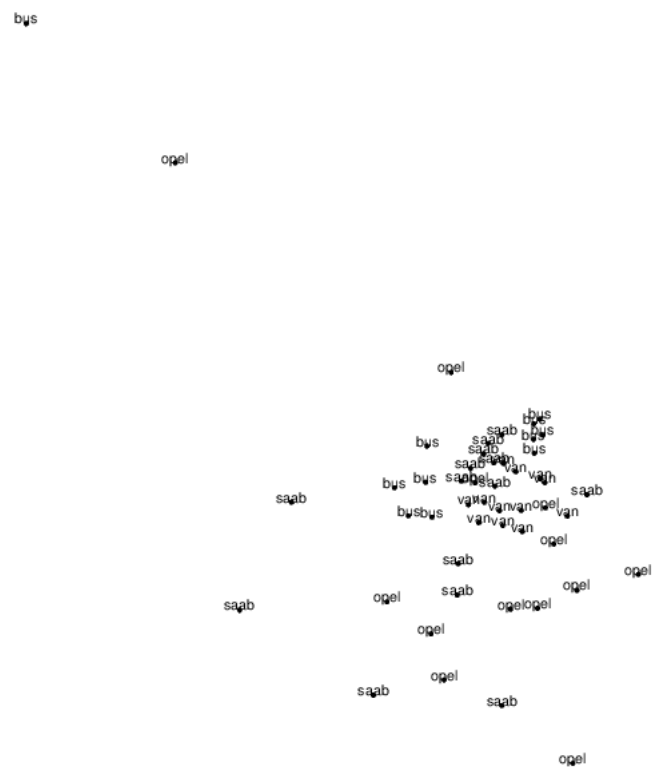


Figura 4.5: LVQ 50 *prototipos* pareja 2



Figura 4.6: LVQ 50*prototipos* pareja 3

Como se puede apreciar en las figuras, con 50 prototipos, la distribución de dichos prototipos para cada una de las tres parejas es muy diferente. No hay muchos prototipos por clase, y el porcentaje de acierto en test es significativamente mejor que en el caso anterior en el cual teníamos solo 20 prototipos, ya que ahora alcanzamos el 57 %.

100*prototipos*

Escogiendo 100 prototipos el resultado que obtenemos es el siguiente:

Tabla 4.3: LVQ 100*prototipos*

Pareja	Porcentaje de acierto
1	0.4760
2	0.7808
3	0.7534
Media	0.6700666666

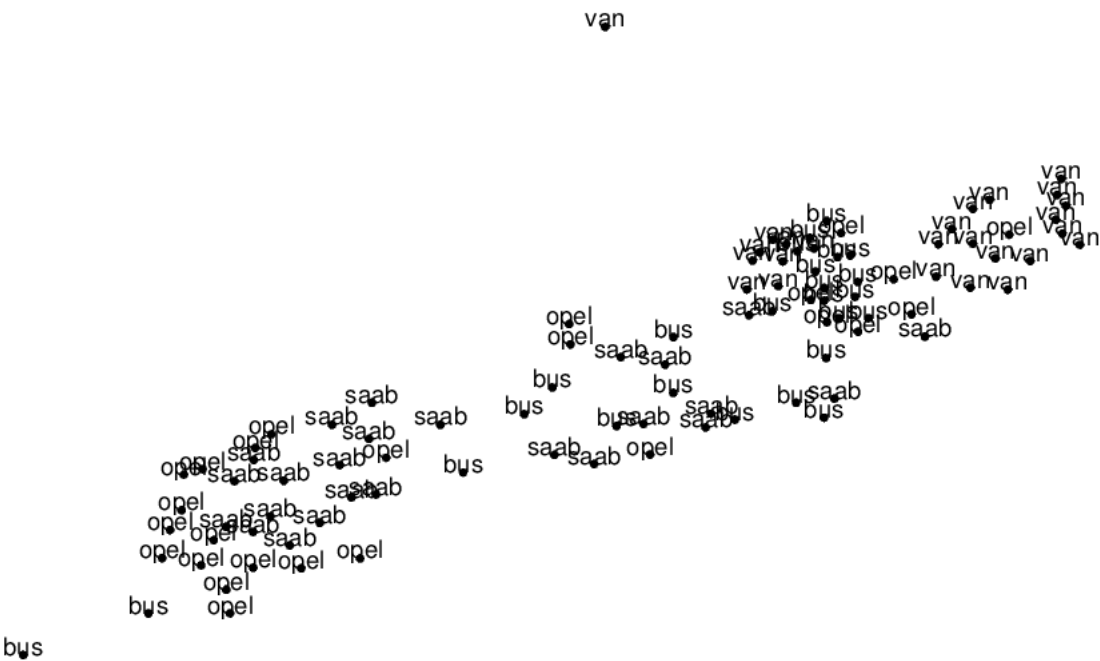
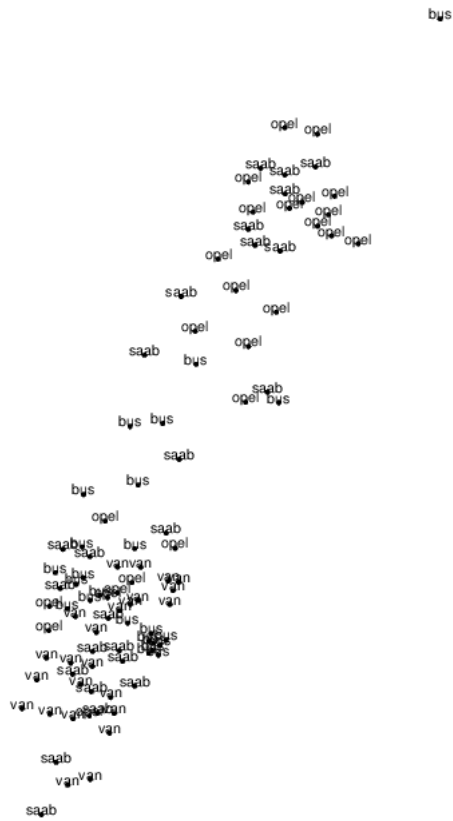
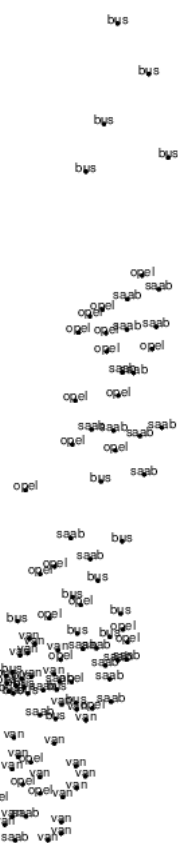


Figura 4.7: LVQ 100*prototipos* pareja 1

Figura 4.8: LVQ 100 *prototipos* pareja 2Figura 4.9: LVQ 100 *prototipos* pareja 3

Utilizando 100 prototipos se aprecia en las figuras que la distribución de los prototipos se asemeja bastante en el caso de las parejas 2 y 3, sin embargo es diferente para la primera pareja. Se observa que se produce gran concentración de prototipos en los tres casos, y el porcentaje de acierto ya asciende a 57 %, un 10 % más que utilizando la mitad de prototipos.

180*prototipos*

Escogiendo 180 prototipos el resultado que obtenemos es el siguiente:

Tabla 4.4: LVQ 180*prototipos*

Pareja	Porcentaje de acierto
1	0.4623
2	0.8630
3	0.8082
Media	0.7111666666



Figura 4.10: LVQ 180*prototipos* pareja 1

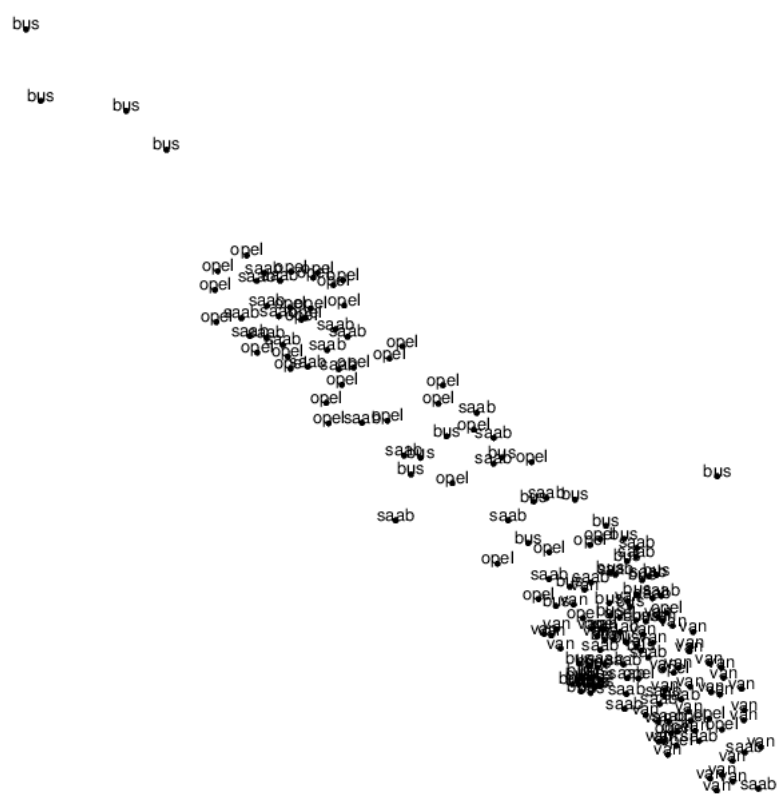


Figura 4.11: LVQ 180prototipos pareja 2



Figura 4.12: LVQ 180prototipos pareja 3

Se puede apreciar utilizando 180 prototipos que la distribución de los prototipos en el espacio de entrada de la primera pareja es inversa a la distribución que se aprecia para la primera pareja de 100 prototipos. Sin embargo el porcentaje de acierto se eleva ya al 71 %, un poco mayor que en el caso de 100 prototipos. Vuelve a suceder lo mismo, la distribución en el caso de la primera pareja es bastante diferente a la distribución apreciada en las otras dos parejas.

196*prototipos*

Escogiendo 196 prototipos el resultado que obtenemos es el siguiente:

Tabla 4.5: LVQ 196*prototipos*

Pareja	Porcentaje de acierto
1	0.4623
2	0.8185
3	0.8459
Media	0.7089

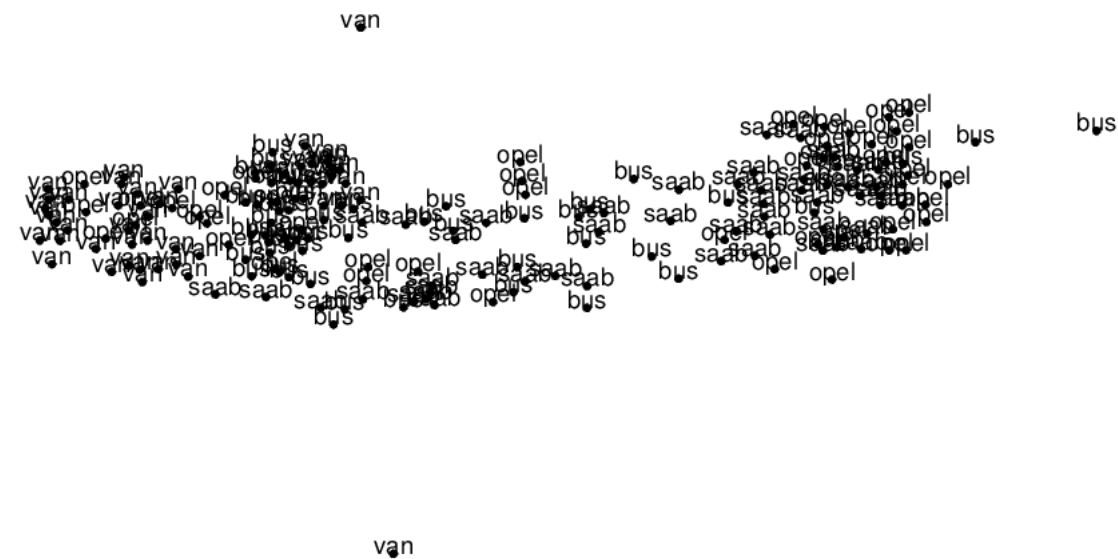


Figura 4.13: LVQ 196*prototipos* pareja 1

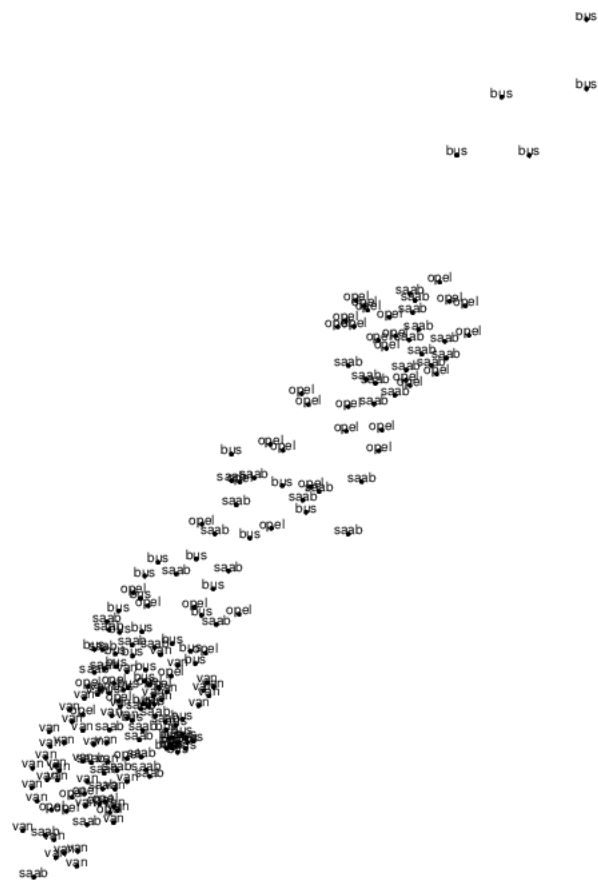


Figura 4.14: LVQ 196 *prototipos* pareja 2

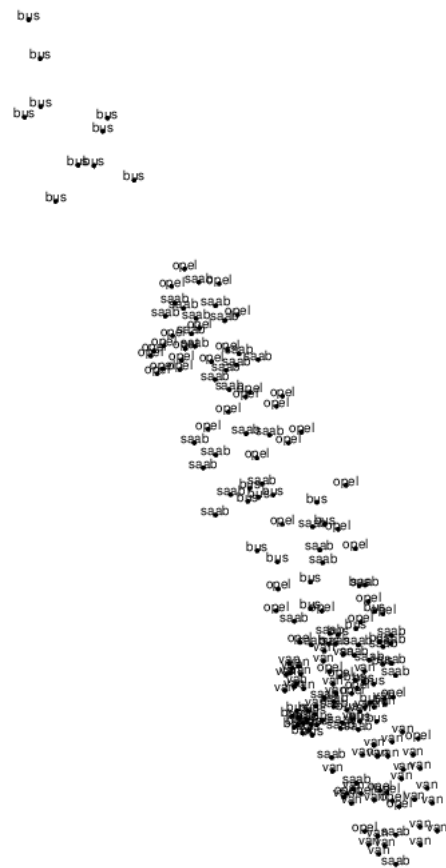


Figura 4.15: LVQ 196 *prototipos* pareja 3

Utilizando 196 prototipos la situación es bastante semejante al caso de 180, pero se empeora ligeramente el porcentaje de acierto, ahora es de 70,8 %, también se han estudiado los casos de 190 y 186 prototipos y no mejoran la tasa de acierto del caso 180, por tanto consideramos que el caso de 180 prototipos es el mejor experimento.

CAPÍTULO 5

Mapas autoorganizados de Kohonen

En este capítulo se va a explicar en primer lugar los mapas autoorganizados de Kohonen, que permiten realizar clasificación no supervisada, a diferencia de los otros dos modelos anteriores. Disponen de dos capas, la capa de entrada y la capa de competición. En este modelo las salidas son calculadas en base a la distancia.

5.1 Kohonen (SOM)

El modelo de mapas autoorganizados de Kohonen son redes de neuronas no supervisadas, formadas por dos capas y que basan sus salidas en la distancia, que puede ser calculada mediante el producto escalar o distancia euclídea (el método más utilizado). Al tratarse de un modelo basado en aprendizaje competitivo, se introduce el vector de entrada, que llega a la capa de competición y luego se escoge aquella neurona que produzca la salida más pequeña.

Se puede representar un mapa de Kohonen en un espacio bidimensional si tiene dos neuronas de entradas, en cuyo caso, las coordenadas de cada uno de los prototipos son los pesos de las neuronas de la capa de competición. La célula o neurona ganadora será aquella que minimice la distancia entre la propia célula y el prototipo.

Para realizar el aprendizaje se aplican prototipos y las células que no ganan, se acercan al prototipo, de esta manera se reorganiza la red.

5.2 Experimentación realizada

En cuanto a la experimentación realizada se ha probado con distintos números de prototipos, escogiendo números proporcionales al número de clases existentes. Se atenderá al porcentaje de acierto realizando la media entre las tres hojas de la validación cruzada y se escogerá el mejor experimento.

Experimentaciones

Tamaño del mapa $x = 1, y = 10$

Con un mapa con coordenada x igual a 1 y coordenada y igual a 10, tenemos un mapa unidimensional con 10 neuronas en la capa oculta y obtenemos los siguientes resultados:

Tabla 5.1: SOM $x = 1y = 10$

Pareja	Porcentaje de acierto
1	0.43493150684932
2	0.51282051282051
3	0.50177935943061
Media	0.48317712636

Viendo que el porcentaje de acierto es muy bajo, no veremos en detalle los mapas devueltos por los programas de monitorización del software utilizado.

Tamaño del mapa $x = 10y = 1$

Con un mapa con coordenada x igual a 10 y coordenada y igual a 1, tenemos un mapa unidimensional con 10 neuronas en la capa oculta y obtenemos los siguientes resultados:

Tabla 5.2: SOM $x = 10y = 1$

Pareja	Porcentaje de acierto
1	0.38356164383562
2	0.51648351648352
3	0.51245551601423
Media	0.47083355877

De nuevo, viendo este segundo mapa unidimensional, el porcentaje de acierto es muy bajo por lo que no nos detendremos en un estudio más a fondo.

Tamaño del mapa $x = 10y = 10$

Con un mapa con coordenada x igual a 10 y coordenada y igual a 10, tenemos un mapa bidimensional con 100 neuronas en la capa oculta y obtenemos los siguientes resultados:

Tabla 5.3: SOM $x = 10y = 10$

Pareja	Porcentaje de acierto
1	0.39041095890411
2	0.43956043956044
3	0.54804270462633
Media	0.45933803436

El resultado obtenido en el caso de 10×10 sigue siendo un porcentaje de acierto muy bajo que no nos ofrece buenos resultados, por lo que seguiremos experimentando por si encontrásemos un mapa bidimensional que clasifique mejor.

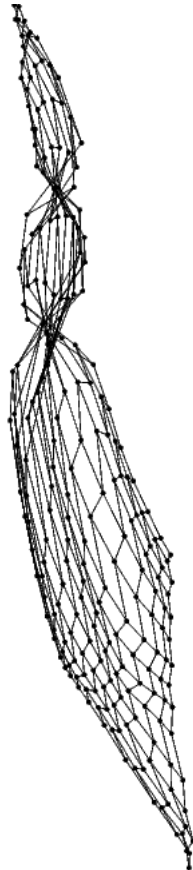
Tamaño del mapa $x = 15y = 15$

Con un mapa con coordenada x igual a 15 y coordenada y igual a 15, tenemos un mapa bidimensional con 225 neuronas en la capa oculta y obtenemos los siguientes resultados:

Tabla 5.4: SOM $x = 15y = 15$

Pareja	Porcentaje de acierto
1	0.39383561643836
2	0.41391941391941
3	0.66548042704626
Media	0.4910784858

El resultado obtenido en el caso de 15x15 es un porcentaje de acierto muy bajo teniendo en cuenta LVQ y MLP pero vamos a ver los resultados de la monitorización:

**Figura 5.1:** SOM $x = 15y = 15$ pareja 1

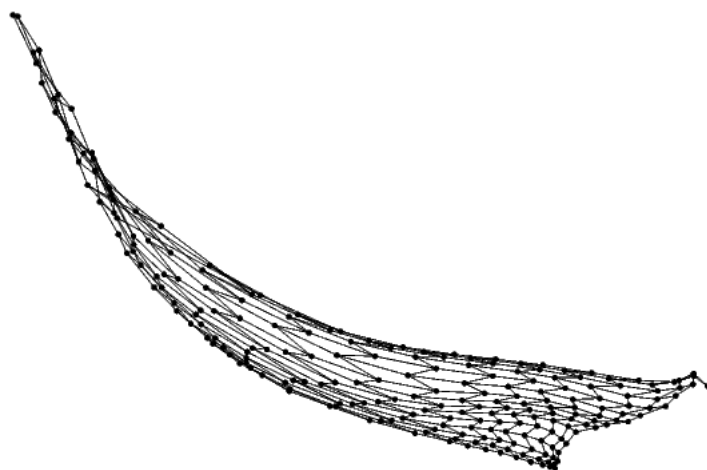


Figura 5.2: SOM $x = 15y = 15$ pareja 2

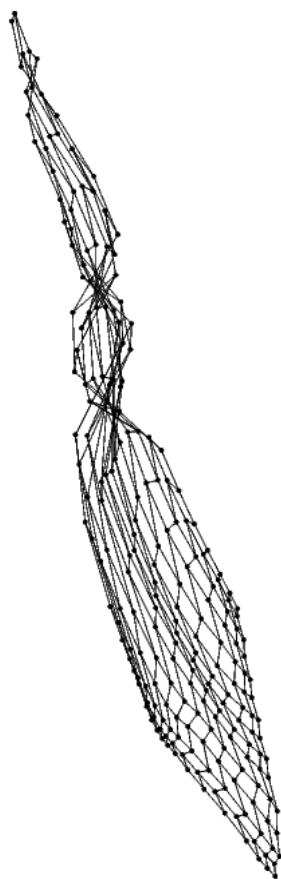


Figura 5.3: SOM $x = 15y = 15$ pareja 3

CAPÍTULO 6

Conclusiones

ver lvq vs som diapositivas lvq

Analizados ambos modelos, estamos en disposición de valorar objetivamente cual de ellos da mejor resultado para el problema que teníamos que resolver: para ello veamos los dos modelos elegidos para cada tipo, y también la tabla de resultados de cada uno de los dos modelos. La diferencia entre ambos modelos es que en Adaline teníamos la posibilidad de modificar ciclos y tasa de aprendizaje mientras que en Perceptrón Multicapa hemos visto que además se pueden añadir capas de neuronas ocultas para hacer el modelo más complejo. El error de test que obtenemos en el mejor modelo escogido por Adaline es de 0.019147481314295 mientras que el error de test que obtenemos en el mejor modelo escogido MLP es de 0.00569856426934457. Como vemos resulta evidente que el error mejora significativamente en un factor de $\frac{1}{4}$ aproximadamente. Consideramos entonces que una red MLP aproxima mejor y es capaz de generalizar mejor en este caso concreto, aunque hay que tener cautela porque es fácil dejarse llevar y añadir las suficientes capas para que MLP aproxime exactamente el conjunto de entrenamiento, que como ya hemos visto en ?? no nos lleva más que a una red que posteriormente fallará en la generalización de otros conjuntos de datos.

Es importante resaltar que aunque parezca que la diferencia es mínima en cuanto se "desnormalizan" los datos vemos que algunas salidas tanto en ?? como en ?? presentan errores de una magnitud bastante importante, y que en estos casos nuestras redes fallan en su generalización.

En cuanto a la práctica en general, pienso que ha habido problemas de rendimiento tanto en el script en R como en mi programa en PHP debido a que el proceso de aprendizaje si los ciclos son muy elevados he llegado a ver ejecuciones de veinte minutos. Lo cual ha provocado tener que invertir más tiempo en la práctica. Sería bueno plantearse mejorar el rendimiento de estos programas aunque se entiende que no es el objetivo principal.

Bibliografía

- [1] Conjunto de datos de siluetas de vehículos, Knowledge Extraction based on Evolutionary Learning. Obtenido de <http://sci2s.ugr.es/keel/dataset.php?cod=68>.
- [2] Peceptrón Multicapa, Inés M. Galván y José M^a Valls. Obtenido de <http://ocw.uc3m.es/ingenieria-informatica/redes-de-neuronas-artificiales/transparencias/material-de-clase.-tema-3/view>.
- [3] Mapas autoorganizados de Kohonen, Inés M. Galván y José M^a Valls. Obtenido de <http://ocw.uc3m.es/ingenieria-informatica/redes-de-neuronas/transparencias/Tema4%20AprendizajeNoSupervisado.pdf>.

