

**REDES DE NEURONAS ARTIFICIALES**  
**Curso 2017-18**

**PRÁCTICA II. PROBLEMA DE CLASIFICACIÓN**  
**Vehicle silhouettes**

## **1. Introducción**

En esta práctica se va a abordar un problema de clasificación con tres modelos diferentes: Perceptron Multicapa, LVQ y Mapas de Kohonen. Se trabajará con el dominio real de clasificación “Vehicle Silhouettes”. Es un dominio compuesto por 846 instancias, 18 atributos enteros y 4 clases que corresponden a distintos tipos de vehículos. El objetivo es clasificar una determinada silueta de un vehículo en uno de los cuatro tipos disponibles en este conjunto de datos: OPEL, SAAB, BUS, VAN.

A partir de imágenes de diferentes vehículos, se extraen una serie de características que son invariantes a las rotaciones y a los cambios de escala. Estas características son los 18 atributos enteros de cada uno de los datos. En el siguiente enlace puede encontrarse una breve descripción del dominio:

<http://sci2s.ugr.es/keel/dataset.php?cod=68#sub2>

De las 846 instancias totales, 212 corresponden a la clase OPEL, 217 a la clase SAAB, 218 a BUS y 199 a VAN.

Los datos para la realización de la práctica pueden encontrarse en la sección “Práctica II” en Aula Global.

## **2. Trabajo a realizar**

### **2.1 Preparación de los datos**

Antes de utilizar los datos para crear los modelos es necesario:

1. **Normalizar** los datos de entrada en el intervalo  $[0,1]$ . La salida tendrá que ser codificada según el modelo utilizado.
2. **Preparar los datos para Validación Cruzada:** en esta práctica se pide utilizar un esquema de validación cruzada de 3 hojas (folds). El procedimiento se detalla a continuación:
  - Se divide el conjunto total de datos en 3 partes (P1, P2 y P3). Se deberá mantener la proporción de instancias de las 4 clases del conjunto total en cada una de las partes P1, P2 y P3 (validación cruzada estratificada).
  - Se obtienen 3 parejas de ficheros (Entrenamiento+Test), de la siguiente forma:
    - P1 como conjunto de test y (P2+P3) como conjunto de entrenamiento.
    - P2 como conjunto de test y (P1+P3) como conjunto de entrenamiento.
    - P3 como conjunto de test y (P1+P2) como conjunto de entrenamiento.
  - Se aleatorizan (desordenan) los tres conjuntos de entrenamiento generados.
  - Los conjuntos de datos correspondientes (las 3 parejas de Test + Entrenamiento) deberán guardarse para ser utilizados por los tres métodos: Perceptron Multicapa, LVQ y Mapas Kohonen. De esta forma, los

resultados obtenidos por los diferentes modelos podrán compararse directamente.

- El resultado final de cada experimento (variando los parámetros de cada método) será **la media sobre las 3 hojas (folds)** de la tasa de aciertos en los tres conjuntos de test. También podrá obtenerse la tasa de aciertos en entrenamiento como la media de la tasa de aciertos de los tres conjuntos de entrenamiento.

## 2.2 Clasificación con Perceptron multicapa

Como en la práctica 1, se utilizará el **simulador snns** bajo el lenguaje de programación R (paquete RSNNS). Se facilitará el script básico a utilizar para entrenar el Perceptron Multicapa, así como para calcular su error en diferentes conjuntos de datos, realizar un gráfico que muestre la evolución de los errores a lo largo de los ciclos de aprendizaje y obtener los ficheros de resultados que contendrán las salidas obtenidas por la red para entrenamiento y test.

Se probarán diferentes arquitecturas de redes cambiando el número de neuronas ocultas y la razón de aprendizaje, obteniendo los porcentajes de aciertos en entrenamiento y test para cada experimento (media de los porcentajes de aciertos para cada hoja de la validación cruzada).

## 2.3 Clasificación con LVQ

El trabajo a realizar consiste en construir modelos de prototipos utilizando el algoritmo LVQ. Se variará el número de prototipos a utilizar y se obtendrán los porcentajes de aciertos en entrenamiento y test para cada experimento (media de los porcentajes de aciertos para cada hoja de la validación cruzada).

Para el uso del método LVQ se utilizará el software libre LVQ\_PAK. Para la realización de esta parte de la práctica se deben seguir los pasos indicados en el documento “Pasos básicos para el uso del software de LVQ” disponible en la sección “Práctica II” en Aula Global. Los pasos son:

- **Inicializar los centros o prototipos.** Se utilizará el programa *eveninit* (el mismo número de prototipos para cada clase) o *propinit* (número de prototipos proporcional a las instancias de cada clase).
- **Entrenar con LVQ** para ajustar estos prototipos de forma supervisada. Hay varios algoritmos de entrenamiento, siendo *olvq1* el más apropiado.
- **Obtener la tasa de aciertos de test y de entrenamiento.** Con el programa *accuracy* podemos obtener la tasa de aciertos por clase y total. Se puede utilizar con cualquier fichero que se quiera clasificar, en este caso nos interesan los 3 pares de ficheros de entrenamiento y de test.
- **Generar el fichero con los patrones clasificados.** El programa *classify* genera un nuevo fichero de datos con la clasificación asignada a cada patrón de entrenamiento o test.

## 2.4 Mapas autoorganizados de Kohonen

Al ser **SOM** (Self-Organizing Map) un algoritmo no supervisado orientado principalmente a la obtención de clusters o grupos en los datos, el mecanismo que se utilizará en esta práctica para resolver un problema de clasificación supervisado será el siguiente:

1. Construir un mapa con los datos de entrada del conjunto de entrenamiento.
2. Calibrar el mapa, es decir asignar a cada neurona del mapa (prototipo) una de las clases que aparecen en el conjunto de datos. Para ello y utilizando el programa *vcal*, se asignarán los patrones de entrenamiento a cada una de las neuronas del mapa (según su cercanía) y la clase de la neurona o prototipo será la clase mayoritaria del conjunto de patrones asignados.
3. Para cada patrón de test, se obtiene la neurona del mapa (prototipo) más cercana y se clasificará con la clase de dicha neurona (se realizará con el programa *visual*).
4. Se calcula el porcentaje de acierto sobre el conjunto de test y de entrenamiento. Este paso no puede realizarse con el software SOM-PAK, por lo que deberá realizarse con una hoja de cálculo o un programa independiente realizado por el alumno.

Para el uso de los mapas de Kohonen se utilizará el software libre SOM\_PAK. En el documento "Pasos básicos para el uso del SOM-PAK" disponible en la sección "Práctica II" en Aula Global se puede encontrar una descripción de los pasos para la realización de esta parte de la práctica. Estos pasos son:

- **Inicializar el mapa con *randinit*.**
- **Entrenar el mapa con *vsom*.** Es recomendable hacer el entrenamiento en dos fases, como se explica en la documentación.
- **Evaluar el error de cuantización** (quantization error) con *qerror*.
- **Calibrar el mapa con *vcal*,** asignando a cada prototipo la etiqueta o clase que le corresponde según una muestra de ejemplos etiquetados. Para ello se utilizará el conjunto de entrenamiento.
- **Clasificar los datos de test y de entrenamiento** con el programa *visual*. Este programa asigna a cada patrón del fichero de datos la neurona (o prototipo) del mapa más cercana. Además, como el mapa se ha etiquetado con *vcal*, se le asigna una clase o etiqueta a cada dato.
- Se podrán utilizar herramientas de monitorización y **visualización de los mapas**, como *umatrix*, *sammon*, etc.

Se probarán diferentes mapas variando la dimensión (unidimensional, bidimensional) y el número de neuronas en la capa de competición. Se deberá también, en la medida de lo posible, analizar algunos de los mapas obtenidos (los más representativos) con las herramientas visuales proporcionadas por este software.

### 3. Documentación a entregar

Se entregará una memoria en formato pdf y un conjunto de ficheros que contengan los modelos construidos y los resultados obtenidos.

- La memoria deberá contener, al menos, un capítulo de introducción, un capítulo para explicar la preparación de los datos realizada, un capítulo para cada uno de los modelos (Perceptron Multicapa, LVQ y Mapas de Kohonen) con una descripción de toda la experimentación realizada, así como tablas resúmenes que muestren los resultados de los experimentos realizados y un capítulo final con las conclusiones obtenidas y una comparación de los resultados obtenidos con los diferentes modelos.
- Los ficheros a entregar son los siguientes:
  - Ficheros de datos de entrenamiento y test al realizar la validación cruzada de tres hojas
  - Las tres redes PM del mejor experimento obtenido (una por hoja de validación cruzada) y los tres ficheros de resultados (salidas deseadas y obtenidas por la red) para los patrones de test y solo para la red que se considera más adecuada.
  - Los tres ficheros de prototipos obtenidos (entrenados) con LVQ\_PAK para el mejor experimento (uno por hoja de validación cruzada) y los tres ficheros de resultados correspondientes obtenidos con el programa 'classify' para los patrones de test y solo para el número de prototipos más adecuado.
  - Los tres mapas entrenados y calibrados con SOM\_PAK para el mejor experimento (uno por hoja de validación cruzada) y los tres ficheros de resultados correspondientes obtenidos con el programa 'visual' para los patrones de test y solo para el mapa más adecuado.

**La entrega de la práctica será el día 17 de diciembre del 2017.**