

Sistemas de Ayuda a la Decisión Clínica  
Práctica: Preprocesado de datos  
Master de Ingeniería Biomédica

Antonio José Serrano, Emilio Soria  
Departamento de Ingeniería Electrónica  
E.T.S.E  
Universidad de Valencia

## 1 Análisis de componentes principales

En estadística, el Análisis del Componente Principal (en inglés, PCA) es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Técnicamente, la PCA busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. En términos menos formales, puede usarse para determinar el número de factores subyacentes explicativos tras un conjunto de datos, que expliquen la variabilidad de dichos datos. Esta técnica la usaremos para reducir el tamaño de una imagen (en tonos grises) sin perder demasiada calidad. El código está disponible en el programa 'SADC\_Tema3.PCA.Imagen.m' en la dirección <http://www.uv.es/idal/madm>. Necesitaremos además las imágenes tiger.mat y fruits.mat

## 2 Ejemplos de Mapas autoorganizativos

En este apartado, y antes de comenzar a usar MATLAB, veremos una serie de demostraciones en Internet sobre el uso de esta red para diferentes tipos de problemas. El siguiente enlace contiene un gran número de demostraciones de diferentes modelos neuronales <http://staff.aist.go.jp/utsugi-a/Lab/Links.html>. Otros enlaces relacionados con los mapas autoorganizados son los siguientes:

- <http://www.cis.hut.fi/research/javasomdemo/>. Demostraciones simples pero muy interesantes para comprender las propiedades de los mapas autoorganizativos.
- WEBSOM (<http://websom.hut.fi/websom/>). Es una demostración típica y que refleja las capacidades de esta red. Se trata de ordenar mediante el uso del SOM una serie de documentos procedentes de un grupo de noticias. El vector de características utilizado es el vector de frecuencias de aparición de una serie de palabras clave. Muy aconsejable por la propia demostración y la cantidad de material que se encuentra en dicha WEB.
- <http://davis.wpi.edu/~matt/courses/soms/applet.html>. Aquí se encuentra un ejemplo visual de la capacidad de organizar los vectores de entrada de acuerdo a sus características intrínsecas. Los patrones de entrada son vectores de tres componentes (coordenadas RGB, que definen los diferentes colores que se tienen) y la red realiza un mapeo de dichos vectores de tres dimensionres a las dos dimensiones del mapa autoorganizado.
- Demo GNG. No se da la dirección completa porque es demasiada larga pero hay que ir al primer enlace que proporciona Google buscando con estas palabras clave. Aquí se muestra la capacidad de esta red neuronal para modelizar la distribución de un conjunto de datos. Además en esta página se muestra otros modelos neuronales relacionados con los mapas autoorganizados

### 3 Matlab y SOM

En esta parte de la práctica se verá como utilizar Matlab para implementar los mapas autoorganizados (S.O.M). En la página web <http://www.cis.hut.fi/projects/somtoolbox/> se encuentra una librería para MATLAB para generar mapas autoorganizados. Si se tiene que utilizar este tipo de redes esta librería proporciona todas las funciones necesarias además de un interfaz de usuario sencillo y muy fácil de manejar. Para lanzar el entorno gráfico es necesario escribir *som\_gui* en la línea de comandos (¡tras haber instalado la librería claro!).

La emplearemos en dos ejemplos; en uno de ellos se trata de datos simulados y el otro de datos reales. En cuanto a los datos simulados se ha generado una base de datos de paciente con los valores de peso, talla, índice de masa corporal, colesterol y nivel de grasa; evidentemente no podemos representar de forma conjunta las 5 variables por lo que usaremos el SOM para intentar sacar conclusiones de este conjunto de patrones. El fichero a utilizar es *SADC\_Tema3-SOM\_Colesterol\_data.m*.

Empezaremos con un ejemplo. La organización de la flor de iris. El conjunto consiste en 50 ejemplo de cada una de las tres especies de flor de Iris (setosa, virginica y versicolor). Cuatro características con medidas para cada muestra. Con la longitud y anchura tanto del pétalo como del sépalo.