

Aprendizaje Automático

TP3: Algoritmos de Vectores de Soporte

Profesores: J. Gambini - J. Santos

1. Generar, en forma aleatoria, un conjunto de ejemplos en $[0, 5] \times [0, 5]$ tal que pertenezcan a dos clases linealmente separables 1 y -1 .
 - a) Utilice un perceptrón simple escalón para separar al conjunto linealmente. ¿El hiperplano de separación es óptimo? Justifique la respuesta.
 - b) ¿Cómo obtendría el hiperplano óptimo a partir del hiperplano generado por el perceptrón simple escalón?
 - c) ¿Cómo utilizaría un perceptrón simple para separar dos clases que no son linealmente separables? Contemple la posibilidad de usar un perceptrón simple lineal.

2. El archivo `acath.xls` contiene datos de pacientes que recurrieron al centro médico con dolor en el pecho. Estos datos fueron extraídos de <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/acath.html>.

Se trata de una muestra de 3504 pacientes para los que se recogieron diversas variables cuyos nombres en la base de datos y descripción es la siguiente:

- `sigdz` : variable binaria que toma valores 1 y 0, indicando si el paciente presenta estrechamiento de al menos un 75 % de alguna de las arterias coronarias importante (`enf = 1`) o no (`enf = 0`).
 - `tvdlm` : lo mismo que la anterior pero corresponde a tres arterias con estrechamiento.
 - `sex`: variable categórica que indica el sexo del paciente, 0 masculino, 1 femenino.
 - `age`: variable continua que representa la edad en años del individuo.
 - `choleste`: variable continua que expresa los Mg/dl de colesterol.
 - `duracion`: variable continua que recoge la duración, en días, de los síntomas de la enfermedad coronaria.
- a) Dividir aleatoriamente el conjunto de datos en dos conjuntos, uno de entrenamiento y uno de prueba.

- b) Clasificar la variable categórica `sigdz` que indica si el paciente posee o no una enfermedad coronaria, utilizando el método SVM. Calcular la matriz de confusión.
- c) Proponer (y también realizar) una secuencia de pasos que permita encontrar una clasificación mejor que la anterior, evaluando diferentes valores de C y diferentes núcleos.

3. Segmentación de Imágenes en color:

Considere la imagen `cow.jpg` y las imágenes muestra: `vaca.jpg`, `cielo.jpg` y `pasto.jpg` correspondientes a las clases dentro de la imagen “vaca”, “cielo” y “pasto”, respectivamente.

- a) Construir un conjunto de datos para entrenamiento, indicando para cada muestra a qué clase pertenece.
- b) Dividir aleatoriamente el conjunto de datos en dos conjuntos, uno de entrenamiento y uno de prueba.
- c) Utilizar el método SVM para clasificar los pixels del conjunto de prueba, entrenando con el conjunto de entrenamiento. Utilizar diferentes núcleos y diferentes valores del parámetro C . Construir la matriz de confusión para cada caso.
- d) ¿Cuál es el núcleo que da mejores resultados? Pensar una justificación teórica para la respuesta.
- e) Con el mismo método ya entrenado clasificar todos los pixels de la imagen.
- f) Con el mismo método ya entrenado clasificar todos los pixels de otra imagen.