

Tareas 6 EL708 – Otoño 2006
Clasificación de Base de Datos Iris
(similar a tarea 5 pero con SVM)

El objetivo de esta tarea es construir un clasificador para la base de datos Iris (Fisher, 1936). Como recordarán la base de datos Iris posee 3 clases (*Sestosa*, *Versicolor*, *Virginica*) y cada elemento de la misma está descrito por 4 características (*sepal length*, *sepal width*, *petal length*, *petal width*). El número total de elementos de la base de datos es 150 y existen 50 de cada clase.

Se pide:

1. Crear un conjunto de entrenamiento y uno de prueba (test) para la base de datos Iris. El conjunto de entrenamiento debe contener el 80% de la base de datos y el de prueba el 20% restante. Usar conjuntos similares a los de la tarea 5.
2. Entrenar un clasificador SVM que resuelva el problema usando la base de datos de entrenamiento. Haga todos los supuestos que desee, la única restricción es que estos sean lógicos. Puede usar la implementación de SVM que desee, pero se recomienda la librería SVMTORCH que puede ser bajada desde http://www.idiap.ch/machine_learning.php?content=Torch/en_SVMTorch.txt. Es importante que prueben distintos valores para el parámetro C y para el parámetro del Kernel que escojan (polinomial, Gaussiano). Si tienen dudas sobre el uso de esta librería consulten al Sr. Mauricio Correa (mauricio.knight@gmail.com).
3. Validar los resultados usando la base de datos de prueba.
4. Como los resultados obtenidos dependen de cómo se haya dividido la base de datos inicialmente, haga diez divisiones distintas de la base de datos y repita los pasos 2 y 3 diez veces. Entregue como resultado el promedio y la desviación estándar de la tasa de reconocimiento obtenida.

Los informes deben ser entregados en forma impresa el día Martes 20 de Junio a las 17:00 en la secretaría del tercer piso del edificio de eléctrica. Los informes y los programas deben ser enviados por email a Patricio Loncomilla (ploncomi@ing.uchile.cl) el día Martes 20 de Junio hasta las 17:00.