

EL4106 Inteligencia Computacional

Tarea 3

Profesores: Pablo Estévez y Pablo Huijse

Auxiliar: Ignacio Reyes

Ayudante: Rodrigo Carrasco

Semestre: Primavera 2017

En esta tarea usted deberá resolver un problema de clasificación binaria sobre dos conjuntos de datos bidimensionales generados sintéticamente. Para esto usted utilizará (a) *Naive Bayes Classifier* con función de verosimilitud Gaussiana y (b) *Support Vector Machines* con algoritmo C-SVM. En el caso de SVM se pide que compare SVM lineal y SVM no-lineal con kernel Gaussiano. Adjunto a esta tarea encontrará un notebook de ipython (tarea3.ipynb). Utilícelo para responder los puntos mencionados a continuación. Al final de la sesión usted deberá subir a u-cursos un informe conciso que contenga las respuestas a las preguntas en el enunciado y el análisis de sus resultados. Se recomienda abrir su procesador de texto preferido y completar el informe a medida que avanza en la tarea.

1. Importe las librerías necesarias corriendo el bloque B.1, no es necesario correr este bloque de nuevo mientras el notebook esté activo.
2. Genere el conjunto *two moons* usando el bloque B.2, esto creará conjuntos de entrenamiento y validación de manera automática.
3. Compare los resultados de clasificación para los distintos algoritmos. Tome nota del área bajo la curva roc (AUC) de validación y entrenamiento. Describa lo que observa en términos del hiperplano separador obtenido por los clasificadores, **contrastando adecuadamente con la teoría**. ¿En qué casos se logra separar correctamente los conjuntos? ¿En qué casos se observa sobreajuste? ¿En que casos el modelo utilizado por el clasificador no es adecuado?
4. Para el caso de SVM usted debe calibrar adecuadamente el algoritmo encontrando la combinación de C y γ que maximizan el AUC de validación. Se pide además que compare el caso SVM lineal (donde sólo debe calibrar C) y el caso de SVM con kernel Gaussiano. Describa el efecto inducido por los parámetros en el hiperplano.
5. Repita el punto 3 usando un conjunto aleatorio de dos clases con características bidimensionales cada una separada en dos clusters. Descomente las líneas indicadas en el bloque B.2

IMPORTANTE: El día Viernes 13 de Octubre en horario de clase auxiliar se realizará una sesión de laboratorio en la sala de computación del segundo piso del DIE donde se desarrollará la tarea. La asistencia es OBLIGATORIA. Se recomienda PUNTUALIDAD.

IPYTHON NOTEBOOK: Para iniciar un servidor de *ipython notebook*, abra una consola de anaconda (Menú de inicio, anaconda, open anaconda shell) y escriba *ipython notebook*. Se debería abrir un browser en la dirección <http://localhost:8888/>. Una vez en la interfaz del *ipython notebook* utilice el explorador de carpetas y navegue hasta encontrar el archivo ipynb adjunto o utilice el botón *upload*. Para ejecutar cada bloque de código presione SHIFT+ENTER.