**REDES NEURONALES Y ALGORITMOS GENÉTICOS**

**Trabajo 2: Clasificación binaria**

Este trabajo está orientado a predecir una variable **binaria** a través de diferentes algoritmos de clasificación.

**Normas para la realización del trabajo**

1. El trabajo es individual.
2. Se entregará el trabajo en pdf en una copia digital enviada al Campus Virtual.
3. El trabajo deberá estar explicado (no basta con responder a las cuestiones), indicando, si es necesario, el código SAS utilizado. Se valora la claridad de exposición en el informe y la estructura. Puede contener Anexos de datos y gráficos o no, todo según vuestro criterio. El trabajo es libre, con lo cual se agradece sentido común.
4. El trabajo está automáticamente suspenso (aunque se podría recuperar en Septiembre incluyendo examen oral) si se da al menos una de las siguientes circunstancias:

La presentación y explicaciones son escasas. Los modelos comprobados son demasiado limitados.

Se observan signos de copia de otros trabajos de otros alumnos.

Se construirán los modelos sobre **una** matriz de datos a elegir voluntariamente. La calificación también dependerá de la complejidad de los datos a tratar. Se piden al menos 1000 observaciones y 10 variables.

Una web de donde se pueden descargar es:

http://sci2s.ugr.es/keel/category.php?cat=clas#sub2

**Cuestiones generales a responder**

**Se trata de conseguir obtener el mejor método/algoritmo para predecir, estable en cuanto a su performance, y en comparación con un modelo de regresión logística con selección de variables stepwise.**

1. Se requiere la comparación entre los mejores algoritmos y regresión logística ;
2. Se comprobará el efecto de la variación de los parámetros básicos de cada algoritmo.
3. Se probará si es preciso, modelos con diferentes variables
4. Los algoritmos a utilizar son obligatoriamente y como mínimo:

* Redes Neuronales
* Regresión Logística
* Random Forest
* Bagging
* Gradient Boosting
* Support Vector Machines (si no funciona bien no preocuparse;probar al menos kernel polinómico de grado 1)

También si se quiere y para comprender los datos se puede probar con un simple árbol.

1. Es necesario utilizar validación cruzada o como mínimo training/test repetido. Se comparará con diferentes particiones y semillas.

f) Es necesario hacer alguna prueba de ensamblado.

Se puede utilizar el programa SAS aportado por el profesor y cualquier modificación o uso de código. También se puede utilizar R y/o cualquier otro paquete, siempre que se cumpla (e) .

**Fecha de entrega: Viernes 10 de Junio**