

Guía para seguir el módulo Machine Learning

1) Instalar paquetes en R: basta ejecutar en Rstudio el script "paquetes R a instalar master big data.R".

2) Los scripts "utilidades básicas R y Rstudio.R" y "utilidades minimas R machine learning.R" son opcionales, podéis echarles un vistazo por si hay algo que os puede interesar.

3) En general, cada tema contiene:

a) Un documento escrito con las diapositivas, y otro llamado "referencia" más desarrollado.

b) Scripts de ejemplos a los que se alude en el documento y archivos de datos en formato .Rda. Las carpetas de carácter general para todo el curso "Todos los datasets" y "Todos los scripts de código" pueden utilizarse uniéndolas como workspace porque he incluido en ellas todo el material.

4) En cada tema hay bibliografía pero se espera que el módulo sea autocontenido, es decir que toda la información que se necesita está en los documentos y material aportado.

5) Los temas a abordar en este módulo son los siguientes:

Semana 1

1. Introducción
2. Redes Neuronales

Semana 2

3. Árboles
4. Bagging, Random Forest, Gradient Boosting

Semana 3

5. Support Vector Machines
6. Ensamblado
7. Comparación de algoritmos y estrategias avanzadas en Machine Learning
8. Decisión en clasificación binaria

Los tiempos son aproximados, se intentará enviar la documentación para dos semanas cada vez.

6) La tarea a realizar en este módulo será un trabajo de modelización predictiva:

Un trabajo largo de comparación de algoritmos predictivos sobre una variable dependiente binaria, utilizando redes, algoritmos basados en árboles (RF, GBM), SVM, métodos de ensamblado, etc.

Aunque el enunciado se dará la tercera semana, desde el primer día podéis comenzar a buscar el dataset, depurarlo o prepararlo. En el archivo "recopilación links datasets machine learning.txt" hay repositorios de internet donde se pueden descargar archivos de datos.

