

# Instrucciones Tarea Evaluación de Modelos de Clasificación

## Módulo Análisis de Datos, EMI2016-1

### 1<sup>er</sup> Semestre 2019

1. Los datos para esta tarea se han subido a la plataforma en el archivo Compras\_Retail que contiene:
  - 14 variables explicativas referidas a los datos de una tarjeta de tienda Retail, y de compras asociadas a la misma.
  - 1 variable objetivo que indica si en el núcleo familiar asociado a la tarjeta hay un embarazo.
2. El objetivo de la tarea es proponer un predictor para la existencia de embarazo en la familia basado en el uso de las variables explicativas realizando una evaluación comparativa de dos modelos de clasificación.
3. Los pasos a seguir son los siguientes:
  - (i) Separar los datos en un conjunto de Entrenamiento y otro de Testeo respetando la proporción existente entre 1 y 0 en la variable objetivo.
  - (ii) Utilizar los datos de entrenamiento para estimar dos modelos predictivos (Adaline y Regresión Logística).
  - (iii) Estimar el error esperado fuera de la muestra para los dos predictores constuidos utilizando los datos de Testeo.
  - (iv) Calcular otro estimador del error fuera de la muestra de ambos modelos utilizando Cross Validation (con 5 subconjuntos)
  - (v) Construir la matriz de confusión para ambos modelos. Además plotear la curva ROC y calcular el AUC.
  - (vi) En base a los resultados obtenidos señalar si alguno de los dos modelos es preferible, e indicar un valor de corte para el predictor recomendable de acuerdo a la naturaleza del problema.