Afi Escuela de Finanzas

Regresión Avanzada

Máster en Data Science y Big Data en Finanzas

Javier Nogales – PhD Matemáticas
Profesor Titular, Estadística UC3M
www.est.uc3m.es/nogales @fjnogales

GLM: 40%

- The file fair.csv contains information about extramarital affairs in the US
- More information: https://vincentarelbundock.github.io/Rdatasets/doc/Ecdat/Fair.html
- Objective: develop two GLMs, one to explain the variable nbaffairs and other for the variable rate, including all the other variables in both cases
  Note variable rate is ordinal, hence ordinal GLM is recommended, like MASS:polr
- For both models, provide full interpretation and insights (in terms of coefficients and diagnosis)
- For both models, evaluate the predictive performance by dividing the sample in two parts (training and testing sets)
- Upload to Campus Virtual: *dataset, R or Python code, and report* (no more than 8 pages)

Advanced Regression: 60%

- Predict home sales prices
- The data file is in HomeSalesData.csv : 20000 houses with 19 features (plus the price)

  Real data with some NAs, outliers, complex variables, etc.

  (You could use other data set, with same complexity. But check with me previously!)
- Develop two different models:
  - One to explain the main factors affecting the home prices (using the complete data set)
  - Other to predict the home prices (using different training/testing sets) and associated confidence intervals for the testing set
- Be original. Examples: interactions between variables, dimension reduction in observations or variables, feature engineering, other advanced tools, etc.
- You can take inspiration from other analysis, but always cite the source
- Upload to Campus Virtual: *dataset, R or Python code, and report* (no more than 12 pages)

Grading

- GLM model to explain nbaffairs: 1.5 points
- GLM to predict nbaffairs: 0.5 points
- GLM model to explain rate: 1.5 points
- GLM to predict rate: 0.5 points
- Data cleaning and feature engineering for HomeSalesData: 1 point
- Model to explain home prices, including brief descriptive analysis: 2 points
- Model to predict home prices, including confidence intervals: 2 points
- Originality and report: 1 point

Afi Escuela de Finanzas