

## 1. Análisis estadístico de datos cualitativos

Válido para un alumno

Resumen/contenido: Es muy habitual que los cursos básicos de Estadística presenten técnicas para datos cuantitativos. Las herramientas matemáticas empleadas en el análisis de variables cuantitativas son más sofisticadas y atractivas desde el punto de vista teórico. Sin embargo, en muchísimas áreas de la vida cotidiana (por ejemplo, en biomedicina y en las ciencias sociales) los datos recogidos son cualitativos (también llamados categóricos) y las herramientas estadísticas para el análisis de los datos cuantitativos no se pueden emplear en este contexto. El objetivo de este trabajo es hacer un recorrido por las técnicas estadísticas válidas para analizar datos cualitativos entendiendo los resultados matemáticos (y, por tanto, las suposiciones) que justifican esos procedimientos. Se practicará lo aprendido analizando conjuntos de datos reales con R.

Requisitos: No es necesario saber programar en R (basta tener interés por aprender y haber utilizado algún programa tipo Matlab o Python).

Asignaturas de cuarto relacionadas/compatibles: Estadística II

Bibliografía/referencias:

Agresti, A. (2012). *Categorical Data Analysis*. 3ª ed. Wiley.

Agresti, A. (2019). *An Introduction to Categorical Data Analysis*. 3ª ed. Wiley.

## 2. Clasificación y regresión con bosques aleatorios

Válido para un alumno

Resumen/contenido: En el problema de regresión el objetivo es describir los valores de una variable respuesta cuantitativa en términos de otras variables cuantitativas llamadas regresores o covariadas. La función de regresión de la respuesta sobre los regresores se estima en base a la información proporcionada por una muestra de todas las variables en  $n$  individuos. En el problema de clasificación supervisada, también llamado "reconocimiento de patrones", la variable "respuesta" es cualitativa e indica el grupo al que pertenece el individuo. El objetivo en este caso es clasificar correctamente a un individuo del que se conocen sus valores de unas variables auxiliares llamadas características (*features*), basándose en la información de una muestra de  $n$  individuos correctamente clasificados, en los que se han observado los valores de las características de interés. El objetivo del trabajo será, bien en el contexto de la regresión o en el de la clasificación, estudiar la técnica de árboles y su versión agregada, los bosques aleatorios (*ranfom forests*), que es un procedimiento de referencia en la actualidad. Se practicará lo aprendido analizando conjuntos de datos reales con R.

Requisitos: No es necesario saber programar en R (basta tener interés por aprender y haber utilizado algún programa tipo Matlab o Python).

Asignaturas de cuarto relacionadas/compatibles: Estadística II

Bibliografía/referencias:

Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.  
<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>

Izenman, A.J. (2008). *Modern Multivariate Statistical Techniques. Regression, Classification, and Manifold Learning*. Springer.