**PROPUESTA DE TRABAJO FIN DE MÁSTER (Modalidad A)**

**Curso 2021-2022**

**MÁSTER EN TÉCNICAS ESTADÍSTICAS**

|  |  |
| --- | --- |
| **Título** | Desarrollo de librería de R para la selección óptima de número de categorías |
| **Director/a** | JAVIER ROCA PARDIÑAS e IRANTZU BARRIO BERAZA |
| **Descripción del contenido** | Se ha propuesto previamente una metodología para seleccionar los puntos de corte óptimos para categorizar un predictor continuo en el contexto de  regresión logística. Esta categorización está basada en la capacidad discriminativa máxima (medida mediante el AUC) de la variable categorizada.  Aunque esta metodología permite seleccionar cualquier número posible de puntos de corte, ya sea en un contexto univariante o múltiple,  en muchas circunstancias el número de categorías en las que categorizar la variable predictora no está claro. Recientemente, los directores de este TFM han desarrollado un contraste de hipótesis bootstrap que permite seleccionar el número de puntos de corte óptimos a utilizar en la categorización de una variable continua.  El objetivo de este trabajo fin de máster es doble. En primer lugar se desea comparar los resultados obtenidos con este contraste bootstrap con aquellos obtenidos mediante el paquete CatPredi de R, y en segundo lugar, se desarrollará un paquete de R que incorpore la metodología más reciente, así como los algoritmos de selección disponibles en el paquete CatPredi. |
| **Recomendaciones** | Barrio I, Roca-Pardi~nas J, Arostegui I. (2020) Selecting the number of categories of the lymph node ratio in cancer research: a bootstrap-based hypothesis test. Statistical Methods in Medical Research (in press). DOI: 10.1177/0962280220965631 |
| **Otras observaciones** | El paquete se subirá a la plataforma CRAN y previsiblemente se publlicará en la revista The R Journal |