Identificación de OTUs diferenciales en microbiomas: Extensiones de la función explore_logratios para clasificación binaria y multinomial

R. Alberich, N. A. Cruz, R. Fernández, I. García, A. Mir, F. Rosselló

Departamento de Ciencias Matemáticas e Informática. Universitat de les Illes Balears

Instituto de Investigación Sanitaria Illes Balears (IdISBa)

Presentamos dos extensiones de la función `explore_logratios` de la librería `coda4microbiome` (Calle et al., 2023) de R. Esta librería ha sido diseñada para identificar unidades taxonómicas operativas (OTUs) diferenciales de un microbioma mediante el ajuste de una regresión logística en todos los pares de log-ratios. Los coeficientes se estiman utilizando penalizaciones de los métodos de regresión ridge y lasso, conocidos colectivamente como elastic-net.

La primera extensión consiste en encontrar un conjunto de OTUs con log-ratio relevantes, tales que la media de las correlaciones log-ratio entre ellos sea máxima. Se propone una visualización de las asociaciones encontradas que permite detectar si hay máximos locales de interés. En la práctica, el conjunto de OTUs que da asociación máxima suele separar muy bien las dos categorías de la variable binaria de interés. Probamos esta extensión con varios conjuntos de datos de referencia, demostrando que proporciona un conjunto óptimo de OTUs para clasificar eficazmente las dos clases de la variable.

La segunda extensión permite utilizar `explore_logratios` en problemas de clasificación discreta no binaria. Esta metodología se basa en el área bajo la curva (AUC) de predicción multinomial, como se describe en (Hand y Till, 2001) y permite emplear la primera extensión para la identificación de log-ratios relevantes en contextos con más de dos categorías.

Referencias

- Calle M.L., Pujolassos, M. and Susin A. (2023) `coda4microbiome`: compositional data analysis for microbiome cross-sectional and longitudinal studies. BMC Bioinformatics volume 24, 82
- Hand, D.J., Till, R.J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning 45, 171–186.