



ID de aportación : 41

Tipo: Taller

Machine Learning con Datos Censurados usando Tidymodels

Tidymodels es un metapaquete, donde se han integrado todos los procesos de construcción y evaluación de modelos predictivos, manteniendo la filosofía de programación de tidyverse. Centrado fundamentalmente en problemas de regresión y clasificación, tiene extensiones que permiten trabajar con datos censurados, datos de supervivencia.

El uso de datos censurados cada vez se está extendiendo más, como en problemas relacionados con la fidelización de productos y clientes en las empresas, y por otro lado, es fundamental en algunas áreas de investigación biomédica, como son la oncología y las enfermedades cardiovasculares

Este taller abarcará todas las fases de la construcción de un modelo predictivo con el paquete tidymodels y sus extensiones, con un ejemplo de datos de supervivencia de alta dimensionalidad ($p \gg n$). Se explicarán las etapas de pre-procesamiento de datos con “recipe”, medidas de la capacidad predictiva, como son el c-index, el brier score o los AUCs de las time-dependent ROC curves, con “yardstick”. La construcción de los modelos de supervivencia con el paquete “censored”, que es la extensión de “parsnip” para este tipo de datos. También se explicará la evaluación y optimización de los parámetros del modelo con muestras de training y testing, y con técnicas de remuestreo implementadas en los paquetes “rsample” y “tune”

Se explicarán desde modelos de supervivencia básicos, como es la regresión de Cox, hasta modelos de machine learning con datos censurados, como son Random Survival Forest o Boosting. Además, se establecerán comparaciones entre el rendimiento predictivo de estos modelos, integrándolos en un “workflow”, una de las facilidades fundamentales que permite tidymodels

¿Presentas la comunicación a premio?

Afiliación (del autor)

Senior Biostatician GEICAM

Autor primario: HERRANZ VALERA, Jesús (GEICAM)

Clasificación de pistas: Estadística