



ID de aportación : 37

Tipo: Oral

## Interpretación de modelos de machine learning: estimando la importancia de las variables predictoras y los efectos marginales sobre la variable objetivo

Al igual que durante el siglo pasado la inferencia estadística tuvo que desarrollarse para estimar modelos con los que conseguir extraer la máxima información de una muy corta cantidad de datos, la nueva realidad está caracterizada por disponibilidad de grandes cantidades de datos, y de diferente naturaleza que han dado origen a nuevos modelos y técnicas de análisis (machine learning, Deep learning, inteligencia artificial...). Estos nuevos modelos que podríamos calificar como de básicamente algorítmicos, y cuyo objetivo principal es la predicción, han venido a sumarse y a complementar, a los modelos de regresión tradicional, básicamente inferenciales y confirmatorios que utilizamos en las ciencias empíricas para entender el funcionamiento de los diferentes fenómenos de la realidad, el contraste de hipótesis y la cuantificación de la asociación entre variables dependientes e independientes.

Nos encontramos así ante un escenario de encuentro entre los dos enfoques disponibles para el análisis de datos, el puramente predictivo y algorítmico, por un lado, y el más explicativo e inferencial por otro. Enfoques que representan las dos culturas de la modelización estadística a la que hacía referencia Leo Breiman (2001). Dos culturas que lejos de distanciarse y enfrentarse están siendo objeto de un acercamiento e integración. En efecto, esos modelos algorítmicos, principalmente predictivos, han sido denominados como de “caja negra”, porque en ellos es complicado establecer de manera clara cuál es el efecto que tienen las diferentes variables predictoras o independientes sobre la variable objetivo o dependiente. Parte de los avances recientes en la Ciencia de Datos se encuentra precisamente en iluminar o dotar de luz a esos modelos de caja negra, para permitimos avanzar en el mejor conocimiento del funcionamiento de la realidad.

En este trabajo queremos presentar de una manera práctica y aplicada algunas de las técnicas existentes para estimar el impacto y la importancia de las variables predictoras en los modelos de machine learning. Se presentarán librerías como DALEX o iml que tratan precisamente de dotar de interpretabilidad a estos modelos “de caja negra”. Estas técnicas se compararán con otros desarrollos que mediante el uso de las expresiones algebraicas de las derivadas parciales en algunos de estos modelos posibilitan estimar el impacto marginal que cada una de las variables explicativas sobre la variable predictora.

### ¿Presentas la comunicación a premio?

#### Afiliación (del autor)

Facultad de Estudios Estadísticos –Universidad Complutense de Madrid

**Autor primario:** SANDUBETE, Julio E (Grupo de Investigación UCM 940051: Análisis de Datos en Estudios Sociales, de Género y Políticas de Igualdad. Facultad de Estudios Estadísticos –Universidad Complutense de Madrid)

**Coautores:** FERNÁNDEZ AMORÓS, Marcial (Grupo de Investigación UCM 940051: Análisis de Datos en Estudios Sociales, de Género y Políticas de Igualdad. Facultad de Estudios Estadísticos –Universidad Complutense de Madrid); GÓMEZ HIDALGO, Silvia (Grupo de Investigación UCM 940051: Análisis de Datos en Estudios Sociales, de Género y Políticas de Igualdad. Facultad de Estudios Estadísticos –Universidad Complutense de Madrid); ESCOT, Lorenzo (Grupo de Investigación UCM 940051: Análisis de Datos en Estudios Sociales, de Género y Políticas de Igualdad. Facultad de Estudios Estadísticos –Universidad Complutense de Madrid)

**Presentador:** SANDUBETE, Julio E (Grupo de Investigación UCM 940051: Análisis de Datos en Estudios Sociales, de Género y Políticas de Igualdad. Facultad de Estudios Estadísticos –Universidad Complutense de Madrid)

**Clasificación de pistas:** Estadística