

Trabajo Práctico Final

Introducción

En este último trabajo práctico nos sitúa como parte del equipo de ciencia de datos en una empresa de comercio electrónico que se especializa en vender pescado fresco de alta calidad a una clientela muy selectiva que consiste principalmente en restaurantes de lujo.

Cuando enviamos nuestros productos hay un dato que no puede faltar: el peso del pescado. Esta información es relevante por dos razones. En primer lugar, el monto de dinero que se le factura a nuestros clientes es función del peso del pescado que reciben. En segundo lugar, la empresa de transporte con la que trabajamos tiene diferentes bandas tarifarias según el peso de la mercadería transportada, y algunas de estas bandas pueden resultar bastante caras.

El problema es que nuestra empresa compra pescados al por mayor, por lo que se sabe el peso total del pedido, pero no el de los pescados individualmente. La solución mas obvia consiste en pesar a cada uno de los pescados de manera manual. Pero no es una gran solución: pesar cada pescado manualmente es costoso, requiere mucho tiempo y mano de obra. Entonces, ¿qué podemos hacer?, ¿ideas?

Resulta que hablando con algunos colegas descubrimos que nuestro proveedor mayorista conoce el tamaño (es decir, largo, alto y ancho) de cada pescado de manera individual. Es imposible pesar los peces en el bote, porque el bote siempre está en movimiento, pero tienen una cámara que registra el tamaño de cada pez .

¡Qué noticia! Pero eso no es todo. Nuestra empresa solía pesar cada pez manualmente antes de detenerse por razones de costo, lo que significa que tenemos un conjunto de datos de entrenamiento de diferentes tipos de peces, con su peso medido de manera precisa.

El desafío

El conjunto de datos `fish-market.csv` contiene la información de los pescados mencionados en la introducción e incluye mediciones sobre las siguientes variables:

- **Species:** Nombre de la especie del pescado
- **Weight:** Peso del pescado en gramos
- **Length1:** Longitud vertical en centímetros
- **Length2:** Longitud diagonal en centímetros
- **Length3:** Longitud transversal en centímetros
- **Height:** Altura en centímetros
- **Width:** Ancho diagonal en centímetros

El objetivo final de este trabajo es construir un modelo regresión bayesiano que en base a la información de los pescados prediga su peso. Para ello, se recomienda que se tengan en consideración las siguientes pautas generales que hacen a un análisis bayesiano:

- Introducción
- Análisis exploratorio de datos
- Propuesta y ajuste de modelos
 - Descripción matemática
 - Elicitación de los *priors*
 - Pruebas predictivas *a priori*

- Ajuste del modelo
- Evaluación de la convergencia de las cadenas de Markov
- Exploración de la distribución *a posteriori* de los parámetros
- Pruebas predictivas *a posteriori*
- Evaluación del ajuste del modelo
- Interpretación de parámetros
- Comparación de modelos
- Ajuste del modelo final en un conjunto de entrenamiento y predicción sobre un conjunto de datos de evaluación
- Análisis final y conclusión

Se recomienda fuertemente que se haga uso de diferentes visualizaciones para comunicar los resultados de las diferentes etapas del análisis y que se propongan y evalúen un mínimo de tres modelos. En particular, considere que sucede a medida que se incorporan mas predictores e interacciones entre los mismos.