

Índice general

Capítulo 1

Introducción

El modelado estadístico es una herramienta fundamental para la investigación científica y el análisis de datos, su principal propósito es el de aproximar la realidad a partir de la implementación de modelos matemáticos que tienen en cuenta la incertidumbre. Estos tipos de modelos son capaces de abarcar distintos problemas como pueden ser: la descripción de relaciones entre variables, la predicción de nuevos datos o la comprobación de hipótesis. ?

Hoy en día existen muchos métodos y técnicas para proceder a resolver los problemas antes mencionados, pero en este trabajo nos centraremos en el desarrollo de los Modelos Aditivos Generalizados (MAG). Sin embargo, hasta llegar a ellos, pasaremos por la descripción de los Modelos Lineales, los Modelos Lineales Generalizados (MLG) y los Modelos Aditivos. Esto se debe a que los MAG no son más que una extensión de los anteriores, así que haremos un transcurso desde modelos simples hasta un Modelo Aditivo Generalizado completo.

En el primer capítulo hablaremos de los Modelos Lineales y los Modelos Lineales Generalizados. El concepto de regresión lineal surgió a partir de la necesidad de estudiar la relación entre unas variables, de las cuales se conocen ciertos datos, mediante formalizaciones matemáticas. En concreto lo introdujo Francis Galton y luego fue desarrollado por el estadístico y matemático Karl Pearson a finales del siglo XIX. Sin embargo, ya en el 1805 Legendre proponía la primera forma del método de mínimos cuadrados, por lo que estamos hablando de técnicas que ya llevan más de dos siglos entre nosotros. A pesar de ello, no se desarrollan las nociones de los MLG hasta el 1970, estos modelos relajan las hipótesis que deben asumir los Modelos Lineales y permiten un primer acercamiento a que los modelos tengan un grado de no linealidad.

Comenzaremos el siguiente capítulo introduciendo los Modelos Aditivos y varios resultados básicos sobre la estimación de los MAG, a partir de ellos ya nos podremos adentrar más en profundidad en los resultados necesarios para sus futuras aplicaciones. En particular, añadiremos una sección que hable de las funciones de suavizado (*smoothers*) y los tensores (*tensors*) que aplicaremos luego en la práctica, además de presentar también conceptos sobre la comparación de estos tipos de modelos y sobre el contraste de hipótesis.

Los Modelos Aditivos Generalizados mejoran la metodología de los MLG incorporando la flexibilidad que aporta la regresión no paramétrica y mantienen la interpretabilidad de los datos del análisis de regresión con múltiples variables predictoras pues se modelan

como una suma de términos ‘suaves’. Actualmente, estos modelos son un punto de partida magnífico para el modelado de un problema, un GAM bien ajustado debería funcionar de manera conveniente, incluso comparado con métodos de *boosting* o de *deep learning*. Además, el MAG tiene una base para una mejor interpretabilidad y métricas de incertidumbre más sencillas, por lo que en muchos casos del análisis de datos, los Modelos Aditivos Generalizados son una buena opción.