



Estimación no paramétrica de funciones

(Fecha última actualización:04/10/2024)

Máster en Estadística. CURSO ACADÉMICO 2024-2025	
Título	Regresión no paramétrica: fundamentos y aplicaciones
Tipo	INVESTIGACIÓN <input checked="" type="checkbox"/> ORIENTACIÓN PRÁCTICA <input type="checkbox"/>
Número de alumnos	1
Profesor(es)/ email	María Dolores Martínez Miranda (mmiranda@ugr.es)
Descripción	<p>Los métodos de regresión no paramétrica o técnicas de suavizado constituyen hoy en día una herramienta fundamental para el análisis y la descripción de la posible relación entre una variable de respuesta y una o varias variables explicativas. La regresión asumiendo un modelo paramétrico (lineal, polinómico o en general no lineal) resulta poco flexible en la práctica y en muchos casos inadecuada. Centrándonos por ejemplo en la regresión simple, nuestra experiencia nos dice que no siempre es sencillo describir una nube de puntos mediante una función que dependa de unos pocos parámetros (dos si se trata de regresión lineal), si no imposible. La regresión no paramétrica no presupone un tipo de función concreta, sino que deja que sean los datos los que revelen la estructura de la función de regresión. Permitir esta flexibilidad supone por un lado un mayor coste computacional y por otro la presencia de sesgo en las estimaciones. El sesgo viene controlado por un parámetro denominado parámetro de suavizado (o ancho de banda en estimación tipo núcleo). Este parámetro también determina la varianza de los estimadores. Parámetros de suavizado pequeños dan lugar a estimaciones con sesgo pequeño y varianza grande, mientras que parámetros de suavizado grandes reducen la varianza a costa de incrementar el sesgo. De este modo el parámetro de suavizado representa el balance (trade-off) que existe entre sesgo y varianza, balance que además es característico en este tipo de métodos. Aprender a controlar ese balance es fundamental en la práctica y son varios los métodos que existen para ello.</p> <p>El trabajo que se propone supone una introducción teórica y práctica a los métodos de regresión no paramétrica, haciendo hincapié en los métodos lineales locales. Los métodos que se estudiarán están implementados en R en diversos paquetes especializados. Usando estos paquetes se desarrollarán en el trabajo ilustraciones de los aspectos teóricos usando datos simulados y datos reales. En algún caso esto requerirá programación en R a un nivel medio.</p>
Objetivos particulares	<ol style="list-style-type: none"> 1. Conocer y comprender el modelo de regresión no paramétrica. 2. Formular los métodos regresión no paramétrica más conocidos (tipo núcleo, splines, desarrollos en serie etc.) 3. Profundizar en los estimadores polinomiales locales con dos casos particulares de interés: estimador de Nadaraya-Watson y estimador local lineal. 4. Describir las propiedades teóricas asintóticas de los estimadores (consistencia, sesgo y varianza). 5. Comprender el papel que juega el parámetro de suavizado y formular el problema de selección de dicho parámetro. 6. Describir los selectores del parámetro de suavizado más comunes: validación cruzada, métodos plug-in, regla del pulgar. 7. Ilustrar los métodos estudiados usando implementaciones en R.



Prerrequisitos y recomendaciones	Haber cursado alguna asignatura donde se incluyan contenidos de modelos de regresión y su inferencia desde la perspectiva formal de la estadística matemática. Además se requiere destreza en el uso del entorno de programación y análisis estadístico R. incluyendo habilidad para estudiar paquetes específicos disponibles en CRAN y realizar tareas de programación de funciones.
Plan de trabajo	<ol style="list-style-type: none"> 1. Revisión de los aspectos teóricos. 2. Implementación computacional: estudio de herramientas disponibles R y creación de 3. funciones originales en su caso. 4. Aplicaciones con datos reales y datos simulados. 5. Reflexión sobre limitaciones y extensiones de la metodología considerada. 6. Elaboración de las conclusiones del trabajo.
Competencias generales y específicas	<p>Competencias generales:</p> <p>CG1 - Los titulados han de saber aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.</p> <p>CG7 - Los titulados han de realizar una contribución a través de una investigación original que amplíe las fronteras del conocimiento desarrollando un corpus sustancial, del que parte merezca la publicación referenciada a nivel nacional o internacional.</p> <p>CG8 - Los titulados deben ser críticos en el análisis, evaluación y síntesis de ideas nuevas y complejas.</p> <p>Competencias específicas:</p> <p>CE1 - Conocer métodos para el Análisis de Datos</p> <p>CE5 - Adquirir conocimientos avanzados en Inferencia Estadística</p> <p>CE17 - Adquirir capacidades de elaboración y construcción de modelos y su validación</p> <p>CE18 - Ser capaz de realizar un análisis de datos</p> <p>CE20 - Ser capaz de realizar una correcta representación gráfica de datos</p> <p>CE21 - Conocer, identificar y seleccionar fuentes estadísticas</p> <p>CE22 - Ser capaz de interpretar resultados a partir de modelos estadísticos</p> <p>CE24 - Ser capaz de extraer conclusiones y redactar informes</p> <p>CE25 - Ser capaz de identificar relaciones o asociaciones</p> <p>CE26 - Saber utilizar con destreza entornos de programación y análisis estadístico</p>
Bibliografía	<ol style="list-style-type: none"> 1. Fan, J. y Gijbels, I. (1996). Local polynomial modelling and its applications. Chapman and Hall, London. 2. García Portugués, E. (2022). Notes for Nonparametric Statistics. Version 6.5.9. ISBN 978-84-09-29537-1. https://bookdown.org/egarpor/NP-UC3M/ 3. Wand, M. (2015). KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995). R package version 2.23-15. https://CRAN.R-project.org/package=KernSmooth 4. Wand, M. P. y Jones, M. C. (1995). Kernel Smoothing. Chapman and Hall, London.