



Título del trabajo: Técnicas de regresión no paramétrica y su implementación en R
Tutor/a: María Dolores Martínez Miranda
Cotutor/a:
Departamento responsable: Estadística e I.O.
Perfil y número de estudiantes al que va dirigido (máximo 2): Dirigido a un estudiante con inquietud por aprender métodos modernos del análisis de regresión y su aplicación en la práctica. Número de estudiantes: 1
<i>(Rellenar sólo en caso de que la propuesta esté realizada a través de estudiante)</i>
Estudiante que propone el trabajo (Nombre, Apellidos, DNI):
Tipo de trabajo: 9
Competencias <i>(estas son la mínimas; consultar (**) si se considera añadir otras)</i> <i>Competencias generales:</i> G02, G03, G04, G05, G06, G07, G08, G09. <i>Competencias específicas:</i> E01, E02, E03, E04, E05, E08, E09, E10.
Resultados de aprendizaje <i>(estos son los mínimos; añadir otros si se considera)</i> <ul style="list-style-type: none">• Adquirir competencias globales ligadas al desarrollo y aplicación de los conocimientos del Grado.• Adquirir competencias ligadas a la búsqueda y organización de información y documentación relevante sobre el tema objeto de estudio.• Aplicar el “pensamiento estadístico” y tener capacidad para enfrentarse a las distintas etapas de un estudio estadístico (desde el planteamiento del problema hasta la exposición de resultados).• Saber presentar, de forma escrita y oral, la memoria, los resultados y las conclusiones del trabajo realizado.
Antecedentes y resumen del tema propuesto: 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. El precio a pagar al permitir esta flexibilidad es 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 (<i>trade-off</i>) 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. El trabajo que se propone supone una introducción teórica y práctica a los métodos de regresión



no paramétrica, haciendo hincapié en los métodos tipo núcleo. 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.

Breve descripción de las actividades presenciales y no presenciales a realizar:

Actividades presenciales (15-30%)	Planteamiento, orientación y supervisión	18 horas
	Exposición del trabajo	2 horas
	Otras:	
Actividades no presenciales (70-85%)	Preparación del trabajo	180 horas
	Elaboración de la memoria	100 horas
	Otras:	
Total (12 ECTS)		300 horas

Objetivos que se pretenden alcanzar:

- Conocer y comprender el modelo de regresión no paramétrica.
- Formular los métodos regresión no paramétrica más conocidos (tipo núcleo, splines, desarrollos en serie etc.)
- Profundizar en los estimadores tipo núcleo: estimador de Nadaraya-Watson y estimador local lineal.
- Describir las propiedades teóricas asintóticas de los estimadores (consistencia, sesgo y varianza).
- Comprender el papel que juega el parámetro de suavizado y formular el problema de selección de dicho parámetro.
- Describir los selectores del parámetro de suavizado más comunes: validación cruzada, métodos plug-in, regla del pulgar.
- Ilustrar los métodos estudiados usando implementaciones en R.

Bibliografía básica para la puesta en marcha del trabajo:

- Fan, J. y Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall, London.
- 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>
- Wand, M. P. y Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

Tipo de trabajo (*):

1. Estudio de profundización en algún tema concreto de Estadística, o como proyecto de aplicación de la misma a estudios o problemas de otros ámbitos científicos o sociales.



2. Realización completa de todas las fases de un proyecto estadístico, bien con auxilio de prácticas en empresas o con prácticas propuestas y dirigidas por el tutor.
3. Estudio de casos, teóricos o prácticos, relacionados con la Estadística.
4. Elaboración de un informe o un proyecto de naturaleza profesional.
5. Elaboración de un plan de empresa.
6. Simulación de encargos profesionales.
7. Trabajos bibliográficos sobre el estado actual de una temática relacionada con la Estadística.
8. Creación y/o empleo de herramientas informáticas para su uso en Estadística.
9. Trabajos de inicio a la investigación.
10. Trabajos cuya finalidad sea la divulgación de la Estadística en diversos contextos.
11. Trabajos sobre Historia de la Estadística.
12. Trabajos relacionados con la docencia de la Estadística.

Competencias (**)

Competencias generales:

G01. Poseer los conocimientos básicos de los distintos módulos que, partiendo de la base de la educación secundaria general, y apoyándose en libros de texto avanzados, se desarrollan en la propuesta de título de Grado en Estadística que se presenta.

G02. Saber aplicar los conocimientos básicos de cada módulo a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de la Estadística y ámbitos en que esta se aplica directamente.

G03. Saber reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

G04. Poder transmitir información, ideas, problemas y sus soluciones, de forma escrita u oral, a un público tanto especializado como no especializado.

G05. Haber desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

G06. Saber utilizar herramientas de búsqueda de recursos bibliográficos.

G07. Poder comunicarse en otra lengua de relevancia en el ámbito científico.

G08. Poseer habilidades y aptitudes que favorezcan el espíritu emprendedor en el ámbito de aplicación y desarrollo de su formación académica.

G09. Fomentar y garantizar el respeto a los Derechos Humanos, a los principios de accesibilidad universal, igualdad, y no discriminación; y los valores democráticos, de la cultura de la paz y de igualdad de género.

Competencias específicas:



E01. Conocer los fundamentos básicos del razonamiento estadístico, en el diseño de estudios, en la recogida de información, en el análisis de datos y en la extracción de conclusiones.

E02. Conocer, saber seleccionar y saber aplicar, técnicas de adquisición de datos para su tratamiento estadístico.

E03. Conocer los fundamentos teóricos y saber aplicar modelos y técnicas estadísticas en estudios y problemas reales en diversos ámbitos científicos y sociales.

E04. Saber seleccionar los modelos o técnicas estadísticas para su aplicación en estudios y problemas reales en diversos ámbitos científicos y sociales, así como conocer herramientas de validación de los mismos.

E05. Comprender la importancia de la Investigación Operativa como metodología de optimización, toma de decisiones y diseño de modelos particulares para la resolución de problemas en situaciones específicas.

E06. Comprender y utilizar básicamente el lenguaje matemático.

E07. Conocer los conceptos y herramientas matemáticas necesarias para el estudio de los aspectos teóricos y prácticos de la Probabilidad, la Estadística y la Investigación Operativa.

E08. Conocer y saber utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, bases de datos, visualización gráfica y optimización, que sean útiles para la aplicación y desarrollo de las técnicas estadísticas.

E09. Conocer los conceptos básicos y habilidades propias de un ámbito científico o social en el que la Estadística o la Investigación operativa sean una herramienta fundamental.

E10. Tomar conciencia de la necesidad de asumir las normas de ética profesional y las relativas a la protección de datos y del secreto estadístico, como premisas que deben guiar la actividad profesional como profesionales de la Estadística.