

Plan 549 GRADO EN ESTADISTICA

Asignatura 47108 METODOS DE SUAVIZADO EN ESTADISTICA

Grupo 1

Tipo de asignatura (básica, obligatoria u optativa)

optativa

Créditos ECTS

6

Competencias que contribuye a desarrollar

Se incluyen a continuación las que aparecen en la memoria de verificación del plan de estudios.

2.1

Generales

- G1. Capacidad para la gestión de la información: Incluye la capacidad para la búsqueda, manejo y exposición de información relevante de diversas fuentes, así como el manejo de las herramientas TIC necesarias.
- G2. Capacidad para la abstracción y el razonamiento crítico: El modelado y análisis de datos de problemas reales exige una elevada capacidad de abstracción, y el razonamiento crítico es necesario para hacer interpretaciones y establecer conclusiones y soluciones con rigor científico.
- G3. Capacidad para la puesta al día y el auto-aprendizaje: Incluye la capacidad para la búsqueda de recursos que permitan la solución de nuevos problemas o de nuevas técnicas, en un medio científico y tecnológico en continua evolución.

2.2

Específicas

- E1. Recogida y tratamiento de datos: Incluye la capacidad para decidir sobre el diseño del procedimiento de obtención de datos. Capacidad para la búsqueda de información de fuentes diversas y para la elaboración de cuestionarios. Capacidad para manejar bases de datos y para llevar a cabo el tratamiento de los mismos.
- E2. Descripción y síntesis de datos: Esta competencia es la que permitirá describir numéricamente y mediante gráficos diferentes características de interés de variables e individuos de los datos objeto de estudio, localización, escala, diferentes tipos de asociación, outliers,...
- E3. Ajuste de modelos estadísticos y de investigación operativa: Incluye la capacidad para crear o reconocer un modelo adecuado al problema objeto de estudio. Capacidad para el ajuste del modelo mediante las herramientas adecuadas.
- E4. Análisis de resultados, interpretación y validación de modelos: Incluye la capacidad para la selección y validación de un modelo. Capacidad para la interpretación de los modelos ajustados y las diferencias entre ellos.
- E5. Extracción de conclusiones: Incluye la capacidad para interpretar los resultados del ajuste de un modelo seleccionado en términos del problema objeto de estudio, viendo su utilidad y/o proponiendo la necesidad de otras orientaciones del estudio.
- E6. Presentación y comunicación de resultados: Incluye la capacidad para presentar los resultados de los análisis realizados, junto a las posibles soluciones a los problemas planteados por los demandantes del estudio en contextos diversos.

Competencias transversales

Instrumentales

- I1. Capacidad de análisis y síntesis
- I2. Capacidad de gestión de la información
- I3. Capacidad de organización y planificación
- I4. Conocimientos de informática relativos al ámbito de estudio
- I5. Resolución de problemas
- I6. Comunicación oral y escrita en lengua nativa
- I7. Conocimiento de lenguas extranjeras
- I8. Toma de decisiones

Personales

P2. Razonamiento crítico

P4. Compromiso ético

Sistémicas

S1. Aprendizaje autónomo

S2. Adaptación a nuevas situaciones

S3. Motivación por el trabajo bien hecho

S4. Iniciativa y espíritu emprendedor

Objetivos/Resultados de aprendizaje

El objetivo general de la asignatura es introducir técnicas no paramétricas, basadas en el suavizado núcleo, y su aplicación al análisis de datos.

Proporcionar técnicas, para el análisis exploratorio de datos y construcción de modelos, que pueden aplicarse a una amplia variedad de tipos de datos, sin imponer suposiciones restrictivas, con el objeto de poner de manifiesto estructuras en los datos que pueden quedar enmascaradas con otras técnicas de la estadística paramétrica.

Desarrollar la intuición y habilidades matemáticas necesarias para la comprensión del suavizador núcleo, en el contexto de la estimación de la función de densidad y de la función de regresión.

Destacar la importancia del parámetro de suavizado en el compromiso entre el error sistemático y el error aleatorio.

Proporcionar resúmenes de datos en dimensiones 1,2 y 3, mediante representaciones gráficas que complementan las de la estadística descriptiva.

Conocer y manejar la librería "sm" de R (web R: <http://www.r-project.org/>).

Contenidos

1.INTRODUCCIÓN A LOS MÉTODOS DE SUAVIZADO

Un compromiso entre los modelos paramétricos y no paramétricos.

Aplicaciones. Análisis de datos preliminares. Construcción de modelos. Bondad de ajuste. Estimación paramétrica.

Modificando la metodología estándar.

2. ESTIMACIÓN DE FUNCIONES DE DENSIDAD UNIVARIANTES

El histograma. Propiedades: El sesgo y la varianza. Amplitud óptima del intervalo. Inconvenientes. El polígono de frecuencias.

El estimador núcleo. Propiedades: El sesgo y la varianza, el error cuadrático medio integrado. Núcleo y parámetro de suavizado óptimos. Inconvenientes: Sesgo en la frontera, Suavizado constante.

Métodos de elección del parámetro de suavizado. Suavizado óptimo normal. Suavizado por validación cruzada.

Reglas plug-in.

Otros estimadores. Parámetro de suavizado variable, métodos basados en vecinos más próximos, métodos basados en series ortogonales.

3.INFERENCIA BASADA EN LA ESTIMACIÓN DE DENSIDADES

Bandas de variabilidad. Banda de variabilidad bootstrap. Test de normalidad. Banda de referencia normal.

4.ESTIMACIÓN DE FUNCIONES DE DENSIDAD MULTIVARIANTES.

Matriz de suavizado. Núcleo producto y radialmente simétrico. Visualizaciones.

Dificultades en grandes dimensiones.

Aplicaciones: Test de independencia. Regla de clasificación núcleo.

5.ESTIMACIÓN DE LA FUNCIÓN DE REGRESIÓN

Ideas básicas. Promedio móvil local. Estimador núcleo de Nadaraya-Watson. Estimador basado en el ajuste local de un polinomio. Estimador basado en vecinos más próximos. Estimador Loess. Estimador basado en series ortogonales.

Regresión no paramétrica en dos dimensiones.

6.INFERENCIA EN REGRESIÓN NO PARAMÉTRICA

El sesgo y la varianza. Estimadores de σ^2 . Bandas de variabilidad.

7.VALIDACIÓN DE MODELOS DE REGRESIÓN PARAMÉTRICOS

Contraste de no efecto. Banda de referencia.

Validando una relación lineal. Test de pseudo razón de verosimilitud. Banda de referencia.

8.COMPARANDO CURVAS Y SUPERFICIES

Comparando estimadores de la densidad en una dimensión.

Riesgo relativo en dos dimensiones.

Comparando curvas y superficies de regresión.

Contraste de paralelismo.

9. APLICACIONES

Verosimilitud local y suavizado en regresión logística.

Aplicaciones al Análisis de Supervivencia.

10. INTRODUCCIÓN A LOS MODELOS ADITIVOS

Modelos aditivos y algoritmo de "backfitting".

Modelos semiparamétricos.

Modelos aditivos generalizados.

Principios Metodológicos/Métodos Docentes

La asignatura se desarrollará mediante la realización de diversas actividades, clases en el aula, tanto teóricas como prácticas, clases prácticas en el laboratorio, tutorías individualizadas, pruebas puntuables y examen final.

Clases en el aula: La teoría básica necesaria será expuesta por el profesor en clases en el aula. Se ilustrará su aplicación mediante ejemplos. Esto hace difícil diferenciar claramente entre clases de teoría y clases prácticas en el aula. No obstante, podemos estimar que la "teoría" ocupará un 80% del tiempo total dedicado a las clases en el aula. Clases prácticas en el laboratorio. Los estudiantes realizarán prácticas de ordenador en el Laboratorio de Estadística para familiarizarse con el manejo de R y la librería sm. Se analizarán conjuntos de datos reales y simulados. Mediante el análisis de datos simulados pondremos de manifiesto el funcionamiento de los métodos introducidos. Con el análisis de conjuntos de datos reales y su interpretación se pretende que el estudiante sea capaz de extraer el conocimiento, sobre el modelo subyacente, que los métodos aprendidos ponen de manifiesto.

Criterios y sistemas de evaluación

Primera convocatoria (JUNIO):

La calificación de cada alumno se hará mediante evaluación continua de las diferentes actividades programadas en la asignatura: realización, presentación y entrega de prácticas, realización de exámenes parciales y examen final.

- En las fechas señaladas en el calendario general de actividades de 4º del Grado en Estadística, (ver <http://www.eio.uva.es/docencia/grado/>) se recogerán los trabajos prácticos propuestos a los estudiantes con la anterioridad suficiente, se realizarán, siempre dentro de las horas de clase, exámenes parciales que consistirán en dar respuesta a cuestiones sobre la materia estudiada hasta el momento, habitualmente con ayuda del ordenador y el examen final (junio).

- Cada uno de los trabajos prácticos se calificará, en una escala de 0 a 10, informando a los estudiantes del resultado obtenido, así como de los errores cometidos, en su caso. La nota media A de todos ellos se utilizará directamente en la calificación final de la asignatura, tanto en la primera convocatoria (JUNIO), como en la segunda convocatoria (JULIO), según se indica más abajo.

- La nota media B de todos los exámenes parciales, en una escala de 0 a 10, se utilizará directamente en la calificación final de la asignatura solo en la primera convocatoria (JUNIO), según se indica más abajo.

- La nota C del examen final (JUNIO), en una escala de 0 a 10, se utilizará directamente en la calificación final de la asignatura solo en la primera convocatoria (JUNIO), según se indica más abajo.

La calificación final de la asignatura = $A*0.2 + B*0.2 + C*0.6$.

Se hará sobre 10, con un decimal y su correspondencia cualitativa.

Segunda convocatoria (JULIO):

- El examen final (JULIO) se realizará en el día y hora fijados en la segunda convocatoria oficial de examen de la asignatura. La nota D del examen final (JULIO), en una escala de 0 a 10, se utilizará directamente en la calificación final de la asignatura en la convocatoria (JULIO), según se indica más abajo.

La calificación final de la asignatura = $A*0.2 + D*0.8$.

Se hará sobre 10, con un decimal y su correspondencia cualitativa.

Recursos de aprendizaje y apoyo tutorial

Tutorías:

- Las tutorías individualizadas podrán ser atendidas en las seis horas oficiales que se podrán consultar en la Web de la UVA a principio de curso o a cualquier otra hora, previa cita con el profesor.

Bibliografía

Bowman, A.W. and Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis. Oxford.

Hastie, T and Tibshirani, R..(1990). Generalized Additive Models. Chapman and Hall.

Silverman B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall.

Scott, D.W. (1992) Multivariate Density Estimation. Wiley.

Simonoff, J.S. (1996). Smoothing Methods in Statistics. Springer-Verlag.

Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing. Chapman and Hall.

Calendario y horario

Ver la página web de la Facultad de Ciencias:

<http://www.cie.uva.es/horarios>

<http://www.cie.uva.es/sites/files/files/horarios/GE.pdf>

Tabla de Dedicación del Estudiante a la Asignatura/Plan de Trabajo

ACTIVIDADES PRESENCIALES

HORAS

ACTIVIDADES NO PRESENCIALES

HORAS

Clases teóricas

26

Estudio y trabajo autónomo individual

75

Clases prácticas

Estudio y trabajo autónomo grupal

15

Laboratorios

26

Prácticas externas, clínicas o de campo

Seminarios

Otras actividades

8

Total presencial

60

Total no presencial

90

Responsable de la docencia (recomendable que se incluya información de contacto y breve CV en el que aparezcan sus líneas de investigación y alguna publicación relevante)

Bonifacio Salvador González, bosal@eio.uva.es

ver la página web del profesor <http://www.eio.uva.es/infor/personas/boni.html>

Idioma en que se imparte

Español

El alumno utilizará recursos para su aprendizaje, como la bibliografía, en Ingles
