

Máster en Técnicas Estadísticas

DATOS IDENTIFICATIVOS DE LA MATERIA

Nombre de la materia: Regresión no paramétrica y semiparamétrica

Número de créditos ECTS: 5

Curso académico: 2019/2020

Profesorado:

Rosa M. Crujeiras Casais (2 créditos)

César A. Sánchez Sello (3 créditos, coordinador)

OBJETIVOS DE LA MATERIA

El objetivo de esta materia es que el alumnado se familiarice con las técnicas de regresión no paramétrica y semiparamétrica, con especial énfasis en su aplicación práctica. Estos objetivos se concretan en los siguientes resultados del aprendizaje:

- Conocer las principales técnicas no paramétricas y semiparamétricas de estimación de la función de regresión.
- Saber escoger el modelo de regresión no paramétrico o semiparamétrico adecuado para analizar la dependencia existente en datos complejos procedentes de situaciones reales.
- Conocer las limitaciones de las técnicas no paramétricas en el análisis de situaciones reales con un alto número de variables.
- Ser autónomo en el análisis de datos en entornos aplicados multidisciplinares utilizando técnicas no paramétricas y semiparamétricas.

CONTENIDOS DE LA MATERIA

Tema 1. Estimación no paramétrica de la función de distribución.

La función de distribución empírica. Suavización de la función de distribución empírica. Selección del parámetro de suavizado.

Tema 2. Estimación no paramétrica de la función de densidad.

Estimación tipo núcleo de la función de densidad. Estimación por vecinos más próximos. Estimación de la sparsity: inversa de la función de densidad. Selección del parámetro de suavización: regla del pulgar, regla plug-in, validación cruzada, métodos bootstrap. Estimador de la densidad con soporte acotado. Estimación de las derivadas de la función de densidad. Análisis exploratorio basado en el estimador núcleo.

Tema 3. Regresión tipo núcleo.

El regresograma. El estimador de Nadaraya-Watson. El estimador polinómico local. Selección del parámetro de suavizado: plug-in y validación cruzada. Estimación de las derivadas de la función de regresión. Estimación no paramétrica de la regresión multivariante. Verosimilitud local: regresión logística no paramétrica.

Tema 4. Estimación de la regresión por vecinos más próximos.

El estimador clásico por los k vecinos más próximos. Variantes del estimador por vecinos más próximos. El estimador Loess. Comparación con la suavización ordinaria.

Tema 5. Estimación de la regresión mediante splines.

Mínimos cuadrados penalizados. Splines cúbicos y splines cúbicos naturales. Spline de interpolación y spline de suavización. Regresión spline: base de B-splines. Regresión wavelet, bases de Fourier y otras estimaciones a través de bases de funciones. Regresión spline penalizada.

Tema 6. Modelos parcialmente lineales y modelos aditivos.

Presentación de los modelos parcialmente lineales y de los modelos aditivos. Algoritmos de estimación: backfitting y mínimos cuadrados iterados. Contraste entre modelos: test de razón de verosimilitudes con grados de libertad aproximados. Interacciones no paramétricas.

Tema 7. Modelos aditivos generalizados.

Presentación de los modelos aditivos generalizados. Extensión de los algoritmos de estimación y contrastes para los modelos aditivos generalizados.

Tema 8. Modelos single-index.

Presentación del modelo single-index. El problema de identificación. Algoritmos de estimación y contrastes.

BIBLIOGRAFÍA BÁSICA Y COMPLEMENTARIA

BIBLIOGRAFÍA BÁSICA

- [1] Bowman, A.W. y Azzalini, A. (1997). Applied smoothing techniques for data analysis. Oxford University Press.
- [2] Härdle, W. (1990). Applied nonparametric regression. Econometric society monographs, Cambridge University Press.
- [3] Härdle, W., Müller, M., Sperlich, S. y Werwatz, A. (2004). Nonparametric and Semiparametric Models. Springer.
- [4] Wand, M.P. y Jones, M.C. (1995). Kernel Smoothing. Chapman Hall.
- [5] Wasserman, L. (2005). All of Nonparametric Statistics. Springer.
- [6] Wood, S.N. (2006). Generalized Additive Models. An Introduction with R. Chapman and Hall.

BIBLIOGRAFÍA COMPLEMENTARIA

- [1] Fan, J. y Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman and Hall.
- [2] Green, P.J. y Silverman, B.W. (1994). Nonparametric regression and generalized linear models: A roughness penalty approach. Chapman and Hall.
- [3] Hastie, T. y Tibshirani, R. (1990). Generalized Additive Models. Chapman and Hall.
- [4] Scott, D.W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley and Sons.
- [5] Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall.
- [6] Simonoff, J.S. (1996). Smoothing Methods in Statistics. Springer.
- [7] Wahba, G. (1990). Spline Models for Observation Data. Society for Industrial and Applied Mathematics.

COMPETENCIAS GENERALES Y ESPECÍFICAS

En esta materia se trabajarán las competencias básicas, generales y transversales recogidas en la memoria del título. Se indican a continuación cuáles son las competencias específicas que se potenciarán en esta materia:

E1 - Conocer, identificar, modelar, estudiar y resolver problemas complejos de estadística e investigación operativa, en un contexto científico, tecnológico o profesional, surgidos en aplicaciones reales.

E2 - Desarrollar autonomía para la resolución práctica de problemas complejos surgidos en aplicaciones reales y para la interpretación de los resultados de cara a la ayuda en la toma de decisiones.

E3 - Adquirir conocimientos avanzados de los fundamentos teóricos subyacentes a las distintas metodologías de la estadística y la investigación operativa, que permitan su desarrollo profesional especializado.

E4 - Adquirir las destrezas necesarias en el manejo teórico-práctico de la teoría de la probabilidad y las variables aleatorias que permitan su desarrollo profesional en el ámbito científico/académico, tecnológico o profesional especializado y multidisciplinar.

E5 - Profundizar en los conocimientos en los fundamentos teórico-prácticos especializados del modelado y estudio de distintos tipos de relaciones de dependencia entre variables estadísticas.

E6 - Adquirir conocimientos teórico-prácticos avanzados de distintas técnicas matemáticas, orientadas específicamente a la ayuda en la toma de decisiones, y desarrollar capacidad de reflexión para evaluar y decidir entre distintas perspectivas en contextos complejos.

E8 - Adquirir conocimientos teórico-prácticos avanzados de las técnicas destinadas a la realización de inferencias y contrastes relativos a variables y parámetros de un modelo estadístico, y saber aplicarlos con autonomía suficiente un contexto científico, tecnológico o profesional.

E9 - Conocer y saber aplicar con autonomía en contextos científicos, tecnológicos o profesionales, técnicas de aprendizaje automático y técnicas de análisis de datos de alta dimensión (big data).

E10 - Adquirir conocimientos avanzados sobre metodologías para la obtención y el tratamiento de datos desde distintas fuentes, como encuestas, internet, o entornos "en la nube".

METODOLOGÍA DOCENTE: ACTIVIDADES DE APRENDIZAJE Y SU VALORACIÓN EN CRÉDITOS ECTS

La enseñanza constará de clases expositivas e interactivas, así como de la tutorización del aprendizaje y de los trabajos encomendados al alumnado. Se proporcionará material para el seguimiento del curso, así como otro material orientativo del aprendizaje del software. En las clases expositivas e interactivas se resolverán ejemplos mediante el software R, por lo que es necesario que el alumnado disponga en el aula de un ordenador.

La actividad presencial, junto con el correspondiente y necesario trabajo personal del alumnado para su preparación, es valorada con tres créditos ECTS. Esta carga de trabajo incluye el examen final. Se considera que es suficiente una hora y media de trabajo personal para la preparación de cada sesión presencial de tipo teórico-práctico. Los otros dos créditos ECTS de la materia corresponden a trabajos prácticos que el alumnado tendrá que elaborar a lo largo del curso.

A continuación se presenta una aproximación de las horas que se dedicarán a cada tema.

TEMA 1. DISTRIBUCIÓN (1h expositiva, 1h de laboratorio)

TEMA 2. DENSIDAD (3h expositivas, 4h de laboratorio)

TEMA 3. REG NÚCLEO (2h expositivas, 3h de laboratorio)

TEMA 4. KNN (2h expositivas, 2h de laboratorio)

TEMA 5. SPLINES (2h expositivas, 3h de laboratorio)

TEMA 6. PLM Y ADITIVOS (2h expositivas, 3h de laboratorio)

TEMA 7. GAM (1h expositiva, 2h de laboratorio)

TEMA 8. SINGLE-INDEX (2h expositivas, 2h de laboratorio)

CRITERIOS Y MÉTODOS DE EVALUACIÓN

Evaluación continua (40%): la evaluación continua se realizará en base a la resolución de casos prácticos por parte del alumnado. En estos problemas, el alumnado utilizará el programa R y redactará los correspondientes informes. También se contempla la resolución de ejercicios en clase y su presentación cumplimentando formularios de evaluación. La calificación obtenida se conservará entre las oportunidades (ordinaria y Extraordinaria) dentro de la convocatoria de cada curso. Con las distintas actividades que se propondrán a lo largo del curso, se valorará el nivel de adquisición de las competencias básicas y generales, CB6-CB10 y CG1-CG5. También se evaluará el nivel alcanzado en las competencias transversales CT1-CT5 y de las competencias específicas E2, E6, E9 y E10.

Examen final (60%): el examen final constará de varias cuestiones teórico-prácticas sobre los contenidos de la materia, dentro de las que se podrá incluir la interpretación de resultados obtenidos con el programa utilizado en la docencia interactiva. En el examen se evaluará la adquisición de las competencias específicas E1, E3, E4, E5 y E8.

Presentación a la evaluación: se considera que la/el alumna/o concurre a una convocatoria cuando participa en actividades que le permiten obtener al menos un 50% de la evaluación final. El peso de la evaluación continua en la oportunidad extraordinaria de recuperación (pruebas de julio) será el mismo que en la evaluación ordinaria. En la segunda oportunidad de evaluación (recuperación), se realizará un examen y la nota final será el máximo de tres cantidades: la nota de la evaluación ordinaria, la nota del nuevo examen y la media ponderada del nuevo examen y la evaluación continua.

TIEMPO DE ESTUDIO Y DE TRABAJO PERSONAL QUE DEBE DEDICAR UN ESTUDIANTE PARA SUPERAR LA MATERIA

Se estima que el alumnado necesitará una hora y media para preparar el material correspondiente a cada hora de una clase teórica de tipo presencial. El alumnado dedicará el equivalente a dos créditos ECTS a preparar los diversos trabajos propuestos a lo largo del curso.

RECOMENDACIONES PARA EL ESTUDIO DE LA MATERIA

Para superar con éxito la materia es aconsejable la asistencia a las clases, siendo fundamental el seguimiento diario del trabajo realizado en el aula. También es recomendable que el alumnado sea capaz de utilizar el software estadístico R para explorar las posibilidades de las diversas técnicas no paramétricas explicadas a lo largo del curso. Además, para un mejor aprendizaje de la materia, es conveniente tener presente el sentido práctico de los métodos introducidos a lo largo del curso.

RECURSOS PARA EL APRENDIZAJE

El alumnado dispondrá de notas elaboradas por el profesorado de la materia, así como de la bibliografía recomendada en este programa. Para la difusión del material propio de la asignatura, se hará uso del Campus Virtual de la USC o bien de la web del máster.

OBSERVACIONES

El desarrollo de los contenidos de la materia se realizará teniendo en cuenta que las competencias a adquirir por el alumnado deben cumplir con el nivel MECES3. Los contenidos que se incluyen en esta materia son novedosos y altamente especializados. Se trabajará sobre la correcta formulación de modelos, la construcción de estimadores y la validación y análisis de las distintas propuestas estudiadas.