



UNIVERSIDADE DA CORUÑA *Universidade de Vigo*

Máster en Técnicas Estadísticas

DATOS IDENTIFICATIVOS DE LA MATERIA

Código de la materia: 614493125

Nombre de la materia: Aprendizaje Estadístico

Número de créditos ECTS: 5

Curso académico: 2021/2022

Profesorado:

Rubén Fernández Casal (coordinador, 2.5 créditos)

Manuel Oviedo de la Fuente (2.5 créditos)

OBJETIVOS DE LA MATERIA

En esta materia se pretende acercar al alumno al modelado y resolución de problemas de aprendizaje estadístico procedentes de aplicaciones reales. Los objetivos a alcanzar como resultado del aprendizaje son:

- Ser capaz de identificar y modelar problemas complejos de aprendizaje estadístico surgidos en aplicaciones reales.
- Adquirir conocimientos avanzados sobre aprendizaje estadístico tanto supervisado como no supervisado.
- Conocer los principios generales de las nuevas metodologías de aprendizaje supervisado para clasificación y regresión.
- Identificar y adquirir destreza en el uso de las principales herramientas de R de aprendizaje estadístico.
- Saber analizar datos utilizando técnicas de aprendizaje estadístico en contextos multidisciplinares.
- Ser capaz de aplicar con autonomía los resultados adquiridos en el análisis de datos masivos o de alta dimensión.
- Saber comunicar los resultados propios del Aprendizaje Estadístico a un público especializado o no especializado.

CONTENIDOS DE LA MATERIA

1. Introducción al aprendizaje estadístico.

Aprendizaje Estadístico y Aprendizaje Automático. Métodos de Aprendizaje Estadístico: supervisados y no supervisados. Construcción y evaluación de los modelos. La maldición de la dimensionalidad. Análisis e interpretación de los modelos. Paquetes de R.

2. Métodos de aprendizaje supervisado para clasificación.

Métodos basados en árboles. Bagging, bosques aleatorios y boosting. Máquinas de soporte vectorial (support vector machines). Otros métodos de clasificación.

3. Métodos de aprendizaje supervisado para regresión.

Modelos lineales generalizados y modelos aditivos. Métodos no paramétricos. Selección de variables y métodos de regularización: regresión contraída (ridge) y lasso. Métodos de reducción de la dimensión: componentes principales y mínimos cuadrados parciales. Regresión por projection pursuit. Redes neuronales. Métodos colaborativos.

BIBLIOGRAFÍA BÁSICA Y COMPLEMENTARIA

Básica

Fernández-Casal, R. y Costa, J. (2020). [Aprendizaje Estadístico](#). Libro online (con código en abierto en [github](#)).

James, G., Witten, D., Hastie, T. y Tibshirani, R. (2017). [An Introduction to Statistical Learning: with Applications in R](#). Springer.

Kuhn, M. y Johnson, K. (2013). [Applied predictive modeling](#). Springer.

Williams, G. (2011). *Data Mining with Rattle and R*. Springer.

Complementaria

Burger, S. V. (2018). *Introduction to Machine Learning with R: Rigorous Mathematical Analysis*. O'Reilly.

Efron, B. y Hastie, T. (2016). [Computer Age Statistical Inference](#). Cambridge University Press.

Hastie, T., Tibshirani, R. y Friedman, J. (2009). [The Elements of Statistical Learning: Data Mining, Inference, and Prediction](#). Springer.

Hastie, T., Tibshirani, R. y Wainwright, M. (2015). *Statistical Learning with Sparsity: the Lasso and Generalizations*. CRC press.

Torgo, L. (2011). *Data Mining with R: Learning with Case Studies*. Chapman & Hall/CRC Press.

COMPETENCIAS

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.
- E7 - Adquirir conocimientos teórico-prácticos avanzados de distintas técnicas de optimización matemática, tanto en contextos unipersonales como multipersonales, y saber aplicarlos con autonomía suficiente en un contexto científico, tecnológico o profesional.
- 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

La enseñanza constará de clases expositivas e interactivas, así como de la tutorización del aprendizaje y de las tareas encomendadas al alumnado. En las clases expositivas e interactivas se resolverán ejemplos mediante software especializado, por lo que es conveniente que el alumnado disponga en el aula de un ordenador.

Se propondrán actividades para el alumnado que consistirán en el modelado y resolución de problemas de aprendizaje supervisado de clasificación y de regresión.

CRITERIOS Y MÉTODOS DE EVALUACIÓN

La evaluación consistirá en la realización de trabajos prácticos, individuales o en grupo, que se les propondrán a los alumnos (evaluación continua), y en un examen escrito teórico-práctico. Los trabajos permitirán valorar el nivel de adquisición de las competencias básicas CB6-CB10, generales CG1-CG5, transversales CT1-CT5 y específicas E1-E10. Asimismo, el examen permitirá valorar el nivel de adquisición de las competencias específicas E1-E10.

Para aprobar la materia será necesario superar las dos partes (evaluación continua y examen). La nota del examen escrito representará el 50% de la calificación global, mientras que el 50% restante corresponderá a la evaluación continua (prácticas).

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

Cada crédito ECTS se traduce en 7 horas de clase de tipo presencial. Se estima que el alumnado necesitará una hora para preparar el material correspondiente a cada hora presencial, previa a la propia clase. Posteriormente, precisará de hora y media para la comprensión global de los contenidos, incluyendo las actividades asociadas a ejercicios y otras tareas. En total resultarán 24.5 horas por crédito ECTS.

RECOMENDACIONES PARA EL ESTUDIO DE LA MATERIA

Es recomendable que el alumno posea conocimientos básicos de regresión y de estadística multivariante. Disponer de conocimientos generales del entorno estadístico R facilitará el aprovechamiento de las clases y la realización de las prácticas.

Se aconseja participar activamente en el proceso de aprendizaje de la materia: asistencia y participación a las clases teóricas, prácticas, y de ordenador, utilización de horas de tutorías y la realización de un esfuerzo responsable de trabajo y asimilación personal de los métodos estudiados.

RECURSOS PARA EL APRENDIZAJE

Bibliografía y apuntes. Uso del repositorio de material docente del máster y de otros repositorios abiertos (GitHub o similar).

OBSERVACIONES

Se recomienda haber cursado una materia con contenidos de regresión, por ejemplo, *Modelos de Regresión* del itinerario aplicado o *Regresión Generalizada y Modelos Mixtos* del itinerario teórico del presente máster.

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. Esta materia tendrá una gran componente práctica, con énfasis en la identificación y modelado de problemas reales complejos y altamente especializados. Además de adquirir conocimientos sobre las herramientas fundamentales del aprendizaje estadístico, el alumno deberá ser capaz de resolver de manera autónoma problemas prácticos con la ayuda de los paquetes especializados en aprendizaje estadístico del software R.

COVID19. La metodología docente y los criterios de evaluación expuestos en esta guía docente se utilizarán independientemente del grado de presencialidad bajo el que se imparta la asignatura. En caso de que sea necesario suspender la docencia presencial, esta se impartirá de forma telemática (expositiva e interactiva) con sesiones síncronas y tutorías a través de Teams combinadas con material electrónico (videos y apuntes en formato electrónico). Los porcentajes de calificación de los trabajos prácticos y del examen escrito podrían verse modificados para adaptarse a las eventuales directrices de la Universidade da Coruña.