

Modelización de Datos Complejos

Código: 104864
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503852 Estadística Aplicada	OB	3	2

Contacto

Nombre: Rosario Delgado de la Torre
Correo electrónico: rosario.delgado@uab.cat

Uso de idiomas

Lengua vehicular mayoritaria: catalán (cat)
Algún grupo íntegramente en inglés: No
Algún grupo íntegramente en catalán: Sí
Algún grupo íntegramente en español: No

Equipo docente

Rosario Delgado de la Torre

Prerequisitos

Se supone que el alumno que cursa esta asignatura ha adquirido las competencias propias de las asignaturas de

- Càlcul 1,
- Eines informàtiques per a l'Estadística i Introducció a la Programació,
- Introducció a la Probabilitat i Inferència Estadística 1, i
- Aprenentatge Automàtic 1.

Será necesario un buen nivel y práctica en la programación con R.

Objetivos y contextualización

Aprender qué son y cómo se utilizan las Redes Bayesianas: son un modelo probabilístico que se utiliza en el Aprendizaje Automático Supervisado y que describe las relaciones de tipo probabilístico entre las variables que afectan a determinado fenómeno de interés (que puede ser un sistema complejo) y se pueden utilizar como clasificadores.

Entender cómo las Redes Bayesianas se utilizan para evaluar y cuantificar riesgos, entre otras aplicaciones. Conocer diferentes metodologías que se tendrán que aplicar, o no, cuando se trabaje con estos modelos, dependiendo de sus características, en la fase de pre-proceso de la base de datos o bien en la fase de construcción del modelo predictivo.

Conocer diferentes métricas de comportamiento para realizar la validación del modelo y entender su utilidad y adecuación, en función de las características de la base de datos.

Aprender cómo construir scripts de R que permitan aprender estos modelos a partir de una base de datos y hacer su validación, utilizando las librerías pertinentes. Aplicarlo a datos reales.

Competencias

- Analizar datos mediante la aplicación de métodos y técnicas estadísticas, trabajando con datos de diversas tipologías.
- Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otras personas.
- Diseñar un estudio estadístico o de investigación operativa para la resolución de un problema real.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Formular hipótesis estadísticas y desarrollar estrategias para confirmarlas o refutarlas.
- Interpretar resultados, extraer conclusiones y elaborar informes técnicos en el campo de la estadística.
- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
- Resumir y descubrir patrones de comportamiento en la exploración de los datos.
- Seleccionar y aplicar procedimientos más apropiados para la modelización estadística y el análisis de datos complejos.
- Utilizar correctamente un amplio espectro del software y lenguajes de programación estadísticos, escogiendo el más apropiado para cada análisis y ser capaz de adaptarlo a nuevas necesidades.
- Utilizar eficazmente la bibliografía y los recursos electrónicos para obtener información.

Resultados de aprendizaje

1. Analizar datos mediante otros modelos para datos complejos (datos funcionales, datos de recuento, etc.).
2. Analizar datos mediante técnicas de inferencia usando software estadístico.
3. Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otros.
4. Elaborar informes técnicos específicos del ámbito de la modelización estadística.
5. Emplear gráficos de visualización del ajuste y de la adecuación del modelo.
6. Establecer las hipótesis experimentales de la modelización.
7. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
8. Identificar las etapas en los problemas de modelización.
9. Identificar las suposiciones estadísticas asociadas a cada procedimiento.
10. Modificar ligeramente el software existente si el modelo estadístico propuesto lo requiere.
11. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
12. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.
13. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.
14. Utilizar eficazmente bibliografía y recursos electrónicos para obtener información.
15. Validar los modelos utilizados mediante técnicas de inferencia adecuadas.

Contenido

1. Introducción a las Redes Bayesianas (RBs).
Definición.
Inferencia con RBs.
Aprendizaje de RBs (tanto de estructura como de parámetros).
2. Las RBs como clasificadores.
La labor de clasificación dentro del Aprendizaje Automático Supervisado.

- El criterio MAP. Tipo de RB (Naive Bayes, Augmented Naive, TAN).
- Tipo de clasificación: binaria, multi-class, multi-label.
- 3. La validación y las métricas de comportamiento.
 - Validación cruzada.
 - Métricas para el caso binario y multi-class.
 - Métricas por el caso de clasificación ordinal.
- 4. Otros aspectos.
 - Clasificación multi-label: las cadenas de clasificadores.
 - La aproximación coste-sensitive.
 - El problema del desequilibrio de la base de datos: oversampling, thresholding,...
 - Ensamblajes de clasificadores.
 - RBs Gaussianas e híbridas.
 - RBs dinámicas.

Metodología

La asignatura se estructura a partir de clases teóricas, de problemas y prácticas. El seguimiento de la asignatura es presencial, pero será necesario ampliar las explicaciones del profesor con el estudio autónomo del alumno, con el apoyo de la bibliografía de referencia y del material facilitado por el profesor.

La clase de problemas se dedicará a la resolución orientada de algunos problemas propuestos. En las clases prácticas se trabajará con R y sus librerías. Se valorará especialmente la participación de los estudiantes en las clases de problemas y prácticas.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Problemas	14	0,56	1, 6, 5, 8, 9, 11, 15
Prácticas (entregas, controles)	12	0,48	2, 7, 4, 10
Teoría	26	1,04	1, 2, 3, 7, 4, 6, 5, 8, 9, 10, 11, 12, 13, 14, 15
Tipo: Supervisadas			
Tutorías	10	0,4	3, 7, 12, 13, 14
Tipo: Autónomas			
Estudio y pensar problemas	40	1,6	6, 5, 8, 9, 11, 15
Trabajo práctico con ordenadores	30	1,2	2, 7, 4, 10, 15

Evaluación

La calificación final de esta asignatura se obtiene como la media ponderada de las notas de:

- PAC1 (20%)
- PAC2 (20%)

- Examen (60%)

Las pruebas de evaluación continuada PAC1 y PAC2 consisten en una entrega de problemas/ejercicios de prácticas/trabajos con R, que se concretarán a lo largo del curso.

Sólo se tendrán en cuenta en el cálculo de la media ponderada aquéllas notas que sean como mínimo 3.5 sobre 10 (las que no lo cumplan ponderarán 0).

Para aprobar la asignatura es necesario que esta media sea de al menos 5.0 sobre 10.

En caso de no aprobar la asignatura en primera convocatoria, el alumno puede presentarse a recuperación. El examen de recuperación representa el 100% de la nota final para aquéllos alumnos que se presenten a la recuperación, que sólo pueden ser alumnos que no han superado la asignatura en primera convocatoria (el examen de recuperación no sirve para mejorar nota a alumnos que ya han aprobado).

Se considerará evaluable el alumno que haya presentado las entregas PAC1 o PAC2, o se haya presentado al examen o al examen de recuperación. En caso contrario constará en el acta como No Evaluable.

Para la eventual asignación de Matrículas de Honor no se tendrá en cuenta las notas de la segunda convocatoria.

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Examen	60%	3	0,12	1, 2, 3, 7, 4, 6, 5, 8, 9, 10, 11, 12, 13, 14, 15
PAC1	20%	6	0,24	1, 2, 3, 7, 4, 6, 5, 8, 9, 10, 11, 12, 13, 14, 15
PAC2	20%	9	0,36	1, 2, 3, 7, 4, 6, 5, 8, 9, 10, 11, 12, 13, 14, 15

Bibliografía

- Norman Fenton and Martin Neil, "Risk Assessment and Decision Analysis with Bayesian Networks", CRC Press. A Chapman & Hall Book, 2013. (Disponible en línea)
- Radhakrishnan Nagarajan, Marco Scutari and Sophie Lèbre, "Bayesian Networks in R with applications in Systems Biology", Springer, 2013. (Disponible en línea)
- Oliver Porret, Patrick Naïm and Bruce Marcot, "Bayesian Networks. A practical guide to applications". Series: Statistics in Practice. Wiley, 2008. (Disponible en línea)
- Richard E. Neapolitan, "Learning Bayesian Networks", Prentice Hall Series in Artificial Intelligence, 2004.

Software

Se utilizará el software R con algunas librerías que se indicarán oportunamente a lo largo del curso. Preferiblemente en el entorno RStudio.