

Aprendizaje Automático 1

Código: 104870
Créditos ECTS: 6

| Titulación | Tipo | Curso | Semestre |
|------------------------------|------|-------|----------|
| 2503852 Estadística Aplicada | OB | 3 | 1 |

Contacto

Nombre: Antonio Lozano Bagen

Correo electrónico: Antonio.Lozano.Bagen@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

Joan Valls Marsal

Prerequisitos

Esta asignatura supone que el alumno ha obtenido los conocimientos que se imparten en diferentes asignaturas sobre los siguientes temas:

- Cálculo en varias variables.
- Probabilidad.
- Modelos lineales.
- Programación en R.

Objetivos y contextualización

Esta asignatura pretende familiarizar al alumno con diferentes métodos de aprendizaje automático aplicando el punto de vista utilizado cuando se dispone de grandes cantidades de datos.

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.
- Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
- Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
- 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 las fuentes y técnicas de adquisición y gestión de datos adecuadas para su tratamiento estadístico.
- Seleccionar los modelos o técnicas estadísticas para aplicarlos a estudios y problemas reales, así como conocer las herramientas de validación de los mismos.
- Seleccionar y aplicar procedimientos más apropiados para la modelización estadística y el análisis de datos complejos.
- Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
- 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 metodología de aprendizaje automático.
2. Aplicar el espíritu crítico y el rigor para validar o refutar argumentos tanto propios como de otros.
3. Caracterizar grupos homogéneos de individuos mediante análisis multivariante.
4. Describir las ventajas e inconvenientes de los métodos algorítmicos frente a los métodos convencionales de la inferencia estadística.
5. Descubrir comportamientos y tipologías de individuos mediante técnicas de minería de datos.
6. Emplear gráficos de resumen de datos multivariados o más complejos.
7. Evaluar de manera crítica y con criterios de calidad el trabajo realizado.
8. Identificar las suposiciones estadísticas asociadas a cada procedimiento avanzado.
9. Identificar, emplear e interpretar los criterios para evaluar el grado de cumplimiento de los requisitos necesarios para aplicar cada procedimiento avanzado.
10. Implementar programas en lenguajes adecuados para la minería de datos.
11. Obtener y gestionar bases de datos complejas para su posterior análisis.
12. Proyectar un estudio en base a metodologías multivariantes y/o minería de datos para resolver un problema contextualizado en la realidad experimental.
13. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.
14. Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
15. 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.
16. 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.
17. Trabajar cooperativamente en un contexto multidisciplinar asumiendo y respetando el rol de los diferentes miembros del equipo.
18. Utilizar eficazmente bibliografía y recursos electrónicos para obtener información.
19. Utilizar métodos de minería de datos para validar y comparar posibles modelos.

Contenido

- Introducción al aprendizaje automático
- Regresión lineal i logística
- Aprendizaje estadístico
- Máquinas de soporte vectorial

- K-vecinos más cercanos
- Navie Bayes
- Árboles de decisión
- Ensembles
- Minería de textos
- Análisis de redes

Metodología

La asignatura tiene programadas, a la semana, dos horas de teoría y dos horas de prácticas.

- Clase de teoría: se definen y se explican los diferentes métodos con sus características particulares y se muestran ejemplos concretos.

- Clase de prácticas: se trabajan los métodos explicados en clase de teoría con diversos conjuntos de datos utilizando el lenguaje de programación R.

Se considera que, para cada hora de teoría y prácticas, el alumno deberá dedicar una hora adicional para la preparación y/o finalización de la sesión.

Actividades

| Título | Horas | ECTS | Resultados de aprendizaje |
|-------------------------------------|-------|------|---|
| Tipo: Dirigidas | | | |
| Sesiones de prácticas | 50 | 2 | 1, 7, 3, 5, 4, 6, 8, 9, 10, 11, 12, 13, 15, 16, 19 |
| Tipo: Supervisadas | | | |
| Sesiones de teoría | 50 | 2 | 1, 2, 3, 5, 4, 6, 8, 9, 12, 13 |
| Tipo: Autónomas | | | |
| Preparación de un trabajo en equipo | 50 | 2 | 1, 2, 7, 3, 5, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |

Evaluación

La evaluación de la asignatura se llevará a cabo con dos exámenes (parcial y final) y un trabajo en equipo. La nota final se calculará con la fórmula:

$$NF = 0,5 * NE + 0,5 * NT$$

donde NT es la nota del trabajo en equipo y NE se calcula como:

$$NE = \max(0,4 * NP + 0,6 * NF, NF)$$

donde NP es la nota del examen parcial y NF la nota del examen final.

A final de curso habrá una prueba de recuperación para aquellos alumnos tales que NE sea menor que 5. En este caso, la nota final se calculará con la fórmula:

$$NF = 0,5 * NR + 0,5 * NT$$

donde NR es la nota del examen de recuperación.

Para que una actividad sea tenida en cuenta en la evaluación, su nota debe ser igual o superior a 3,5.

Actividades de evaluación

| Título | Peso | Horas | ECTS | Resultados de aprendizaje |
|-------------------|------|-------|------|---|
| Examen final | 30% | 0 | 0 | 2, 3, 4, 6, 8, 9, 13, 15, 18 |
| Examen parcial | 20% | 0 | 0 | 2, 3, 4, 6, 8, 9, 13, 15, 18 |
| Trabajo en equipo | 50% | 0 | 0 | 1, 2, 7, 3, 5, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |

Bibliografía

Bibliografía básica:

- An Introduction to Statistical Learning with Applications in R - Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

Bibliografía complementaria:

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction - Trevor Hastie, Robert Tibshirani and Jerome Friedman

- Data Science from Scratch - Joel Grus

- Computer Age Statistical Inference: Algorithms, Evidence and Data Science - Trevor Hastie and Bradley Efron