Universitat Autònoma de Barcelona

**2023/2024**

# Aprendizaje Automático 1

Código: 104870
Créditos ECTS: 6

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

## Contacto

Nombre: Juan Ramon Gonzalez Ruiz

Correo electrónico: juanramon.gonzalez@uab.cat

## Idiomas de los grupos

Puede consutarlo a través de este enlace. Para consultar el idioma necesitará introducir el CÓDIGO de la asignatura. Tenga en cuenta que la información es provisional hasta el 30 de noviembre del 2023.

## Prerrequisitos

Esta assignatura 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

Estos son los contenidos de la asignatura*

- Introducción a Tidyverse
- Introducción al aprendizaje automático
- Elastic net, ridge and lasso regression: mejorando la regresión lineal y logística
- Tractamiento de Big Data con R
- La librería caret
- Métodos de aprendizaje automático
    - K-vecinos más próximos
    - Análisis discrimiante
- Métodos para tratar datos no balanceados
- Árboles de decisión
    - Árboles de clasificación
    - Árboles de regresión
    - Bagged trees
    - Random Forest
- Boosting
    - AdaBoost
    - GBM clásico
    - GBM estocástico
    - XGBoost
    - Otros

**A menos que las restricciones impuestas por las autoridades sanitarias obliguen a una priorizacion o reduccion de estos contenidos.*

## Metodología

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

- Clase de teoria: 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 teoria con diversos conjuntos de datos utilizando el lenguaje de programación R.

Se considera que, para cada hora de teoria y prácticas, el alumno deberá dedicar una hora adicional para la preparación y/o finalización de la sesión. Se realizarán preguntas de autoevaluación en el Moodle para consolidar los conocimientos aprendidos en clase

NOTA: La metodología docente propuesta puede experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

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 | | | |
| Sesiones de prácticas | 50 | 2 | 1, 7, 3, 5, 4, 6, 8, 9, 10, 11, 12, 13, 15, 16, 19 |

| | | | |
|---|---|---|---|
| Tipo: Supervisadas | | | |
| Sesiones de teoria | 50 | 2 | 1, 2, 3, 5, 4, 6, 8, 9, 12, 13 |
| Tipo: Autónomas | | | |
| Trabajos semanales + autoevaluación | 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 un examen (final) una serie de trabajos semanales y preguntas de autoevaluación. La nota final se calculará con la fórmula:

NF = 0,5* NE + 0,4 * NP + 0.1 *NA

donde NP es la nota promedio de los trabajos semanales, NA la nota promedio de las preguntas de autoevaluación y NE la nota del examen que tiene que ser un 5 como mínimo.

A final de curso habrá una prueba de recupración para aquellos alumnes tales que NE sea menor que 5 y/o NF menor que 5. En este caso, la nota final se calculará con la fórmula:

NF = 0,7 * NR + 0,3 * NP

donde NR es la nota del examen de recuperación.

Evaluación única:

Se llevará a cabo un examen (4 horas) que permitirá evaluar de manera integral los conocimientos y habilidades adquiridos durante el curso. Este examen estará diseñado para evaluar la capacidad del alumno para aplicar los análisis estadísticos aprendidos y su comprensión de los conceptos teóricos.

El examen constará de dos partes principales: análisis estadístico y preguntas teóricas. En la sección de análisis estadístico, se proporcionarán datos relevantes que requerirán del alumno la aplicación de las técnicas y herramientas estadísticas aprendidas durante el curso. Se espera que el alumno realice los siguientes pasos:

1. Identificación del problema: El alumno deberá comprender la naturaleza de los datos y los objetivos de análisis.

2. Selección y aplicación de técnicas: El alumno utilizará los conocimientos adquiridos para seleccionar y aplicar las técnicas estadísticas apropiadas para analizar los datos. Esto puede incluir la determinación de medidas de tendencia central, dispersión, correlación, regresión, pruebas de hipótesis, entre otros.

3. Interpretación de los resultados: Una vez realizados los análisis, el alumno deberá interpretar los resultados de manera adecuada, explicando su significado en el contexto del problema planteado.

La segunda parte del examen consistirá en preguntas teóricas que requerirán respuestas escritas. Estas preguntas estarán relacionadas con los conceptos fundamentales de la estadística, su aplicabilidad en diferentes situaciones y su importancia en la toma de decisiones. El alumno deberá demostrar su comprensión de los conceptos y su capacidad para explicarlos de manera clara y coherente.

La evaluación de este examen se realizará teniendo en cuenta varios criterios:

1. Precisión y corrección en los análisis: Se evaluará la capacidad del alumno para realizar los análisis estadísticos de manera precisa y correcta, seleccionando las técnicas adecuadas y utilizando los procedimientos correctos.

2. Interpretación de resultados: Se valorará la capacidad del alumno para interpretar y explicar de manera coherente los resultados obtenidos en los análisis estadísticos realizados.

3. Completitud de respuestas teóricas: Se evaluará la capacidad del alumno para proporcionar respuestas claras y completas a las preguntas teóricas, demostrando un dominio de los conceptos y su aplicación.

4. Organización y claridad en la presentación: Se tendrá en cuenta la organización general del examen, la claridad de las respuestas escritas y la calidad de la presentación de los resultados estadísticos.

NOTA: La evaluación propuesta puede experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

## Actividades de evaluación continuada

| Título | Peso | Horas | ECTS | Resultados de aprendizaje |
|---|---|---|---|---|
| Autoevaluación | 10% | 0 | 0 | 1, 7, 5, 8, 9, 13, 15, 19 |
| Examen final | 50% | 0 | 0 | 2, 3, 4, 6, 8, 9, 13, 15, 18 |
| Prácticas | 40% | 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

- El bookdown de la asignatura: https://isglobal-brge.github.io/Aprendizaje_Automatico_1/

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


## Software

Tanto la teoría como la parte práctica se hará con R.