

23-24

MÁSTER UNIVERSITARIO EN
INGENIERÍA Y CIENCIA DE DATOS

GUÍA DE ESTUDIO PÚBLICA



APRENDIZAJE AUTOMÁTICO II

CÓDIGO 31110130

UNED

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN	
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA	
EQUIPO DOCENTE	
HORARIO DE ATENCIÓN AL ESTUDIANTE	
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE	
RESULTADOS DE APRENDIZAJE	
CONTENIDOS	
METODOLOGÍA	
SISTEMA DE EVALUACIÓN	
BIBLIOGRAFÍA BÁSICA	
BIBLIOGRAFÍA COMPLEMENTARIA	
RECURSOS DE APOYO Y WEBGRAFÍA	

Nombre de la asignatura	APRENDIZAJE AUTOMÁTICO II
Código	31110130
Curso académico	2023/2024
Título en que se imparte	MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA DE DATOS
Tipo	CONTENIDOS
Nº ETCS	4
Horas	100.0
Periodo	SEMESTRE 2
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

En esta segunda asignatura dedicada al aprendizaje automático, abordaremos algunas de las técnicas más recientes y más exitosas de la inteligencia artificial, aquellas que, en los últimos años, han protagonizado algunos de los avances más significativos en la ciencia de datos.

La asignatura tiene cuatro bloques diferenciados: así, construyendo sobre lo aprendido en las asignaturas Modelado estadístico de datos y Aprendizaje automático I, en el primer bloque se expondrán y se trabajarán los modelos basados en la idea del *bootstrap aggregation* o *bagging*, en especial los bosques aleatorios. En el segundo bloque, se verán estrategias de intensificación como *boosting*. El tercer bloque está dedicado a otros paradigmas de agregación de modelos, mientras que el cuarto y último bloque cubre técnicas de aprendizaje no supervisado o autoorganizativo, esenciales para aquellas situaciones en las que interesa encontrar estructuras subyacentes u ocultas en un conjunto de datos.

Todos estos contenidos se cubrirán de forma necesariamente escueta: las ciencias que les dan lugar y base son ciencias recientes pero extensas. El enfoque será eminentemente práctico, pero además de que el alumnado pueda incorporar a su perfil profesional las herramientas que se exponen, se pretende que alcancen una comprensión profunda de su funcionamiento.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Los contenidos de esta asignatura están concebidos teniendo en cuenta una cierta secuencialidad con las asignaturas Modelado estadístico de datos y Aprendizaje automático I, cuyo aprovechamiento será de mucha utilidad al alumnado.

Debido entre otras cosas a su novedad, la práctica totalidad de la literatura sobre el tema está en inglés, se hará necesario un nivel de lectura suficiente como para entender contenidos técnicos en dicha lengua.

Se promoverá el uso de software libre para la realización de las actividades y las prácticas propuestas.

EQUIPO DOCENTE

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

LUIS MANUEL SARRO BARO
lsb@dia.uned.es

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

JOSE MANUEL CASTILLO CARA (Coordinador de asignatura)
manuelcastillo@dia.uned.es

HORARIO DE ATENCIÓN AL ESTUDIANTE

El equipo docente atenderá preferentemente al alumnado a través de los foros del curso virtual. De esta manera, las respuestas a cada estudiante particular serán de utilidad al resto. Por supuesto, también atenderá dudas a través del correo electrónico, si se prefiere.

El alumnado puede proponer videoconferencias en grupo para aclarar aspectos particulares del temario, pero dichas videoconferencias deberán prepararse con anterioridad a través de los foros. La interacción habitual consiste en que unx o varixs estudiantes proponen una videoconferencia sobre un tema particular, definimos el problema de forma colectiva y las cuestiones sobre las que queremos debatir o que pretendemos aclarar y finalmente se fija un día y una hora que sea lo más compatible con las restricciones temporales de equipo docente y estudiantes.

Los horarios de guardia del equipo docente son:

- Luis M. Sarro Baro (lsb@dia.uned.es): Lunes, de 10:00 a 14:00. Teléfono: 913988715
- Manuel Castillo Cara (manuelcastillo@dia.uned.es): Martes, de 10:00 a 14:00. Teléfono: 913989688

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

COMPETENCIAS BÁSICAS

CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y

sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

COMPETENCIAS GENERALES

CG1 - Identificar los métodos apropiados para la solución de problemas asociados a la ciencia de datos y la analítica de información

CG2 - Ser capaz de aplicar diferentes técnicas de aprendizaje máquina, seleccionando el algoritmo óptimo que genere modelos precisos y permita el desarrollo de soluciones predictivas en diferentes ámbitos de uso

CG5 - Utilizar las habilidades de científico de datos y/o ingeniero de datos en entornos de trabajo multidisciplinares y ser capaz de distinguir/organizar las diferentes actividades de los roles en dicho entorno

COMPETENCIAS ESPECÍFICAS

CE2 - Desarrollar aplicaciones/servicios/scripts orientados a la analítica de datos y analizar el uso de diferentes librerías para el desarrollo e implementación de métodos numéricos, algoritmos y modelos asociados a los datos

CE5 - Desarrollar modelos de aprendizaje máquina (Machine Learning) basados en las diferentes categorías de clasificación: supervisada, no supervisada y semi-supervisada

CE6 - Diseñar mecanismos de evaluación de modelos de aprendizaje y comprender las métricas usadas para dicha evaluación

RESULTADOS DE APRENDIZAJE

Tras el estudio de esta asignatura, el alumnado habrá adquirido las siguientes destrezas:

- Reconocer los modelos de potenciación (“boosting”) y sus variantes.
- Aplicar la potenciación de árboles para problemas de clasificación y regresión.
- Identificar los conceptos de regularización, submuestreo y reducción.
- Explicar los fundamentos teóricos de los bosques aleatorios, así como los conceptos básicos de muestras “fuera del saco”, importancia de variables y sobreajuste.
- Aplicar y analizar bosques aleatorios para problemas de regresión y clasificación
- Describir los principios del aprendizaje conjunto y sus aplicaciones.
- Identificar los problemas de aprendizaje no supervisado y poder elegir de entre las técnicas disponibles la más adecuada.
- Aplicar análisis de agrupamiento automático mediante k-medias y sus variantes.
- Explicar los mapas autoorganizados y sus utilidades.

CONTENIDOS

Tema 1: Bosques aleatorios

En este tema abordaremos el estudio del método estadístico de *bootstrap* y su aplicación para reducir la varianza de algoritmos de aprendizaje automático, en la técnica de agregación de *bootstrap* llamada *bagging*. De ahí, introduciremos los bosques aleatorios, un modelo enormemente exitoso y sencillo de utilizar, pero con particularidades matemáticas que lo hacen muy interesante.

Tema 2: Intensificación de gradiente

La intensificación o *boosting* es una familia de algoritmos que construyen modelos generales "fuertes" a partir de modelos parciales (o "aprendices débiles"). En este tema abordaremos este paradigma de agregación de modelos que permiten mejorar las predicciones de cualquier algoritmo de aprendizaje automático.

Tema 3: Aprendizaje conjunto

En este tema abordaremos las técnicas de reducción de varianza englobadas bajo el término *ensemble learning*. Para ello haremos uso de la idea de compromiso entre sesgo y varianza, y mostraremos cómo es posible reducir la varianza de modelos que suelen presentarla alta mediante la agregación de conjuntos de esos mismos modelos.

Tema 4: Aprendizaje no supervisado

En este tema abordaremos el estudio de aquellas técnicas de aprendizaje automático que permiten abordar la búsqueda de patrones en conjuntos de datos no etiquetados. Veremos la taxonomía de estas técnicas y las diferentes métricas de evaluación interna y externa de los resultados.

METODOLOGÍA

Esta asignatura ha sido diseñada para la enseñanza a distancia. Por tanto, el sistema de enseñanza-aprendizaje estará basado en gran parte en el estudio independiente o autónomo del alumnado. Para ello, este contará con diversos materiales que permitirán su trabajo autónomo y la Guía de Estudio de la asignatura, que incluye orientaciones para la realización de las actividades prácticas. Asimismo, mediante la plataforma virtual de la UNED, existirá un contacto continuo entre el equipo docente y el alumnado (así como entre el propio alumnado) a través de los foros, algo que es esencial en la enseñanza no presencial. El estudio de esta asignatura se realizará a través de los materiales y enlaces que el Equipo

Docente publicará en el curso virtual.

Las actividades formativas para el estudio de la asignatura son las siguientes:

- Estudios de contenidos (50 horas)
- Actividades en la plataforma virtual (5 horas)
- Prácticas evaluables (45 horas)
- Total: 100 horas

Los medios necesarios para el aprendizaje son:

1. Materiales teórico-prácticos seleccionados por el Equipo Docente para cubrir los conceptos básicos del temario, que forman parte de la bibliografía básica.

2. Bibliografía complementaria. El estudiante puede encontrar en ella información adicional para completar su formación.

3. Curso Virtual de la asignatura, donde el estudiante encontrará:

- Una **guía de la asignatura** en la que se hace una descripción detallada del plan de trabajo propuesto.
- Un **calendario** con la distribución temporal orientativa de los temas propuesta por el Equipo Docente.
- Enunciado de las **actividades teórico-prácticas** propuestas y una zona donde depositar los entregables asociados a dichas actividades.
- Los **foros** por medio de los cuales el Equipo Docente aclarará las dudas de carácter general y que se usarán también para comunicar todas aquellas novedades que surjan a lo largo del curso. Éste será el principal medio de comunicación entre los distintos participantes en la asignatura.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen	Examen tipo test
Preguntas test	10
Duración del examen	60 (minutos)
Material permitido en el examen	

Se permite el uso de cualquier tipo de material.

Criterios de evaluación

Cada cuestión tendrá un valor de un punto en caso de contestar de forma correcta, y restará 0.3 puntos en caso de contestarse de forma errónea.

% del examen sobre la nota final	30
Nota del examen para aprobar sin PEC	
Nota máxima que aporta el examen a la calificación final sin PEC	3
Nota mínima en el examen para sumar la PEC	4,5
Comentarios y observaciones	

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad

Si

Descripción

La prueba presencial se trata de un cuestionario de 10 preguntas teórico-prácticas que versarán sobre los contenidos de la asignatura. Cada cuestión tendrá un máximo de cuatro respuestas posibles, siendo sólo correcta una.

Criterios de evaluación

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?

Si,PEC no presencial

Descripción

Se pedirá al alumnado que realice varios trabajos de aplicación práctica del contenido de la asignatura. El objetivo de estos trabajos es que el alumnado muestre que ha interiorizado los contenidos y que es capaz de ponerlos en juego y evaluar los resultados con capacidad crítica.

Criterios de evaluación

Los definidos en la rúbrica que acompaña cada enunciado.

Ponderación de la PEC en la nota final 60%

Fecha aproximada de entrega

Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s?

Si,no presencial

Descripción

Se valorará la participación del alumnado en el curso virtual, particularmente si no se reduce a preguntar dudas sino que además se intenta resolver las preguntas de las demás y se traba conversación sobre los temas que se propongan.

Criterios de evaluación

Ponderación en la nota final 10%

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La nota final se obtiene como la suma ponderada de la nota del examen, de las pruebas de evaluación continua y de la evaluación de la participación en el foro de la asignatura. Para aprobar, es necesario sacar más de un 4.5 en cada una de las pruebas de evaluación continua así como en el examen.

En caso de no aprobar en junio, en septiembre se permitirá la entrega de todas aquellas PECs que no hayan sido aprobadas, reteniéndose la nota de aquellas que sí.

BIBLIOGRAFÍA BÁSICA

ISBN(13):9780387848570

Título:THE ELEMENTS OF STATISTICAL LEARNING (Second)

Autor/es:Tibshirani, Robert J. ; Hastie, Trevor ; Friedman, Jerome ;

Editorial:Springer

El libro está disponible para su consulta de forma gratuita en su web.

BIBLIOGRAFÍA COMPLEMENTARIA

Como bibliografía complementaria, en la guía de estudio y en el foro se propondrán lecturas de artículos científicos en los que se presenten las técnicas estudiadas. Por ejemplo, se sugerirá la lectura detenida del artículo *Statistical Modeling: the Two Cultures*, que Leo Breiman publicó en el número 16 de la revista *Statistical Science*.

RECURSOS DE APOYO Y WEBGRAFÍA

El alumnado dispondrá de los siguientes recursos de apoyo al estudio:

- **Guía de la asignatura.** Incluye el plan de trabajo y orientaciones para su desarrollo. Esta guía será accesible desde el curso virtual.
- **Curso virtual.** A través de esta plataforma el alumnado tiene la posibilidad de consultar información de la asignatura, realizar consultas al Equipo Docente a través de los foros correspondientes, consultar e intercambiar información con el resto de personas matriculadas.
- **Biblioteca.** El estudiante tendrá acceso tanto a las bibliotecas de los Centros Asociados como a la biblioteca de la Sede Central y en ellas podrá encontrar un entorno adecuado para el estudio, así como distinta bibliografía que podrá ser de utilidad durante el proceso de aprendizaje. Además, desde la biblioteca digital de la UNED, el estudiante tendrá acceso a O'Reilly for Higher Education (New Safari), una biblioteca digital con más de 30.000 libros técnicos en constante actualización y a todas las revistas científicas a las que la UNED está suscrita.

- Sesiones de acompañamiento.** El profesorado convocará sesiones de acompañamiento por videoconferencia, a fin de que el alumnado pueda plantear dudas o intercambiar pareceres.
-

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.