26-27

GUÍA DE ESTUDIO PÚBLICA

APRENDIZAJE AUTOMÁTICO II

CÓDIGO 31110130

UNED

26-27

APRENDIZAJE AUTOMÁTICO II

CÓDIGO 31110130

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
IGUALDAD DE GÉNERO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

En esta segunda asignatura dedicada al aprendizaje automático, abordaremos algunas de las técnicas más recientes y exitosas de la Inteligencia Artificial (IA), responsables de muchos de los avances más relevantes en ciencia de datos en los últimos años.
La asignatura se estructura en cuatro bloques diferenciados y se construye sobre lo aprendido previamente en las asignaturas de *Modelado Estadístico de Datos* y *Aprendizaje Automático I*. El primer bloque se centra en los modelos basados en la agregación por *Bootstrap* (conocida como *Bootstrap Aggregation* o *Bagging*), en especial Bosques Aleatorios. En el segundo bloque, se explorarán estrategias de intensificación (*Boosting*), y en el tercero, otros paradigmas de combinación como el apilamiento de modelos (*Stacking*). Estos tres primeros bloques conforman una visión integrada de la **taxonomía ensemble de agregación de modelos**, una de las líneas más activas y efectivas del aprendizaje automático actual. El cuarto y último bloque está dedicado al *Aprendizaje No Supervisado* (*Unsupervised Learning*), clave en escenarios donde se busca descubrir estructuras subyacentes u ocultas en los datos.
Todos estos contenidos se cubrirán de forma necesariamente concisa, ya que las disciplinas en las que se basan son recientes pero muy amplias. El enfoque será eminentemente práctico, con el objetivo de que el alumnado no solo incorpore estas herramientas a su perfil profesional, sino que también desarrolle una comprensión profunda de su funcionamiento.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Los contenidos de esta asignatura se han diseñado como una continuación natural de las asignaturas *Modelado Estadístico de Datos* y *Aprendizaje Automático I*, por lo que su seguimiento y aprovechamiento previos resultan altamente recomendables.
Dado que la mayoría de la literatura especializada en estos temas es reciente y se publica en inglés, se requiere un nivel de lectura suficiente para comprender documentación técnica en dicho idioma.
Asimismo, se fomentará el uso de software libre en las actividades y prácticas propuestas a lo largo del curso.

EQUIPO DOCENTE

Nombre y Apellidos	LUIS MANUEL SARRO BARO
Correo Electrónico	lsb@dia.uned.es
Teléfono	91398-8715
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

Nombre y Apellidos	JOSE MANUEL CASTILLO CARA (Coordinador/a de asignatura)
Correo Electrónico	manuelcastillo@dia.uned.es
Teléfono	
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

Nombre y Apellidos	DAVID BERNAL CASAS
Correo Electrónico	dbernal@dia.uned.es
Teléfono	
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

HORARIO DE ATENCIÓN AL ESTUDIANTE

El equipo docente atenderá preferentemente al alumnado a través de los foros del curso virtual, con el fin de que las respuestas a dudas individuales puedan ser también de utilidad para el resto del grupo. No obstante, si se prefiere, también se podrá contactar por correo electrónico.

Horarios de atención directa del equipo docente:

•**Manuel Castillo Cara** (manuelcastillo@dia.uned.es): Martes, de 10:00 a 14:00.

•**Luis M. Sarro Baro** (lsb@dia.uned.es): Lunes, de 10:00 a 14:00.

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: Potenciación ("Boosting") y árboles aditivos

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	20
Duración del examen	60 (minutos)

Material permitido en el examen

No se permite el uso de ningún tipo de material.

Criterios de evaluación

Cada cuestión tendrá un valor de 0,5 punto en caso de contestar de forma correcta, y restará 0,2 puntos en caso de contestarse de forma errónea.

% del examen sobre la nota final	50
Nota del examen para aprobar sin PEC	
Nota máxima que aporta el examen a la calificación final sin PEC	5
Nota mínima en el examen para sumar la PEC	4,5

Comentarios y observaciones

Para que las PECs se sumen a la nota final, es necesario obtener al menos un 4.5 en el examen.

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad	Si

Descripción

La prueba presencial se trata de un cuestionario de 20 preguntas teórico-prácticas que versarán sobre los contenidos de la asignatura.
Cada pregunta tiene 4 posibles respuestas, de las cuales solamente una es correcta.
Las respuestas erróneas descuentan 0,2 puntos.
Las preguntas con más de una respuesta marcada o en blanco no suman ni restan puntos.

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.

Por tanto, estas pruebas serán de naturaleza teórico/práctica para realizar en el transcurso del semestre. No es necesario asistir a ningún centro asociado para realizarlas. La entrega de cada PEC será únicamente por vía telemática dentro del curso virtual en el plazo indicado en el mismo y a lo largo del curso.

Criterios de evaluación

Los definidos en la rúbrica que acompaña cada enunciado de las PECs.

Ponderación de la PEC en la nota final: La PEC1 tendrá una ponderación del 30% y la PEC2 del 20% sobre la nota final. En conjunto, ambas actividades (PEC1 + PEC2) representarán el 50% de la calificación total de la asignatura.

Fecha aproximada de entrega: Meses de abril y mayo

Comentarios y observaciones

Las prácticas de laboratorio se calificarán de 1 a 10 puntos, siendo el 10 la máxima puntuación. Se debe obtener un mínimo de 4,5 por separado en cada una de las prácticas para que se contabilicen en la nota final para poder aprobar la asignatura. **En caso de haber aprobado las PECs pero no haber aprobado el Examen, la nota de estas se guardará para la convocatoria extraordinaria de septiembre en el curso presente.**

Estas fechas de entrega son aproximadas. En el curso virtual se organizarán tareas para la entrega de dichas prácticas con las fechas exactas.

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? No

Descripción

Criterios de evaluación

Ponderación en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La calificación final **solo se calculará si se obtiene al menos un 4,5 sobre 10** tanto en el Examen como en **cada una de las PECs**. En caso contrario, **no se podrá superar la asignatura**, independientemente de la media ponderada.
Por tanto, la nota final se calcula con la siguiente fórmula:
Nota Final = 50% Examen + 30% PEC1 + 20% PEC2
En caso de no superar la asignatura en la convocatoria de junio, en la convocatoria extraordinaria de septiembre se permitirá la entrega de todas aquellas PECs no aprobadas, manteniéndose la calificación de las ya superadas.

BIBLIOGRAFÍA BÁSICA

ISBN(13): 9780387848570
Título: THE ELEMENTS OF STATISTICAL LEARNING Second edición
Autor/es: Tibshirani, Robert J.; Hastie, Trevor; Friedman, Jerome
Editorial: Springer

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

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.