

A7. Métodos supervisados

**MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN
INTELIGENCIA ARTIFICIAL**

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.



DATOS GENERALES

Breve descripción

Esta asignatura ahondará en las principales técnicas dentro del campo de la clasificación supervisada. Técnicas para el aprendizaje de distintos tipos de modelos (desde procedimientos clásicos hasta los más recientes), técnicas de selección de variables multivariadas específicas para entornos supervisados, procedimientos para la comparativa estadística entre los porcentajes de acierto de distintos modelos, etc. Todo ello condimentado con aplicaciones y distintas tecnologías.

Título asignatura

A7. Métodos supervisados

Código asignatura

102470

Curso académico

2018-19

Planes donde se imparte

[MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL](#)

Créditos ECTS

4,5

Carácter de la asignatura

OPTATIVA

Duración

Anual

Idioma

Castellano

CONTENIDOS

Contenidos

Desde sus inicios la Inteligencia Artificial ha prestado atención a cómo hacer que aprendan los ordenadores, más que a programar tareas simples. Este campo, denominado aprendizaje automático, se aplica a tareas que se desea que hagan los ordenadores y sobre las que tenemos más datos que conocimientos detallados de cómo han de llevarse a cabo.

Se presentan métodos de aprendizaje supervisado que son altamente utilizados en aplicaciones reales cuyo planteamiento es un problema de optimización y cuyo fin es resolver una tarea de clasificación, regresión y ranking. Se incluyen métodos tales como Regresión Logística, modelos Bayesianos, Redes de Neuronas Artificiales, Máquinas de Vectores Soporte (SVM) y métodos de Factorización de matrices. También se estudiarán en esta materia las técnicas de meta-aprendizaje:

- Aprendizaje como optimización.
- Regresión y Clasificación.
- Regresión Logística, Métodos Bayesianos, Redes neuronales, Métodos Kernel (SVM), Modelos lineales generalizados, Factorización.
- Meta-aprendizaje: boosting, bagging, random-forest.

Unidades

1. Módulo 1: Introducción a la clasificación supervisada y aplicaciones
2. Módulo 2: Principales algoritmos de aprendizaje supervisado
3. Módulo 3: Uso de tests estadísticos para la comparativa de modelos supervisados
4. Ejercicio final: flujo de análisis en clasificación supervisada

Elabora y estudia el material de manera secuencial, siguiendo el orden de unidades-temas propuesto. Realiza los ejercicios de cada tema-unidad tras estudiar la teoría correspondiente a éste. Anunciaremos los plazos de entrega de los ejercicios durante el curso.

RESULTADOS DE APRENDIZAJE Y DE FORMACIÓN

Generales

CG1 - Entender los conceptos, los métodos y las aplicaciones de la inteligencia artificial.

CG3 - Gestionar de manera inteligente los datos, la información y su representación.

Específicas

CE2 - Aplicar las técnicas de aprendizaje automático utilizando la metodología de validación y presentación de resultados más apropiada en cada caso.

CE5 - Analizar las fuentes documentales propias del ámbito de la investigación en Inteligencia Artificial para poder determinar cuáles de ellas son relevantes en la resolución de problemas concretos.

PLAN DE APRENDIZAJE

Actividades formativas

A1 - **Sesiones presenciales virtuales (clases en vídeo)**: visionado del material audiovisual que constituye las lecciones de la asignatura. Se asume 1.5 veces el tiempo real de vídeo, puesto que el estudiante deberá parar, repetir, etc. algunas secuencias (17 horas).

A2 - **Trabajos individuales**: realización de trabajos prácticos individuales. Se ofrecen una serie de trabajos optativos a los alumnos. Sólo reflejamos la cantidad de trabajo referente al trabajo obligatorio (50 horas).

A3 - **Trabajo autónomo**: estudio del material básico (videos y transparencias), lecturas complementarias y otros contenidos. Estudio encaminado al test de respuestas múltiples (32 horas).

A4 - **Foros y chats**: participación, lectura y contestación de cuestiones y temas para la discusión general (8 horas).

A5 - **Tutorías**: consultas y resolución de dudas, aclaraciones, etc (5,5 horas).

Puede consultar en este enlace el [Cronograma de Carga de Trabajo](#).

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

E1 - **Valoración de los cuestionarios de evaluación:** test de 20 preguntas de respuestas múltiples sobre las lecciones (vídeos-transparencias). El peso en la nota final de este apartado es del 40% del total.

E2 - **Valoración de la participación en foros y chats:** se valorará el nivel de participación/debate de los estudiantes. El peso en la nota final de este apartado es del 10% del total.

E3 - **Valoración de los trabajos individuales:** el alumno programará un "pipeline"-flujo que cubra las principales etapas de un análisis supervisado. El alumno comentará las etapas del "pipeline", las decisiones tomadas, y analizará los resultados: a partir de una base de datos escogida por el alumno desde la plataforma "kaggle.com". Para guiarle en la creación del "pipeline", dispondrá de uno básico ofrecido por los profesores: el alumno deberá variarlo y enriquecerlo. El peso en la nota final de este apartado es del 50% del total. Asimismo, se ofrecerá al alumno una serie de trabajos individuales optativos, cuya realización computará en la nota del apartado E3.

Para aprobar el apartado E1 son necesarios 20 de un máximo de 40 puntos. Para aprobar el apartado E3 son necesarios 25 puntos de un máximo de 50 puntos. Para aprobar la asignatura, el alumno debe aprobar los apartados E1 y E3.

Calendario de exámenes

Para la **convocatoria ordinaria**, habrá 3 fechas de entrega de trabajos final de curso. Los alumnos podrán entregar sus trabajos en cualquier momento, pero sólo en estas fechas se recogerán y evaluarán los que se hayan entregado. Las fechas serán: 20/12/18 - 27/2/19 - 31/5/19

Habrà una **convocatoria extraordinaria** en todas las asignaturas. Para su evaluación, la entrega de trabajos se cerrará el día: 01/07/19

Para los **Trabajos Fin de Master** habrá dos convocatorias: 15/7/19 (convocatoria ordinaria) y 15/9/19 (convocatoria extraordinaria)

PROFESORADO

Profesor responsable

Larrañaga Múgica, Pedro

*Licenciado en Matemáticas. Doctor en Informática.
Catedrático de Universidad (ciencias de la Computación e Inteligencia Artificial).
Universidad Politécnica de Madrid.*

Profesorado

Inza Cano, Iñaki

*Doctor en Informática.
Profesor Pleno -- equivalente laboral a Catedrático de Universidad.
Área: Ciencias de la Computación e Inteligencia Artificial.
UPV-EHU.*

Mihaljevic , Bojan

*Doctor en Inteligencia Artificial.
Profesor Permanente Laboral.
Universidad Politécnica de Madrid.*

Bielza Lozoya, Concha

*Doctora en Informática.
Catedrática de Universidad (área de Estadística e Investigación Operativa).
Universidad Politécnica de Madrid / Escuela Técnica Superior de Ingenieros Informáticos.*

Atienza González, David

*Teaching Assistant
Contratado Predoctoral
Universidad Politécnica de Madrid*

Hernández González, Jerónimo

Teaching Assistant

Investigador Postdoctoral en Ciencias de la Computación e Inteligencia Artificial

Universidad del País Vasco

HORARIO

Horario

Todas las asignaturas estarán en la plataforma a disposición de los estudiantes desde octubre hasta julio.

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Bibliografía

M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim (2014). Do we need hundreds of classifiers to solve real world classification problems? Journal of Machine Learning Research 15, 3133-3181. [\[link\]](#)

G. Brown, A. Pocock, M-J. Zhao, M. Luján (2012). Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. Journal of Machine Learning Research 13, 27-66. [\[link\]](#)

C. Bielza, P. Larrañaga (2014). Discrete Bayesian network classifiers: a survey. ACM Computing Surveys 47(1), 1-43. [\[link\]](#)

Y. Saeys, I. Inza, P. Larrañaga (2007). A review of feature selection techniques in bioinformatics. Bioinformatics 23(19), 2507-2517. [\[link\]](#)

M. Kuhn, K. Johnson (2013). Applied Predictive Modeling. Springer. [\[link\]](#)

Alpaydin. Introduction to Machine Learning. MIT Press.

R. Duda, P.E. Hart, D.G. Stork. Pattern Classification. Wiley. 2001.

J. Hernández-Orallo, M.J. Ramírez, C. Ferri. Introducción a la Minería de Datos. Pearson Educación. 2004.

L. Kuncheva. Combining Pattern Classifiers. Wiley. 2004.

S. Sharma. Applied Multivariate Techniques. Wiley. 1996.

A. Webb. Statistical Pattern Recognition. Wiley. 2002.

I. Witten, E. Frank. Data Mining. Morgan Kaufmann. 2ª ed. 2005.

Castillo, E., Gutiérrez, J.M., Hadi, A.S. (1997) Expert Systems and Probabilistic Network Models. Springer, New York. Versión en español, disponible en la red: Sistemas Expertos y Modelos de Redes Probabilísticas, Academia de Ingeniería, Madrid.

Koller, D, Friedman, N. (2009) Probabilistic Graphical Models Principles and Techniques. The MIT Press.

Koski, T., Noble, J. (2009) Bayesian Networks: An Introduction, Wiley.

Murphy, K.P. (2012) Machine Learning: A Probabilistic Perspective. The MIT Press.

Neapolitan, R., (2004) Learning Bayesian Networks, Prentice Hall.

Webs-enlaces

Software

- paquete caret (R): "classification and regression training with R" [\[link\]](#) [\[link CRAN\]](#)
- bnClassify (R): "learning discrete Bayesian network classifiers from data" [\[link en GitHub\]](#)
- a list of classification software for data mining and analytics [\[link\]](#)
- Bayesian networks and Bayesian classifier software [\[link\]](#)
- MEKA: a multilabel extension to WEKA [\[link\]](#)
- MULAN: a Java library for multilabel learning [\[link\]](#)

Lecturas - Reading

- 60+ free books on big data, data science, data mining, machine learning, python, R and more. [\[link\]](#)

Colección con aplicaciones de la minería de datos [\[link\]](#)