

A3. Ciencia de datos y aprendizaje automático

**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

Título asignatura

A3. Ciencia de datos y aprendizaje automático

Código asignatura

102120

Curso académico

2016-17

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

En esta materia se estudiarán los fundamentos del proceso para descubrir patrones en conjuntos de datos. Se estudiarán algoritmos y técnicas de preparación de los datos, algunos algoritmos básicos de aprendizaje automático y métodos de evaluación de estos algoritmos:

- Objetivos y aplicaciones de la ciencia de datos.
- Preprocesamiento de datos: Selección de variables, discretización, selección de instancias, valores imperfectos (ruido, datos perdidos).
- Técnicas de validación: entrenamiento, hold-out, cross-validation, etc.
- Algoritmos de aprendizaje supervisado: árboles de decisión, técnicas de vecinos más cercanos, Naive Bayes, Perceptrón.

Unidades

1. Módulo 1: Introducción a la minería de datos y ciencia de datos

1.1. Motivación

1.2. Minería y ciencia de datos, ejemplos

1.3. El proceso de KDD. CRISP-DM

1.4. Tareas, técnicas y herramientas

2. Módulo 2: Técnicas de validación y evaluación

2.1. Entrenamiento y validación, hold-out, cross-validation

2.2. Evaluación con costes y desbalanceo

2.3. Análisis ROC

3. Módulo 3: Algoritmos básicos de aprendizaje supervisado

3.1. Métodos basados en instancias/vecinos (kNN)

3.2. Árboles de decisión

3.3. Clasificación probabilística - Naive Bayes

3.4. Redes neuronales

4. Módulo 4: Preprocesamiento de datos

4.1. Integración, manipulación y visualización

4.2. Selección de variables

4.3. Discretización

4.4. Selección de instancias (prototipos)

4.5. Valores imperfectos, ruido, datos perdidos

5. Módulo Práctico: Weka, R y Kaggle

COMPETENCIAS

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**: visionado inicial del material audiovisual (vídeos introductorios, presentaciones, animaciones) que se elabore en cada una de las materias y que servirán presentación de cada uno de los temas a los estudiantes (12 horas - 100% presencialidad).

A2 - **Trabajos individuales**: realización de ejercicios, resolución de problemas, realización de prácticas y/o trabajos/proyectos individuales (17 horas - 0% presencialidad).

A3 - **Trabajo autónomo**: estudio del material básico, lecturas complementarias y otros contenidos y estudio (72 horas - 0% presencialidad).

A4 - **Foros y chats**: lanzamiento de cuestiones y temas para la discusión general (5,5 horas - 0% presencialidad).

A5 - **Tutorías**: consultas y resolución de dudas, aclaraciones, etc (6 horas - 100% presencialidad).

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

E1 - Valoración de los cuestionarios de evaluación: los estudiantes realizarán por cada unidad didáctica un cuestionario de evaluación que será objeto de puntuación en la nota final (ponderación mínima 20% y máxima 40%).

E2 - Valoración de la participación en foros y chats: se valorará el nivel de participación/debate de los estudiantes que contará para la nota final (ponderación mínima 10% y máxima 20%).

E3 - Valoración de los trabajos individuales: se valorarán los problemas, proyectos, trabajos realizados y entregados a través de la plataforma, y apoyado en los casos que sea necesario (sobre todo cuando se trate de desarrollo de código) por plataformas de gestión de código como GitHub. También se incluirá el video que el alumno deberá enviar al profesor para cada asignatura (ponderación mínima 40% y máxima 70%).

PROFESORADO

Profesor responsable

Gámez Martín, José Antonio

*Catedrático de Lenguajes y Sistemas Informáticos
Universidad de Castilla-La Mancha*

Profesorado

Del Jesús Díaz, María José

*Catedrática de Ciencias de la Computación e Inteligencia Artificial
Universidad de Jaén*

Hernández Orallo, José

*Catedrático Lenguajes y Sistemas Informáticos
Universidad Politécnica de Valencia*

Charte Ojeda, Francisco

*Profesor Ayudante Doctor de Arquitectura y Tecnologías de Computadoras
Universidad de Jaén*

Martínez Plumed, Fernando

*Teaching Assistant
Investigador Postdoctoral en Informática
Universidad Politécnica de Valencia*

HORARIO

Horario

Las sesiones se desarrollarán en noviembre de 2016.

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Libros

Foster Provost and Tom Fawcett "Data Science for Business: Fundamental principles of data mining and data analytic thinking", O'Reilly Media, 2013

Jeffrey Stanton "Introduction to Data Science", 2012.
https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf

Lars Nielsen, Noreen Burlingame "A simple introduction to data science", 2013 (ultra-short introduction)

Rachel Schutt "Doing data science", O'Reilly 2013

Jiawei Han "Data Mining: Concepts and Techniques".

José Hernández-Orallo, M.José Ramírez-Quintana, Cèsar Ferri, "Introducción a la minería de datos", Pearson 2004

Peter Flach "Machine learning: the art and science of algorithms that make sense of data", Cambridge 2013

Trevor Hastie, Robert Tibshirani , Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series. 2009. ISBN 978-0-387-84858-7

Pang-Ning Tan, Michael Steinbach, Vipin Kumar "Introduction to Data Mining", Addison-Wesley, 2005. ISBN : 0321321367

Xindong Wu, Vipin Kumar "The Top Ten Algorithms in Data Mining". Chapman and Hall/CRC, 2009. ISBN: 9781420089646

Mark Hall, Ian Witten, Eibe Frank "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann Publishers, 2011. ISBN: 978-0-12-374856-0

Salvador García, Julián Luengo, Francisco Herrera "Data Preprocessing in Data Mining" Springer, 2015. ISBN: 978-3-319-10246-7

Enlaces a documentos disponibles en línea (por tema)

Árboles de decisión

Tan, Steinbach y Kumar, 2005. Chapter 4. Classification: Basic Concepts, Decision Trees, and

Model Evaluation. <http://www-users.cs.umn.edu/~kumar/dmbook/ch4.pdf>

Naive Bayes

Tom Mitchell, 2015. Chapter 2. Estimating Probabilities: MLE and MAP.
http://www.cs.cmu.edu/%7Etom/mlbook/Joint_MLE_MAP.pdf

Tom Mitchell, 2015. Chapter 3. Naive Bayes and Logistic Regression.
<http://www.cs.cmu.edu/%7Etom/mlbook/NBayesLogReg.pdf>

Selección de atributos

Isabelle Guyon and André Elisseeff. "An introduction to variable and feature selection". Journal of Machine Learning Research 3:1157-1182, 2003.
www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

Gavin Brown, Adam Pockock, Ming-Jie Zhao, Mikel Luján. "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection". Journal of Machine Learning Research 13:27−66, 2012. <http://www.jmlr.org/papers/volume13/brown12a/brown12a.pdf>

Lab (R and Kaggle)

CRAN manuals: <http://cran.r-project.org/doc/manuals/R-intro.pdf>, <http://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>

Luis Torgo "Data Mining with R", CRC Press 2010

Wikibooks: http://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R,
http://en.wikibooks.org/wiki/R_Programming

Graham Williams: Hands-On Data Science with R, <http://onepager.togaware.com/>

www.kaggle.com

Lab (WEKA)

WEKA software: <http://www.cs.waikato.ac.nz/ml/weka/index.html>

The Weka Wiki: <https://weka.wikispaces.com/>

A presentation demonstrating all graphical user interfaces (GUI) in Weka.
<http://prdownloads.sourceforge.net/weka/weka.ppt>