

Curso Académico: ( 2018 / 2019 )

Fecha de revisión: 01/05/2018 13:47:53

Departamento asignado a la asignatura: Departamento de Teoría de la Señal y Comunicaciones

Coordinador/a: CID SUEIRO, JESUS

Tipo: Obligatoria Créditos ECTS : 6.0

Curso : 3 Cuatrimestre : 2

## REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Para cursar esta asignatura es recomendable haber superado las asignaturas sobre fundamentos matemáticos de primer curso (Cálculo I y II, Álgebra Lineal, Probabilidad y Análisis de datos), las asignaturas relativas a programación y algoritmos (Programación, Estructuras de Datos y Algoritmos, Optimización y Análítica), así como la asignatura Aprendizaje Estadístico.

## OBJETIVOS

CB1: Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB2: 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

CB5: Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.

CE1: Capacidad para la resolución de los problemas matemáticos que puedan plantearse en la ingeniería y ciencia de datos. Aptitud para aplicar los conocimientos sobre: álgebra; geometría; cálculo diferencial e integral; métodos numéricos; algorítmica numérica; estadística y optimización.

CE13: Capacidad para aplicar y diseñar métodos de aprendizaje automático en problemas de clasificación, regresión y agrupación y para tareas de análisis supervisado, no supervisado y por refuerzo.

CE15: Capacidad para diseñar soluciones basadas en aprendizaje automático en aplicaciones en dominios específicos como los sistemas de recomendación, el tratamiento del lenguaje natural, la Web o las redes sociales online.

CE2: Capacidad para identificar correctamente problemas de tipo predictivo correspondientes a unos objetivos y unos datos determinados y emplear los resultados básicos del análisis de regresión como fundamento básico de los métodos de predicción.

CE3: Capacidad para identificar correctamente problemas de clasificación correspondientes a unos objetivos y unos datos determinados y emplear los resultados básicos del análisis multivariante como fundamento básico de los métodos de clasificación, clustering y reducción de la dimensión.

CE4: Capacidad para la modelización matemática, implementación algorítmica y resolución de problemas de optimización relacionados con la ciencia de datos.

CG1: Conocimientos y habilidades adecuados para analizar y sintetizar problemas básicos relacionados con la ingeniería y la ciencia de datos, resolverlos y comunicarlos de forma eficiente.

CG2: Conocimiento de materias básicas científicas y técnicas que capaciten para el aprendizaje de nuevos métodos y tecnologías, así como que le dote de una gran versatilidad para adaptarse a nuevas situaciones.

CG4: Capacidad para la resolución de los problemas tecnológicos, informáticos, matemáticos y estadísticos que puedan plantearse en la ingeniería y ciencia de datos.

CG5: Capacidad para resolver problemas formulados matemáticamente aplicados a diversas materias, empleando algoritmos numéricos y técnicas computacionales.

CG6: Capacidad para sintetizar las conclusiones obtenidas de los análisis realizados y presentarlas de manera clara y convincente tanto por escrito como oralmente.

CT1: Capacidad de comunicar los conocimientos oralmente y por escrito, ante un público tanto especializado como no especializado.

RA1 Haber adquirido conocimientos avanzados y demostrado una comprensión de los aspectos teóricos y prácticos y de la metodología de trabajo en el campo de la ciencias e ingeniería de datos con una profundidad que llegue hasta la vanguardia del conocimiento

RA2 Poder, mediante argumentos o procedimientos elaborados y sustentados por ellos mismos, aplicar sus conocimientos, la comprensión de estos y sus capacidades de resolución de problemas en ámbitos laborales complejos o profesionales y especializados que requieren el uso de ideas creativas e innovadoras

RA3 Tener la capacidad de recopilar e interpretar datos e informaciones sobre las que fundamentar sus conclusiones incluyendo, cuando sea preciso y pertinente, la reflexión sobre asuntos de índole social, científica o ética en el ámbito de su campo de estudio;

RA4 Ser capaces de desenvolverse en situaciones complejas o que requieran el desarrollo de nuevas soluciones tanto en el ámbito académico como laboral o profesional dentro de su campo de estudio;

RA5 Saber comunicar a todo tipo de audiencias (especializadas o no) de manera clara y precisa, conocimientos, metodologías, ideas, problemas y soluciones en el ámbito de su campo de estudio;

RA6 Ser capaces de identificar sus propias necesidades formativas en su campo de estudio y entorno laboral o profesional y de organizar su propio aprendizaje con un alto grado de autonomía en todo tipo de contextos (estructurados o no).

## DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

Este curso se divide en 3 bloques temáticos. El primero concierne el problema de la adaptación y limpieza de una base de datos, paso previo a cualquier aplicación de aprendizaje automático que se quiera abordar. Los dos bloques siguientes abordan dos aplicaciones relevantes para la industria donde las técnicas de aprendizaje automático han supuesto una revolución en su desarrollo. La comprensión de cómo las distintas técnicas de aprendizaje automático vistas a lo largo del grado han de adaptarse para resolver problemas concretos de interés para la industria y la sociedad dotará al alumno de una visión práctica y general de los conocimientos adquiridos.

### PARTE I: TÉCNICAS DE ADAPTACIÓN Y LIMPIEZA DE DATOS

1. Introducción al problema.
2. Organización e integración de bases de datos provenientes de distintas fuentes.
3. Limpieza de datos: caracterización de datos, detección e imputación de datos corruptos. Detección de puntos atípicos.

### PARTE II: PROCESADO DE LENGUAJE NATURAL

4. Procesamiento de texto con modelado de tópicos
5. Procesado secuencial de texto utilizando redes neuronales. Representación vectorial de texto y modelos de traducción automática.

### PARTE III: SISTEMAS DE RECOMENDACIÓN

6. Sistemas de recomendación guiados por contenido.
7. Descomposición mediante Factorización de Matrices. Sistemas de recomendación basados en filtrado colaborativo.

## ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

AF1: CLASES TEÓRICO-PRÁCTICAS. En ellas se presentarán los conocimientos que deben adquirir los alumnos. Estos recibirán las notas de clase y tendrán textos básicos de referencia para facilitar el seguimiento de las clases y el desarrollo del trabajo posterior. Se resolverán ejercicios, prácticas problemas por parte del alumno y se realizarán talleres y prueba de evaluación para adquirir las capacidades necesarias.

AF2: Actualizado a alegación

AF3: TRABAJO INDIVIDUAL O EN GRUPO DEL ESTUDIANTE.

AF9: EXAMEN FINAL. En el que se valorarán de forma global los conocimientos, destrezas y capacidades adquiridas a lo largo del curso.

MD1: CLASE TEORÍA. Exposiciones en clase del profesor con soporte de medios informáticos y audiovisuales, en las que se desarrollan los conceptos principales de la materia y se proporcionan los materiales y la bibliografía para complementar el aprendizaje de los alumnos.

MD2: PRÁCTICAS. Resolución de casos prácticos, problemas, etc. planteados por el profesor de manera

individual o en grupo.

MD3: TUTORÍAS. Asistencia individualizada (tutorías individuales) o en grupo (tutorías colectivas) a los estudiantes por parte del profesor.

#### SISTEMA DE EVALUACIÓN

**Peso porcentual del Examen/Prueba Final:** 60

**Peso porcentual del resto de la evaluación:** 40

SE1: EXAMEN FINAL. En el que se valorarán de forma global los conocimientos, destrezas y capacidades adquiridas a lo largo del curso.

SE2: EVALUACIÓN CONTINUA. En ella se valorarán los trabajos, presentaciones, actuación en debates, exposiciones en clase, ejercicios, prácticas y trabajo en los talleres a lo largo del curso.

#### BIBLIOGRAFÍA BÁSICA

- C.C. Aggarwal Recommender Systems: The Textbook, Springer, 2016
- D. Juravsky, J.H. Martin Speech and Language Processing, Prentice Hall; 2nd edition, 2008
- J. Ham, M. Kamber Data Mining: Concepts and Techniques (3rd. ed), Morgan Kaufman, 2011

#### BIBLIOGRAFÍA COMPLEMENTARIA

- C. Manning, H. Schütze Foundations of Statistical Natural Language Processing, MIT Press, 1999
- K. Murphy Machine Learning: A probabilistic Perspective, The MIT Press, 2012
- M. W. Berry Survey of Text Mining Clustering, Classification, and Retrieval, Springer, 2004
- S. Bird, E. Klein, E. Loper Natural Language Processing with Python, O'Reilly, 2009