Aplicaciones del aprendizaje automático

Curso Académico: (2022 / 2023) **Fecha de revisión: 20/05/2022 00:04:18**

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

Coordinador/a: ARENAS GARCIA, JERONIMO

Tipo: Obligatoria Créditos ECTS : 6.0

Curso : 4 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 Análitica), así como la asignatura de Aprendizaje Estadístico. Asimismo, es conveniente haber cursado las asignaturas de Aprendizaje Máquina I y II.

OBJETIVOS

- Diseñar un modelo de datos adecuado a una tarea de análisis
- Elegir y utilizar correcta y eficientemente uno o varios métodos de análisis de datos incluyendo técnicas estadísticas o algorítmicas
- Evaluar los resultados del análisis y proponer modificaciones al proceso de análisis
- Saber diseñar y aplicar métodos de inferencia no supervisada para modelos con variables latentes
- Saber diseñar y aplicar técnicas de adaptación y limpieza de datos
- Saber diseñar y aplicar métodos de tratamiento de lenguaje natural
- Saber diseñar y aplicar sistemas de recomendación

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 compresión 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.

La asignatura concluye con un bloque final en el que se presentarán dos herramientas de visualización que los alumnos habrán de utilizar en la elaboración de un trabajo final de la asignatura.

PARTE I: TÉCNICAS DE ADAPTACIÓN Y LIMPIEZA DE DATOS

1. Introducción al problema. Representación y visualización de datos.
2. Organización e integración de bases de datos provenientes de distintas fuentes.
3. Extracción y selección de características. Métodos de Análisis de múltiples variables y métodos basados en Información Mutua.
4. Limpieza de datos: caracterización de datos, detección e imputación de datos corruptos. Detección de puntos atípicos.

PARTE II: PROCESADO DE LENGUAJE NATURAL
5. Pipelines para procesamiento de textos. Representación vectorial de textos.
6. Modelado de Tópicos: Latent Semantic Indexing, Latent Dirichlet Allocation.
7. Representación vectorial de texto y modelos de traducción automática usando redes neuronales.

PARTE III: SISTEMAS DE RECOMENDACIÓN
8. Sistemas de recomendación guiados por contenido.
9. Sistemas de recomendación basados en filtrado colaborativo. ALS y Prod2Vec.

CONTENIDO ADICIONAL: HERRAMIENTAS AVANZADAS DE VISUALIZACIÓN DE DATOS
* Visualización de grafos con Gephi
* Herramientas de Inteligencia de Negocio

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:	30
Peso porcentual del resto de la evaluación:	70

La Evaluación Continua supone el 70% de la calificación del alumno y constará de los siguientes elementos:

* 3 Pruebas sobre los ejercicios de laboratorio (30%): resolución de ejercicios similares a los planteados en los notebooks de la asignatura utilizando python.
* Proyecto final (40%)

El examen final (30%) consistirá en una prueba escrita sobre los contenidos teóricos y prácticos de la asignatura

Para la evaluación extraordinaria, los alumnos realizarán un examen final por valor de 6 puntos (prueba escrita + laboratorio) y, adicionalmente, se les propondrá un nuevo proyecto final por valor de 4 puntos.

BIBLIOGRAFÍA BÁSICA

- null Data Visualization with Python for Beginners: Visualize Your Data using Pandas, Matplotlib and Seaborn, AI Publishing LLC, 2020

- C.C. Aggarwal Recommender Systems: The Textbook, Springer, 2016

- D. Juravsky, J.H. Martin Speech and Language Processing, Prentice Hall; 2nd edition, 2008

- J. Eisenstein Introduction to Natural Language Processing, MIT Press, 2019

- J. Ham, M. Kamber Data Mining: Concepts and Techniques (3rd. ed), Morgan Kaupfman, 2011

- S. Bird, E. Klein, E. Loper Natural Language Processing with Python, O'Reilly Media, 2009

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

RECURSOS ELECTRÓNICOS BÁSICOS

- . Pandas Tutorials: https://pandas.pydata.org/pandas-docs/stable/getting_started/intro_tutorials/

- J. Arenas-García, J. Cid-Sueiro, V. Gómez-Verdejo . Introductory Notebooks on Machine Learning topics.: https://github.com/ML4DS/ML4all