



Asignatura	MÉTODOS MATEMÁTICOS APLICADOS AL DESARROLLO DE SISTEMAS Y SERVICIOS DE INTERNET		
Materia	SISTEMAS, APLICACIONES Y SERVICIOS DE INTERNET		
Módulo	TECNOLOGÍAS INFORMÁTICAS		
Titulación	MASTER EN INGENIERIA INFORMÁTICA		
Plan	510	Código	53191
Periodo de impartición	1 ^{er} CUATRIMESTRE	Tipo/Carácter	OPTATIVA
Nivel/Ciclo	MASTER	Curso	SEGUNDO
Créditos ECTS	3 ECTS		
Lengua en que se imparte	CASTELLANO		
Profesor/es responsable/s	CARLOS MARIJUÁN LÓPEZ		
Datos de contacto (E-mail, teléfono...)	DESPACHO: 2D041 TELÉFONO: 983 423731 E-MAIL: marijuan@mat.uva.es		
Horario de tutorías	Véase www.uva.es → Centros → Campus de Valladolid → Escuela Técnica Superior de Ingeniería Informática → Tutorías		
Departamento	MATEMÁTICA APLICADA		

1. Situación / Sentido de la Asignatura

1.1 Contextualización

En esta asignatura nos interesamos por un problema relativo a la clasificación de grandes volúmenes de información interrelacionada. Estudiaremos técnicas matemáticas que soportan el funcionamiento de los algoritmos con que actualmente se abordan estos problemas y las utilizaremos para mostrar el interés de su conocimiento aplicándolas a casos reales.

1.2 Relación con otras materias

Las técnicas espectrales estudiadas en esta asignatura son útiles en materias relacionadas con *Spectral Clustering*, *Data Mining* o *Big Data*.

1.3 Prerrequisitos

Nociones básicas de teoría de grafos, álgebra lineal y cálculo infinitesimal generalmente estudiados en los títulos de grado que permiten el acceso a este Máster.

2. Competencias

2.1 Generales

Código	Descripción
CG4	Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática.
CG8	Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos.

2.2 Específicas

Código	Descripción
CET2	Capacidad de comprender y saber aplicar el funcionamiento y organización de Internet, las tecnologías y protocolos de redes de nueva generación, los modelos de componentes, software intermediario y servicios.
CET9	Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.

3. Objetivos

Código	Descripción
CET2.1	Conocer los fundamentos matemáticos que permiten comprender la organización de grandes redes y servicios de Internet.
CET9.1	Conocer los fundamentos de la Teoría de Particionado Espectral en que se basan la clasificación de grandes volúmenes de información.
CET9.2	Ser capaz de aplicar técnicas de particionado espectral para la clasificación y comprensión de información de grandes redes en Internet.

4. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	14	Estudio y trabajo individual	20
Clases prácticas de aula (A)		Estudio y trabajo en grupo	25
Laboratorios (L)	8		
Prácticas externas, clínicas o de campo			
Seminarios (S)	6		
Tutorías grupales (TG)			
Evaluación (fuera del periodo oficial de exámenes)	2		
Total presencial	30	Total no presencial	45

5. Bloques temáticos

Bloque 1: Clasificación de información interrelacionada en Internet

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

En este bloque pretendemos clasificar grandes volúmenes de información en grupos cuyos elementos están densamente relacionados entre sí dentro de cada grupo, pero escasamente relacionados si pertenecen a grupos distintos. Desarrollaremos la Teoría de Particionado Espectral por ser especialmente eficiente cuando se trata de clasificar volúmenes de información de inmensa magnitud. Analizaremos una implementación real en un sistema informático y trataremos con herramientas apropiadas para su aplicación sobre distintas áreas de trabajo en Internet.

b. Objetivos de aprendizaje

Código	Descripción
CET2.1	Conocer los fundamentos de la Teoría de Grafos que permiten comprender la organización de grandes redes y servicios de Internet.
CET9.1	Conocer los fundamentos de la Teoría de Particionado Espectral en que se basan la clasificación de grandes volúmenes de información.
CET9.2	Ser capaz de aplicar técnicas de particionado espectral para la clasificación y comprensión de información de grandes redes en Internet.

c. Contenidos

TEMA 1: Introducción y objetivos

- 1.1 Introducción al clustering
- 1.2 Objetivos de desarrollo teórico.
- 1.3 Objetivos de desarrollo práctico.

TEMA 2: Fundamentos Matemáticos

- 2.1 Nociones algebraicas básicas



- 2.2 Teoría de Grafos para clustering.
- 2.3 Introducción al particionado de conjuntos,
Algoritmo *k-means*.

TEMA 3: Teoría de Particionado Espectral

- 3.1 Grafos de similaridad
- 3.2 Matrices Laplacianas
- 3.3 Algoritmos de Particionado Espectral
- 3.4 Análisis y justificación de los algoritmos

TEMA 4: Consideraciones prácticas

- 4.1 Computación de autovectores. Método de Arnoldi
- 4.2 Acondicionamiento de matrices
- 4.3 Elección del número de núcleos, de la matriz Laplaciana y de la función de similaridad

TEMA 5: Exploración en la web

- 5.1 *Web crawlers*
- 5.2 La araña WIRE
- 5.3 Aplicación a un problema real

d. Métodos docentes

Ver 8. Anexo: Métodos docentes

e. Plan de trabajo

Ver 9. Anexo: Cronograma de actividades previstas

f. Evaluación

Ver 7. Sistema de calificaciones – Tabla resumen.

g. Bibliografía básica

U. von Luxburg, *A Tutorial on Spectral Clustering*. Technical Report No. TR-149. Max-Planck-Intitut fur biologische Kybernetik. Department for Empirical Inference, 2006.

A. Ng, M. Jordan and Y. Weiss. *On spectral clustering: analysis and an algorithm*. En T. Dietterich, S. Becker and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14*. MIT Press, 2002.

J. Shi and J. Malik. *Normalized cuts and image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, **22** (8), 888-905.

C. Castillo. *Effective Web Crawling* (Ph. D. in Computer Science), 2004.

h. Bibliografía complementaria

F. Chung. *Spectral graph theory*, Washington: Conference board of the Mathematical Sciences, 1997.

D. Arthur and S. Vassilvitskii, *How Slow is the k-means Method?* Proceedings of the 2006 Symposium on Computational Geometry (SoCG), 2006.

D. Arthur and S. Vassilvitskii, *k-means++ The Advantages of Careful Seeding*. Symposium on Discrete Algorithms (SODA), 2007.

Z. Markov and D.T. Larose, *Data Mining the Web: uncovering patterns in web content, structure and usage*. Wiley Series on methods and Applications in Data Mining. Wiley-Interscience. A John Wiley & Sons, Inc. Publication, 2007.

B. Mohar, *Some applications of Laplace eigenvalues of graphs*. En G. Hahn and G. Sabidussi (Eds.), *Graph simmetry: Algebraic Methods and applications* (Vol. NATO ASI Ser. C 497, p. 225-275). Kluwer 1991.

D.S. Watkins. *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*. Siam, 2007.

i. Recursos necesarios

El profesor de la asignatura proporcionará los recursos necesarios para realizar este bloque.

6. Temporalización (por bloques temáticos)

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Bloque 1: Clasificación de información interrelacionada en Internet	3 ECTS	Semanas 1 a 8

7. Sistema de calificaciones – Tabla resumen

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Proyecto 1	1/3	Entrega en la semana 3
Proyecto 2	1/3	Entrega en la primera sesión de la semana 6
Proyecto 3	1/3	Entrega en la semana 7
Seminarios	adicional	En las sesiones indicadas en el Anexo 9
Examen final		Periodo de exámenes

CRITERIOS DE CALIFICACIÓN

- Evaluación continua:**
 Se puede aprobar esta asignatura mediante la realización de los tres proyectos indicados en la tabla anterior. Se requiere obtener al menos el 40% en cada uno de los tres proyectos y que la suma de las tres no sea inferior al 50% de la nota total.
 La participación activa y creativa en los Seminarios será evaluada positivamente y la correspondiente calificación se añadirá a la nota total obtenida mediante los proyectos.
 En la sesión de clase de la semana 8 cada alumno se podrá presentar discrecionalmente a un examen para mejorar la calificación de uno de sus proyectos a su elección.
- Convocatoria ordinaria:**
 Cada alumno se examinará, al menos, de los proyectos considerados en la evaluación continua en los que no haya obtenido el 40% de la calificación. También podrá examinarse, a su elección, de los proyectos en los que haya obtenido al menos el 40% de la calificación. Cuando haya dos calificaciones de una misma proyecto se considerará la superior de las dos.
- Convocatoria extraordinaria:**
 Realización de un examen por valor del 100% de la calificación de la asignatura.

8. Anexo: Métodos docentes

Actividad	Metodología
Clases teórico-prácticas	<ul style="list-style-type: none"> Clase magistral participativa Estudio de casos Resolución de problemas
Laboratorio	<ul style="list-style-type: none"> Clase magistral participativa Resolución de casos prácticos con apoyo informático Realización de proyectos en grupo (2/3 alumnos) guiado por el profesor, siguiendo un enfoque colaborativo.
Seminarios	<ul style="list-style-type: none"> Sesiones de debate entre alumnos y profesor sobre los proyectos presentados.

9. Anexo: Cronograma de actividades previstas

Semana	Sesión	Contenido	Actividad	Estudio individual	Estudio en grupo
1	1 2h.	Tema 1	Clase T/M	2h	
	2 2h.	Tema 2.1	Clase T/M	2h	
2	1 2h.	Tema 2.1	Clase T/M	2h	
	2 2h.	Tema 2.1	Laboratorio	2h	
3	1 2h.	Tema 2.2-- 2.3	Clase T/M	2h	
	2 2h.	Temas 1 y 2	Seminario		2h
4	1 2h	Tema 3.1 --3.3	Clase T/M	2h	
	2 2h	Tema 3.4	Laboratorio	2h	
5	1 2h.	Tema 4.1—4.2	Clase T/M	2h	
	2 2h.	Tema 4.3	Laboratorio	2h	
6	1 2h.	Temas 3 y 4	Seminario		2h
	2 2h.	Tema 5.1—5.2	Clase T/M	2h	
7	1 2h.	Tema 5.3	Laboratorio	2h	
	2 2h.	Tema 5	Seminario		2h
8	1 2h.	Temas 1 a 5	Examen	discrecional	2h