

MASTER		Master en Data Science para Finanzas	
ASIGNATURA		<i>Técnicas de agrupación y reducción de la dimensión</i>	
Nº de ECTS		6	
Nº de horas docentes		Parte II: 22.5h (3 ECTS, 15 sesiones)	
Nº de horas actividades académicas dirigidas			
Profesor responsable de la asignatura		Juan Manuel López Zafra	
Curso académico			
Cuatrimestre		1º Cuatrimestre	

1.- DESCRIPCIÓN GENERAL DE LA ASIGNATURA Y OBJETIVOS DE DOCENCIA:

En Técnicas de análisis estadístico en Data Science (I) parte II se pretende un doble objetivo: mostrar al alumno las características del método de análisis en data science y adentrarle en las técnicas estadísticas imprescindibles para la aplicación de algoritmos de machine learning, tanto supervisados como no supervisados.

Técnicas de análisis estadístico en Data Science (I) parte II.

- Primera parte: Métodos de reducción de la dimensión
 - Análisis de componentes principales
 - Escalas multidimensionales
 - Análisis factorial
 - Análisis factorial de correspondencias
- Segunda parte: Métodos de agrupación
 - Análisis cluster: métodos jerárquicos y jerárquicos
 - Métodos no jerárquicos
 - Partitioning around Medoids
 - Agglomerative Nesting
 - Divisive Analysis

2.- FORMA DE EVALUACIÓN PREVISTA:

Participación	10%
Actividades académicas dirigidas	40%
Prueba objetiva final	50%

Nota: para aprobar la asignatura será imprescindible obtener al menos un 5 en examen final; las actividades académicas dirigidas no serán reevaluables. La asistencia a clase es obligatoria, admitiéndose hasta un 20% de ausencias sin justificación; será criterio del profesor admitir o no la justificación; una asistencia menor del 80% supondrá la pérdida del derecho a examen en convocatoria ordinaria.

PROGRAMA DETALLADO

Nº de sesión	Detalle del contenido docente: temas, casos prácticos, actividades académicas dirigidas que se verán en dicha sesión,...	Lecturas recomendadas o referencias bibliográficas relativas a los conceptos-temas desarrollados en la sesión

1	Técnicas de reducción de la dimensión: el análisis de componentes principales (ACP). Aproximación geométrica y características esenciales. Práctica 1.	Zeltermán. Mayor. Dell. Apuntes. Iris Dataset. Identificación de relaciones comunes.
2	ACP. Práctica 2. Estructura subyacente de tipos de interés. Introducción de variables y observaciones suplementarias.	3 años de datos diarios de emisiones de deuda USA sobre 10 plazos diferentes.
3	Técnicas de reducción de la dimensión: las escalas multidimensionales (MDS). Aproximación geométrica y características esenciales.	Borg et al. Dell.
4	Técnicas de reducción de la dimensión: MDS. Práctica 3	Posicionamiento de ciudades y generación de distancias.
5	Técnicas de reducción de la dimensión: el análisis factorial (ANFAC). Aproximación geométrica y características esenciales.	Zeltermán. Dell. Apuntes
6	Técnicas de reducción de la dimensión: ANFAC. Práctica 4	Vehículos. Exploración de la matriz de correlaciones, medidas de bondad, extracción de factores.
7	Técnicas de reducción de la dimensión: ANFAC. Práctica 5	Banca. Ratios de solvencia, liquidez rentabilidad. Estructura subyacente, rotaciones factoriales ortogonales y oblicuas.
8	Técnicas de reducción de la dimensión: el análisis factorial de correspondencias (ANACOR). Aproximación geométrica y características esenciales. Práctica 6	Greenacre. Dell. Apuntes Selección de características de producto en función de la edad del consumidor. Análisis exploratorio, contrastes de independencia, mapas factoriales.
9	Métodos de agrupación: el análisis cluster (AC). Aproximación geométrica y características esenciales.	Everitt et al. Mayor. Dell. Apuntes
10	El análisis cluster (AC). Práctica 7	Ratios financieros y quiebras bancarias: identificación inicial de candidatos.
11	Métodos de agrupación: el análisis cluster (AC). Práctica 8	SP 500. Datos financieros e identificación de grupos. Análisis exploratorio, bondad del análisis, número óptimo de clusters.
11	AC. Métodos jerárquicos vs no jerárquicos.	SP 500. Datos financieros e identificación de grupos. Análisis exploratorio, bondad del análisis, número óptimo

	Práctica 9	de clusters.
13	AC. Métodos no jerárquicos. Partitioning around Medoids (PAM) Agglomerative Nesting (AGNES) Divisive Analysis (DIANA) Práctica 10	Churn, small dataset. Identificación de grupos de clientes con mayor probabilidad de abandono: características diferenciadoras.
14	AC. Métodos avanzados. Introducción a la segmentación borrosa (fuzzy clustering), Clustering LARge Applications (CLARA) y Práctica 11.	Churn, complete dataset. Segmentación del dataset aplicando CLARA. Aplicación de segmentación borrosa (algoritmos Fanny y c-medias)
15	AC. Métodos avanzados. El problema de las variables en escala no métrica. Práctica 12	Introducción de variables no métricas; distancia de Gower.

INFORMACION ADICIONAL

Bibliografía básica	<ul style="list-style-type: none"> • R for data science. Wickham and Golemund. O'Reilly, 2017. ISBN 978 1 491 91039 9 • Applied Multivariate Statistics with R. Zelterman, Daniel. Springer International Publishing (2015) Hardcover ISBN: 9783319140926 • Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. Provost y Fawcett. Publisher: O'Reilly Media (2013). ISBN-10: 1449361323
Bibliografía Complementaria	<ul style="list-style-type: none"> • Introduction to data science. Jeffrey M. Stanton, Syracuse University. Descargar gratuitamente aquí: http://surface.syr.edu/istpub/165/ • Tips for Building a Data Science Capability. Booz Allen Hamilton Descargar gratuitamente aquí: http://www.boozallen.com/consulting/strategic-innovation/nextgen-analytics-data-science/building-a-data-science-capability • Análisis de datos con R. Sébastien Lê y Jérôme Pagès François Husson. Publisher: ESCUELA COLOMBIANA DE INGENIERIA. 2013. ISBN-10: 9588726050 • Cluster Analysis 5th Edition. Brian S. Everitt, Sabine Landau, Morven Leese, Daniel Stahl. Wiley (2011) • Statistics – Textbook. Dell. Disponible online en https://documents.software.dell.com/statistics/current/textbook • Learning Predictive Analytics with R. Eric Mayor. Packt Publishing Limited (2015) • Mastering Data Analysis with R. Daroczi. Packt Publishing Limited (2015)

	<ul style="list-style-type: none">• Applied Multidimensional Scaling. Borg, Groenen, Mair. Springer (2013)• Correspondence Analysis in Practice, 2nd ed. Greenacre. Chapman and Hall/CRC, 2007
Actividades Complementarias	
Localización del profesor	Despacho E 6.7 jmlopezzafra@cunef.edu