

MASTER		Master en Data Science para Finanzas	
ASIGNATURA		<i>Técnicas de clasificación</i>	
Nº de ECTS		6	
Nº de horas docentes			
Nº de horas actividades académicas dirigidas			
Profesor responsable de la asignatura		José María Sarabia Alegría	
Cuatrimestre		1º Cuatrimestre	

MASTER		Master en Data Science para Finanzas	
PARTE		<ul style="list-style-type: none"> • Técnicas de clasificación I: Regresión logística • Técnicas de clasificación II: Análisis discriminante • Técnicas de clasificación III: Árboles de decisión • Técnicas de clasificación IV: Clasificación mediante el método de Bayes ingenuo • Técnicas de clasificación V: Algoritmo kNN: k vecinos más próximos • Técnicas de clasificación VI: Máquinas de soporte vectorial 	
Nº de ECTS		2	
Nº de horas docentes		15	
Nº de horas actividades académicas dirigidas			
Profesor responsable de la asignatura		José María Sarabia Alegría	
Curso académico		2020 / 2021	
Cuatrimestre		1º Cuatrimestre	

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

El objetivo de esta asignatura es el estudio y aplicación en situaciones reales de algunas de las metodologías de clasificación más usadas en Data Science. Las diferentes metodologías serán presentadas detalladamente, para a continuación continuar con su aplicación práctica mediante el programa R. Las técnicas que se estudiarán son: Regresión logística, Análisis discriminante, Árboles de decisión, clasificación mediante el método de Bayes ingenuo, el algoritmo kNN (k vecinos más próximos) y máquinas de soporte vectorial.

Regresión logística

Introducción a los modelos de elección discreta. Especificación.
 Estimación, contraste y validación en regresión logística.
 Modelo de regresión logística multinomial.
 Casos Prácticos con R.

Análisis discriminante

Introducción al análisis discriminante. Clasificación de dos grupos y una variable predictora.
 Función discriminante de Fisher. Validación y extensiones del modelo.
 Casos Prácticos con R.

Arboles de decisión

Introducción: arboles de clasificación y de regresión.
 Construcción del árbol. Métricas de selección.
 Control de tamaño, poda del árbol y validación. Predicción
 Casos Prácticos con R.

Clasificación mediante el método de Bayes ingenuo

Metodología de clasificación.
 Estimación y validación del método.
 Casos Prácticos con R.

Algoritmo kNN: k vecinos más próximos

Explicación del algoritmo.
 Caso de dos inputs y caso general. Validación
 Elección del k.
 Casos Prácticos en R.

Máquinas de soporte vectorial

Algoritmo de construcción.
 Hiperplanos de separación y alternativas. Validación.
 Casos Prácticos con R.

2.- FORMA DE EVALUACIÓN PREVISTA:

Tareas de clase	20%
Trabajo de aplicación de técnicas de clasificación en Data Science	30%
Prueba objetiva final	50%

Nota: para aprobar la asignatura será imprescindible obtener al menos un 5 en el examen final. 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	Regresión logística Introducción a los modelos de elección discreta. Especificación. Estimación, contraste y validación en regresión logística. Casos Prácticos con R.	Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i> . Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes</i> . Mc Graw Hill
2	Regresión logística	Hastie, T., Tibshirani R. and Friedman, J.

	<p>Modelo de regresión logística multinomial. Casos Prácticos con R.</p>	<p>(2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i>. Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes</i>. Mc Graw Hill</p>
3	<p>Análisis discriminante Introducción al análisis discriminante. Clasificación de dos grupos y una variable predictora. Casos de dos predictores. Casos Prácticos con R.</p>	<p>Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). <i>Multivariate Analysis</i>. Academic Press, London. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i>. Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes</i>. Mc Graw Hill</p>
4	<p>Análisis discriminante Función discriminante de Fisher. Validación y extensiones del modelo. Casos Prácticos con R. Arboles de decisión Introducción: arboles de clasificación y de regresión.</p>	<p>Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). <i>Classification and Regression Trees</i>. Chapman & Hall - CRC. Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). <i>Multivariate Analysis</i>. Academic Press, London. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i>. Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes</i>. Mc Graw Hill</p>
5	<p>Arboles de decisión Construcción del árbol. Métricas de selección. Control de tamaño, poda del árbol y validación. Predicción Casos Prácticos con R.</p>	<p>Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). <i>Classification and Regression Trees</i>. Chapman & Hall - CRC. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i>. Second Edition. Springer.</p>
6	<p>Clasificación mediante el método de Bayes ingenuo Metodología de clasificación. Estimación y validación del método. Casos Prácticos con R.</p>	<p>Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). <i>Multivariate Analysis</i>. Academic Press, London. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i>. Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes</i>. Mc Graw Hill</p>
7	<p>Algoritmo kNN: k vecinos más próximos Explicación del algoritmo. Caso de dos inputs y caso general. Validación Elección del k. Casos Prácticos en R.</p>	<p>Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction</i>. Second Edition. Springer.</p>

8	Máquinas de soporte vectorial Algoritmo de construcción. Hiperplanos de separación y alternativas. Validación.	Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction.</i> Second Edition. Springer.
9	Presentación trabajos de aplicación de técnicas de clasificación en Data Science. Revisión de técnicas y conceptos.	Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). <i>Classification and Regression Trees.</i> Chapman & Hall - CRC. Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). <i>Multivariate Analysis.</i> Academic Press, London. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction.</i> Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes.</i> Mc Graw Hill
10	Presentación trabajos de aplicación de técnicas de clasificación en Data Science. Revisión de técnicas y conceptos.	Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). <i>Classification and Regression Trees.</i> Chapman & Hall - CRC. Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). <i>Multivariate Analysis.</i> Academic Press, London. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction.</i> Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes.</i> Mc Graw Hill

INFORMACION ADICIONAL

Bibliografía básica	Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). <i>Classification and Regression Trees.</i> Chapman & Hall - CRC. Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). <i>Multivariate Analysis.</i> Academic Press, London. Hastie, T., Tibshirani R. and Friedman, J. (2008). <i>The Elements of Statistical Learning. Data Mining, Inference and Prediction.</i> Second Edition. Springer. Peña, D. (2002). <i>Análisis de Datos Multivariantes.</i> Mc Graw Hill
Bibliografía Complementaria	https://cran.r-project.org/web/views/
Actividades Complementarias	
Localización del profesor	Correo electrónico : josemaria.sarabia@cunef.edu