



Propuesta de Trabajo Fin de Grado en Matemáticas (curso 2022–2023)

<i>Responsable de tutorización:</i> José Luis Romero Béjar <i>Departamento:</i> Estadística e Investigación Operativa <i>Correo electrónico:</i> jlrbejar@ugr.es
<i>Responsable de cotutorización:</i> José Antonio Sáez Muñoz <i>Departamento:</i> Estadística e Investigación Operativa <i>Correo electrónico:</i> joseasaezm@ugr.es
<i>(Rellenar sólo en caso de que la propuesta esté realizada a través de un estudiante):</i> <i>Estudiante que propone el trabajo:</i> José Pablo Rodríguez Gómez

<i>Título del trabajo:</i> Modelos de Aprendizaje Supervisado Aplicados al Procesamiento del Lenguaje Natural.
<i>Tipología del trabajo (marcar una de las siguientes casillas):</i> <input checked="" type="checkbox"/> <i>Complemento de profundización</i> <input type="checkbox"/> <i>Divulgación de las Matemáticas</i> <input type="checkbox"/> <i>Docencia e innovación</i> <input checked="" type="checkbox"/> <i>Herramientas informáticas</i> <input checked="" type="checkbox"/> <i>Iniciación a la investigación</i>
<i>Materias del grado relacionadas con el trabajo:</i> Probabilidad, Estadística Multivariante,...
<i>Descripción y resumen de contenidos:</i> El <i>Machine Learning</i> es una disciplina del campo de la Inteligencia Artificial que dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones. Distintos algoritmos, basados en modelos probabilísticos, son la base del <i>Aprendizaje Supervisado</i> . Estos modelos permiten un aprendizaje previo basado en un etiquetado de los datos, que permite tomar decisiones o hacer predicciones en función de un conjunto de variables explicativas. La identificación del carácter cuantitativo o cualitativo de las variables de entrada es el punto de partida para la elección de una metodología de clasificación adecuada. La transformación digital, en la que se encuentra la sociedad actual, tiene en el <i>Machine Learning</i> a uno de sus pilares fundamentales por su potencial aplicabilidad. Entre algunas de las aplicaciones más destacadas se encuentra el procesamiento del lenguaje natural (PLN) a través de la comprensión del lenguaje humano. En este trabajo se pretende que el alumno, a partir de una discusión general sobre la problemática de la clasificación estadística, profundice en el análisis de datos categóricos por medio del mejor modelo según el carácter de las variables de entrada, dando una visión completa y actualizada de sus aspectos fundamentales, su implementación, su validación y su aplicación.

Actividades a desarrollar:

- Revisión bibliográfica sobre metodologías consolidadas de clasificación estadística, con especial referencia a su estado actual, y exposición sintética del conocimiento desde una perspectiva global en el contexto del análisis estadístico de datos multivariantes.
- Profundización en el análisis de datos categóricos por medio del mejor modelo, según el carácter de los datos multivariantes usados como regresores, con una clara identificación de los elementos conceptuales, y exposición de sus fundamentos matemáticos y aspectos metodológicos.
- Aplicación para un problema de PLN mediante el uso y desarrollo eventual de procedimientos computacionales y gráficos relacionados.

Objetivos matemáticos planteados

Identificación de los aspectos conceptuales inherentes a los enfoques de clasificación objeto de estudio y su formalización matemática.

Análisis pormenorizado de los fundamentos matemáticos que sustentan el desarrollo de la metodología objeto central del trabajo.

Desarrollo de un estudio aplicado e interpretación de resultados con referencia precisa a la fundamentación matemática de la metodología.

Bibliografía

- [1] C. C. AGGARWAL, *Data Mining: The Textbook*, (Springer, Switzerland, 2015).
- [2] F. E. HARREL, *Regression Modeling Strategies*, (Springer, New York, 2001).
- [3] J. M. HILBE, *Practical Guide to Logistic Regression*, (CRC Press, Boca Raton, 2015).
- [4] D. W. HOSMER, S. LEMESHOW, R. X. STURDIVANT, *Applied Logistic Regression*, (John Wiley & Sons, New Jersey, 2013).
- [5] B. RATNER, *Statistical and Machine-Learning Data Mining. Techniques for Better Predictive Modelling and Analysis of Big Data*, (CRC Press, Boca Raton, 2011).
- [6] D. ZELTERMAN, *Applied Multivariate Statistics with R*, (Springer, Switzerland, 2015).

Firma del estudiante

(sólo para trabajos propuestos por estudiantes)

Firma del responsable de tutorización

(sólo para trabajos propuestos por estudiantes)

Firma del responsable de cotutorización

(sólo para trabajos propuestos por estudiantes)

En Granada, a 10 de mayo de 2022.