



1. DATOS BÁSICOS DEL TFG:

Título: Análisis de Datos Clínicos Mediante el Uso de Técnicas Multivariantes

Descripción general (resumen y metodología):

La investigación clínica, en general, viene experimentando un creciente interés por el empleo de técnicas multivariantes vinculadas al aprendizaje automático y al análisis de big data. En este contexto, el uso de herramientas como árboles de decisión, análisis discriminante o modelos de regresión logística para, por ejemplo, predecir la supervivencia de pacientes con diferentes tipos de cáncer, entre otras muchas aplicaciones, se ha vuelto cada vez más común. Sin embargo, en ocasiones, la validación de estos modelos de clasificación resulta deficiente debido a una tasa insatisfactoria de clasificaciones correctas. Esto suele ser resultado de una definición estática de los niveles de la variable respuesta o al uso inadecuado de variables explicativas desde el punto de vista del análisis estadístico.

En tales situaciones, un análisis exploratorio previo utilizando técnicas multivariantes como el clustering, el análisis de componentes principales o el análisis factorial puede resultar útil para definir los niveles de la variable respuesta y así optimizar el rendimiento de los modelos de clasificación.

Este trabajo tiene como objetivo que el alumno proporcione una visión exhaustiva y actualizada sobre diversas técnicas multivariantes utilizadas en el ámbito de la medicina, incluyendo sus fundamentos, implementación y aplicación en este campo específico.

Metodología:

- Exploración bibliográfica sobre metodologías del análisis estadístico de datos multivariantes.
- Profundización en alguna/s de las metodologías relacionadas con el análisis a partir de datos multivariantes, identificando de forma clara los elementos conceptuales inherentes y exponiendo sus fundamentos matemáticos.
- Aplicación a un problema de clasificación en el área de la medicina con datos reales mediante el uso y desarrollo de procedimientos computacionales relacionados.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

1. Identificación de los aspectos conceptuales de alguna/s técnicas multivariantes objeto de estudio y su formalización matemática.
2. Análisis de los fundamentos matemáticos que sustentan el desarrollo de las metodologías objeto central del trabajo.
3. Desarrollo de un estudio aplicado en el campo de la medicina interpretando los resultados con referencia precisa a los fundamentos matemáticos de la metodología empleada.

Bibliografía básica:

1. C. C. Aggarwal, Data Mining: The Textbook, (Springer, Switzerland, 2015).
2. G. D. Garson, Logistic Regression: Binary and Multinomial, (Statistical Associates Publishing, Asheboro, 2014).
3. W. K. Hardle, L. Simar, Applied Multivariate Statistical Analysis, (Springer, Berlin, 2015).

4. B. Ratner, Statistical and Machine-Learning Data Mining. Techniques for Better Predictive Modelling and Analysis of Big Data, (CRC Press, Boca Raton, 2011).
5. A. C. Rencher, W. F. Christensen, Methods of Multivariate Analysis, (John Wiley & Sons, New Jersey, 2012).
6. A. Caballero-Vázquez et al., Risk Factors for Short-Term Lung Cancer Survival, Journal of Clinical Medicine, 10, (2021), 519.
7. D. Zelterman, Applied Multivariate Statistics with R, (Springer, New York, 2015).

Recomendaciones y orientaciones para el estudiante:

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: JOSÉ LUIS ROMERO BÉJAR

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: jlrbejar@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos:

Ámbito de conocimiento/Departamento:

Correo electrónico:

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: FRANCISCO JAVIER GARCIA GOMEZ

Correo electrónico: frangarciafp02@correo.ugr.es