



Regresión en alta dimensión

(Fecha última actualización: 4 de marzo de 2022)

Máster en Estadística. CURSO ACADÉMICO 2021-2022	
Título	Regresión en alta dimensión y técnicas de regularización
Tipo	INVESTIGACIÓN <input checked="" type="checkbox"/> ORIENTACIÓN PRÁCTICA <input type="checkbox"/>
Número de alumnos	2
Profesor(es)/ email	María Dolores Martínez Miranda (mmiranda@ugr.es)
Descripción	<p>Los modelos de regresión en dimensiones elevadas o con un número de observaciones muy grande (<i>big data</i>) suponen un reto tanto computacional como teórico.</p> <p>La inclusión de predictores en un modelo de regresión lineal múltiple no es gratis y existe un precio a pagar en términos de variabilidad añadida, difícil interpretación, además de la posible inclusión de predictores altamente correlados.</p> <p>Seleccionar el subconjunto más adecuado de predictores en regresión múltiple permitiría reducir la dimensión y simplificar la interpretación de modelo. Sin embargo la tarea de encontrar tal subconjunto no es sencilla, ni es única, es más, no existe una solución única que funcione en todos los casos.</p> <p>Este trabajo supone una profundización del estudiante del máster al problema de la regresión con un número elevado de predictores, tomando conciencia de los problemas teóricos y computacionales, y estudiando los métodos más relevantes para la selección variables, en general la selección de modelos en el contexto de la regresión lineal múltiple.</p>
Objetivos particulares	<ul style="list-style-type: none"> • Tomar conciencia del problema de la alta dimensionalidad en la regresión lineal múltiple, tanto desde el punto de vista teórico como práctico. • Conocer los métodos clásicos y estudiar los métodos modernos más relevantes para reducir la dimensión, haciendo especial hincapié en los métodos selección de variables. Ser capaz de identificar sus ventajas e inconvenientes. • Reconocer algunas de estas técnicas dentro del contexto y formulación del Machine Learning. • Realizar experimentos de simulación utilizando el entorno de análisis y programación estadística R y los paquetes de R específicos más relevantes para este tema. • Realizar aplicaciones con datos reales.
Prerrequisitos y recomendaciones	Haber cursado alguna asignatura donde se incluyan contenidos de modelos de regresión y su inferencia. Además de destreza en el manejo del entorno de programación y análisis estadístico R, que incluya programación a un nivel medio.
Plan de trabajo	<ol style="list-style-type: none"> 1. Revisión bibliográfica. 2. Desarrollo de la parte teórica del trabajo. 3. Revisión de paquetes de R específicos en el tema y documentación de los más relevantes. 4. Experimentos de simulación. 5. Aplicaciones con datos reales. 6. Conclusiones del trabajo incluyendo sus limitaciones y posibles extensiones.



Competencias generales y específicas	<p>Competencias generales:</p> <p>CG1 - Los titulados han de saber aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.</p> <p>CG7 - Los titulados han de realizar una contribución a través de una investigación original que amplíe las fronteras del conocimiento desarrollando un corpus sustancial, del que parte merezca la publicación referenciada a nivel nacional o internacional.</p> <p>CG8 - Los titulados deben ser críticos en el análisis, evaluación y síntesis de ideas nuevas y complejas.</p> <p>Competencias específicas:</p> <p>CE1 - Conocer métodos para el Análisis de Datos</p> <p>CE5 - Adquirir conocimientos avanzados en Inferencia Estadística</p> <p>CE17 - Adquirir capacidades de elaboración y construcción de modelos y su validación</p> <p>CE18 - Ser capaz de realizar un análisis de datos</p> <p>CE20 - Ser capaz de realizar una correcta representación gráfica de datos</p> <p>CE21 - Conocer, identificar y seleccionar fuentes estadísticas</p> <p>CE22 - Ser capaz de interpretar resultados a partir de modelos estadísticos</p> <p>CE24 - Ser capaz de extraer conclusiones y redactar informes</p> <p>CE25 - Ser capaz de identificar relaciones o asociaciones</p> <p>CE26 - Saber utilizar con destreza entornos de programación y análisis estadístico</p>
Bibliografía	<ol style="list-style-type: none">1. FARAWAY, J. (2014). Linear Models with R. Chapman & Hall/CRC Texts in Statistical Science.2. HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer Series in Statistics. Springer, New York.3. HASTIE, T., TIBSHIRANI, R. and TIBSHIRANI, R. (2020). Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. Statistical Science, Vol. 35, No. 4, 579-592.4. JAMES, G., WITTEN, D., HASTIE, T., TIBSHIRANI, R. (2017). An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics. Springer, New York.5. SEBER, G.A.F. (1984) Multivariate Observations. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons.