**Título:** Dataset of the paper “Variable selection for linear regression in large databases: exact methods” Applied Intelligence, 51(6), 3736-3756.

**Autoría:** Joaquín Pacheco Bonrostro y Silvia Casado Yusta

**Palabas clave:** variable selection, linear regression, Branch & Bound methods, heuristics.

**Resumen:** The variable selection problem in the context of Linear Regression for large databases is analysed. The problem consists in selecting a small subset of independent variables that can perform the prediction task optimally. This problem has a wide range of applications. One important type of application is the design of composite indicators in various areas (sociology and economics, for example). Other important applications of variable selection in linear regression can be found in fields such as chemometrics, genetics, and climate prediction, among many others. For this problem, we propose a Branch & Bound method. This is an exact method and therefore guarantees optimal solutions. We also provide strategies that enable this method to be applied in very large databases (with hundreds of thousands of cases) in a moderate computation time. A series of computational experiments shows that our method performs well compared with well-known methods in the literature and with commercial software.

**Fecha de publicación de datos en página web (**[**https://www.ubu.es/metaheuristicos-grinubumet/ejemplos-y-datos-de-problemas**](https://www.ubu.es/metaheuristicos-grinubumet/ejemplos-y-datos-de-problemas)**):** 2020

**Financiación:** This work was partially supported by FEDER funds and the Spanish Ministry of Economy and Competitiveness (Projects ECO2016-76567-C4-2-R and PID2019-104263RB-C44), the Regional Government of “Castilla y León”, Spain (Project BU329U14 and BU071G19), the Regional Government of “Castilla y León” and FEDER funds (Project BU062U16 and COV2000375).

**Publicación relacionada:** El artículo relacionado está publicado en “Applied Intelligence”

**Licencia de uso de los materiales (licencia Creative Commons):** CC BY-NC

**Ficheros:**

- Matrices for Paper Variables Selection in Linear Regression Size n = 18-42

- Matrices for Paper Variables Selection in Linear Regression Size n = 45-60

- Matrices for Paper Variables Selection in Linear Regression Size n = 63-72