



PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

Métodos de Regresión en Minería de Datos



(506) 2268.8823 - (506) 8708.9091



info@promidat.com



facebook.com/oldemarrodriguez



www.promidat.com

Duración: Cuatro semanas.

Descripción:



En este curso se presentarán los principales métodos en Minería de Datos, especialmente enfocados en métodos de Regresión. El énfasis principal del curso será examinar dichos métodos desde un punto de vista algorítmico y de sus aplicaciones en casos reales. Se le dará especial importancia al uso de los conceptos de minería de datos en aplicaciones reales con bases de datos de gran tamaño, para esto se utilizarán los programas especializados en Minería de Datos, como son la plataforma de desarrollo R y el paquete **regressor**.

¿A quién va dirigido?

Se orienta especialmente a analistas de riesgo, marketing y mercadeo, personal involucrado en proyectos de pronóstico y predicción o en proyectos de "Business Intelligence", administradores de bases de datos, analistas de sistemas, profesionales en estadística o economía y en general a personas que, basados en los datos de la organización, colaboran en la toma de decisiones.

Objetivos:

En este curso el estudiante será capaz de:

1. Comprender la diferencia entre modelos de clasificación y modelos de regresión.
2. Comprender la diferencia y necesidad de usar bases de datos de aprendizaje y bases de datos de testing.
3. Comprender la necesidad de la utilización de modelos, algoritmos, software para predecir el comportamiento futuro.
4. Conocer los principales modelos de regresión, técnicas y algoritmos utilizados para predecir conductas a partir de grandes volúmenes de datos históricos.
5. Utilizar la plataforma *R* y el paquete **regressor** para analizar y desarrollar ejemplos con datos reales.

Metodología:

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas.

- Una vídeo conferencia semanal, las cuales quedarán grabadas en Webex, para que los alumnos la puedan acceder en cualquier momento.
- Trabajos prácticos semanales.
- Foros para plantear dudas al tutor y compañeros.
- Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Minería de Datos que involucren la predicción de una variable numérica utilizando modelos de Regresión.

Contenido:

1. Conceptos básicos Regresión

- a. Conceptos y diferencias entre clasificación y regresión
- b. Diseño de bases de datos de aprendizaje
- c. Diseño de bases de datos de testing
- d. Manipulación de variables cuantitativas y variables cualitativas
- e. ¿Cómo evaluar la calidad de un modelo un modelo de regresión?
- f. Índices de error. Error cuadrático medio, error absoluto medio, correlación y otros

2. Regresión Lineal Clásica

- a. Regresión Lineal simple y múltiple
- b. Interpretación de coeficientes
- c. Aplicaciones en casos reales con el paquete **regressor**

2. Extensiones al Método de Regresión Lineal

- a. Regresión LASSO

- b. Regresión RIDGE
- c. Regresión combinando métodos de reducción de la dimensión ACP y MCP
- d. Aplicaciones en casos reales con el paquete ***regressor***

3. Árboles de Decisión para Regresión – Método CART

- a. Árboles de Regresión
- b. Bosques de Regresión
- c. Boosting para Regresión
- d. Aplicaciones en casos reales con el paquete ***regressor***

3. Máquinas de Soporte Vectorial para Regresión

- a. Máquinas de Soporte Vectorial en problemas regresión
- b. Truco del Kernel.
- c. Aplicaciones en casos reales con el paquete ***regressor***

5. Deep Learning en Problemas de Regresión

- a. Redes Neuronales y su estructura. Neuronas, capas, funciones de activación
- b. Redes Neuronales para problemas predictivos
- c. Redes Neuronales y Deep Learning para problemas de regresión
- d. Aplicaciones en casos reales con el paquete ***regressor***

Bibliografía:

1. Berry M. and Linoff G. “Data Mining Techniques”. John Wiley & Sonsa, 1997.
2. Hastie, Tibshirani and Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, 2009.
3. Giudici Paolo. “Applied Data Mining: Statistical Methods for Business and Industry”. Wiley, 2005.
4. Dunhan M. “Data mining: Introductory and Advanced Topics”. Prentice Hall, 2002.
5. Han J. and Kamber M. “Data Mining: concepts and techniques”, Morgan Kaufman Publishers 2001.

6. Rodríguez R. Oldemar with contributions from Diego Jiménez A. and Andrés Navarro D. (2018). **regressoR**: Exploratory Data Analysis System. R package version 1.0.1. <https://CRAN.R-project.org/package=RegressoR>
7. R Development Core Team. "R: A Programming Environment for Data Analysis and Graphics". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>
8. R Development Core Team. "Writing R Extensions". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>