



PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

Calibración de Modelos en Python



(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 Machine Learning, especialmente enfocados en métodos predictivos, conocidos también como métodos de aprendizaje supervisado. 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 machine learning en aplicaciones reales con bases de datos de gran tamaño, para esto se utilizarán los programas especializados, como son la plataforma de desarrollo Python y Anaconda Spyder.

¿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 aprendizaje supervisado (aprendizaje supervisado) y modelos de aprendizaje no supervisado (aprendizaje no supervisado).
2. Comprender la diferencia entre 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 predictivos, técnicas y algoritmos utilizados para predecir conductas a partir de grandes volúmenes de datos históricos.

5. Utilizar el lenguaje Python en la plataforma Anaconda Spyder para analizar y desarrollar ejemplos con datos reales.
6. Conocer los principales métodos para medir la calidad de un modelo predictivo, como la Curva ROC, Dejando uno afuera (Leave one Out), Validación Cruzada (Cross Validation).
7. Dado una tabla de datos de gran tamaño ser capaz de seleccionar el mejor modelo predictivo y sus mejores parámetros.

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 y programar proyectos de Machine Learning que involucren predicción utilizando modelos predictivos. Luego hace la calibración y la selección del mejor modelos predictivo dado un “data set” específico.

Contenido:

1. Algoritmos en Deep Learning

- a. Algoritmo del Perceptron.
- b. Redes neuronales de múltiples capas.
- c. Algoritmos basados en Tensor Flow.
- d. Aplicaciones en casos reales con Python, paquetes KERAS y Scikit-Learn.
- e. Algoritmos usados por Google en Python para Deep Learning.

2. Otros modelos predictivos

- a. Método de Bayes.

- b. Regresión Logística.
- c. Regresión Logística Ridge y LASSO.
- d. Análisis discriminante lineal y cuadrático.
- e. Aplicaciones en casos reales con Python con Scikit-Learn.

3. Curva ROC para medir la calidad de un modelo predictivo

- a. Algoritmo basado en la definición.
- b. Algoritmo óptimo.
- c. Programación en Python del Algoritmo óptimo.

4. Calibración y selección de modelos descriptivos y predictivos

- a. Remuestreo (bootstrapping).
- b. Dejando uno afuera (Leave One Out).
- c. Validación cruzada (cross-validation).
- d. Calibración de los parámetros de los modelos predictivos.
- e. Selección de los modelos más adecuados para el análisis una determinada base de datos.
- f. Implementación óptima y paralela en Python.

Bibliografía:

1. Berry M. and Linoff G. "Data Mining Techniques". John Wiley & Sonsa, 1997.
2. Bry X. "Analyses factorielles simples", Ed. Economica, Paris, 1995.
3. Hastie, Tibshirani and Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, 2009.
4. Giudici Paolo. "Applied Data Mining: Statistical Methods for Business and Industry". Wiley, 2005.
5. Dunhan M. "Data mining: Introductory and Advanced Topics". Prentice Hall, 2002.
6. Han J. and Kamber M. "Data Mining: concepts and techniques", Morgan Kaufman Publishers 2001.
7. Jambu M. "Introduccion au Data Mining: Analyse Intelligente des données". Eyrolles, Paris, 1999.

8. Mirkin Boris. "Clustering for Data Mining, a data recovery approach". Chapman & Hall. Boca Raton FL, 2005.
9. Rodríguez O, P.J.F. Groenen S. Winsberg and E. Diday. "I-Scal: Symbolic Multidimensional Scaling of Interval Dissimilarities". COMPUTATIONAL STATISTICS & DATA ANALYSIS the Official Journal of the International Association for Statistical Computing, London, 2006.
10. Andreas C. Müller and Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly, 1st Edition, 2017.
11. Dusty Phillips. Python 3 Object-oriented Programming, Second Edition. Packt Publishing Ltd, 2015.
12. Eric Matthes. Python Crash Course A Hands-On, Project-Based introduction to Programming. No Starch Press, Inc. 2016.
13. Jake VanderPlas. Python Data Science. O'Reilly, 2017.
14. John Paul Mueller (Author) and Luca Massaron. Python for Data Science For Dummies (For Dummies (Computer/Tech)) 1st Edition, 2015.
15. Steven F. Lott. Mastering Object-oriented Python. Packt Publishing Ltd, 2014.
16. Python Software Foundation. 2017. [Python 3.6.2 documentation](https://docs.python.org/3.6.2/). python.org.
17. Anaconda 2017. [Download Anaconda Distribution Python 3.6 version](https://www.anaconda.com/distribution/#python36). Anaconda Inc.
18. Anaconda 2017. [Anaconda Documentation](https://docs.anaconda.com/). Anaconda Inc.
19. Richard, C. (2013). Learning R: A Step-by-Step Function Guide to Data Analysis. Sebastopol: O'Reilly.
20. Matloff, N. (2013). The art of R programming: A tour of statistical software design. San Francisco: No Starch Press.