



ML3000

Modelos Analíticos Avanzados en Python

Descripción

Al hacer análisis de datos se invierten muchos recursos en la correcta preparación de los datos y modelado; no obstante, es primordial presentar los resultados de forma tal que sea muy rápido y fácil interpretarlos; como bien se dice, una buena imagen (gráfico) es más valiosa que mil palabras. De esta forma, en este curso se presentarán los fundamentos del lenguaje Python para visualizar la información por medio del uso de herramientas gráficas.

Adicionalmente se usarán paquetes que permiten visualizar de mejor forma los datos según su tipo y dominio. Es importante recordar que la visualización de la información no es únicamente para presentar resultados finales, sino también para conocer los datos sobre los cuales se plantean los modelos.

Objetivos

En este curso el estudiante será capaz de:

1. Generar un modelo predictivo en Python con Scikit-Learn.
2. Aplicar Validación Cruzada (Cross-Validation) y Remuestreo (bootstrapping) para la medición correcta del error de un modelo predictivo.
3. Calibrar adecuadamente tanto métodos exploratorios como métodos predictivos en Machine Learning.
4. Seleccionar el mejor modelo predictivo dado un conjunto de datos.

Metodología

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

1. Una video conferencia semanal, las cuales quedarán grabadas en Zoom, para que los alumnos la puedan acceder en cualquier momento.

2. Trabajos prácticos semanales.
3. Foros para plantear dudas al tutor y compañeros.
4. Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Machine Learning que involucren evaluar, calibrar y seleccionar métodos exploratorios y predictivos en código Python.

Contenido

1. **Modelos supervisados en clasificación.**
 - a. Método de Bayes.
 - b. Análisis Discriminante Lineal.
 - c. Análisis Discriminante Cuadrático.
2. **Modelos supervisados en regresión.**
 - a. Regresión Clásica – Mínimos Cuadrados.
 - b. Regresión Ridge y LASSO.
 - c. Regresión con árboles de decisión, bosques aleatorios y potenciación.
 - d. Regresión usando Máquinas de Soporte Vectorial.
3. **Medir la calidad de un modelo supervisado.**
 - a. Training-testing-validación (the validation test approach).
 - b. Dejando uno afuera (Leave One Out).
 - c. Validación cruzada (cross-validation).
 - d. Curva ROC.
4. **Calibración y selección en modelos no supervisados y en modelos supervisados.**
 - a. Calibración de los parámetros de los supervisados.
 - b. Selección de los modelos más adecuados para el análisis una determinada base de datos.
 - c. Implementación óptima y paralela en Python.
 - d. Probabilidad de Corte como método para enfrentar problemas desbalanceados.

Evaluación

El curso se evalúa con 4 tareas, una por semana, cada tarea tiene un valor de 25 puntos. La nota mínima de aprobación es de 70.

Bibliografía

- [1] Benjamin Johnston and Ishita Mathur. Applied Supervised Learning with Python: Use scikit-learn to build predictive models from real-world datasets and prepare yourself for the future of machine learning. Google-Books-ID: I_eVDwAAQBAJ. Packt Publishing Ltd, Apr. 27, 2019. 404 pp. isbn: 978-1-78995-583-5.
- [2] Yuxi (Hayden) Liu. Python Machine Learning By Example. Google-Books-ID: 0nc5DwAAQBAJ. Packt Publishing Ltd, May 31, 2017. 249 pp. isbn: 978-1-78355-312-9.
- [3] Osvaldo Martin. Bayesian Analysis with Python. Google-Books-ID: t6PcDgAAQBAJ. Packt Publishing Ltd, Nov. 25, 2016. 282 pp. isbn: 978-1-78588-985-1.
- [4] Andreas C. Müller and Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. "O'Reilly Media, Inc.", Sept. 26, 2016. 400 pp. isbn: 978-1-4493-6990-3.
- [5] Tzu-Tsung Wong. "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation". In: Pattern Recognition 48.9 (Sept. 1, 2015), pp. 2839-2846. issn: 0031-3203. doi: 10.1016/j.patcog.2015.03.009. url: <https://www.sciencedirect.com/science/article/pii/S0031320315000989> (visited on 10/19/2022).



Programa Iberoamericano de
Formación en Minería de Datos