



ML3006 Métodos Supervisados en Python

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 en Python.

Objetivos

En este curso el estudiante será capaz de:

1. Comprender la diferencia entre modelos de aprendizaje supervisado y modelos de aprendizaje no supervisado.
2. Comprender la necesidad de la utilización de modelos, algoritmos, software para predecir el comportamiento futuro.
3. Comprender la diferencia entre los datos de "train" y los datos de "testing".
4. Conocer los principales modelos predictivos, técnicas y algoritmos utilizados para predecir conductas a partir de un conjunto de datos históricos.
5. Utilizar el lenguaje Python 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:

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 y programar proyectos de Machine Learning relacionados con aprendizaje Supervisado utilizando modelos predictivos.

Contenido

1. Conceptos de Aprendizaje Supervisado.

- a. Conceptos y diferencias entre aprendizaje supervisado y aprendizaje no supervisado.
- b. Diseño de bases de datos de aprendizaje.
- c. Diseño de bases de datos de testing.
- d. Variables cuantitativas y variables cualitativas.
- e. ¿Cómo evaluar la calidad de un modelo predictivos?
- f. Cálculo de la Matriz de confusión e índices de calidad.

2. Método de los K vecinos más cercanos.

- a. Estructura General del método.
- b. El mejor valor de K.
- c. Algoritmo de Aprendizaje.
- d. Aplicación con datos reales en Python y Scikit-Learn.

3. Máquinas de Soporte Vectorial.

- a. Hiperplano de separación de las clases.
- b. Vectores de soporte.
- c. Función discriminante lineal.
- d. ¿Cómo resolver un Problema Optimización?
- e. MVS no linealmente separables.
- f. Núcleos en Máquinas de Soporte Vectorial.
- g. Aplicación con datos reales en Python y Scikit-Learn.

4. Árboles de Decisión (Método CART).

- a. Algoritmos ID3, C4.5, C5.0 y CART.
- b. Árboles de auto-regresión.
- c. Aplicación con datos reales en Python y Scikit-Learn.

5. Métodos de Consenso y de Potenciación.

- a. Métodos de Consenso (Bagging).
- b. Bosques Aleatorios (Random forests).
- c. Métodos de impulso (Boosting).
- d. Métodos de Potenciación (ADA Boosting).
- e. Métodos de Potenciación (eXtreme Gradient Boosting).
- f. Aplicación con datos reales en Python y Scikit-Learn.

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

Leo Breiman. "Random Forests". In: Machine Learning 45.1 (Oct. 1, 2001), pp. 5–32. issn: 1573-0565. doi: 10.1023/A:1010933404324. url: <https://doi.org/10.1023/A:1010933404324> (visited on 10/21/2022).

[2] JIANGANG HAO and TIN KAM HO. "Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language". In: Journal of Educational and Behavioral Statistics 44.3 (June 1, 2019). Publisher: American Educational Research Association, pp. 348–361. issn: 1076-9986. doi: 10.3102/1076998619832248. url: <https://doi.org/10.3102/1076998619832248> (visited on 10/21/2022).

[3] 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.

[4] William S. Noble. "What is a support vector machine?" In: Nature Biotechnology 24.12 (Dec. 2006), Number: 12 Publisher: Nature Publishing Group, pp. 1565–1567. issn: 1546- 1696. doi: 10.1038/nbt1206-1565. url: <https://www.nature.com/articles/nbt1206-1565> (visited on 10/21/2022).

[5] Shan Suthaharan. "Decision Tree Learning". In: Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. Ed. by Shan Suthaharan. Integrated Series in Information Systems. Boston, MA: Springer US, 2016, pp. 237–269. isbn: 978-1-4899-7641-3. doi: 10.1007/978-1-4899-7641-3_10. url: https://doi.org/10.1007/978-1-4899-76413_10 (visited on 10/21/2022).

[6] Vaibhav Verdhnan. Supervised Learning with Python: Concepts and Practical Implementation Using Python. Berkeley, CA: Apress, 2020. isbn: 978-1-4842-6155-2 978-1-4842- 6156-9. doi: 10.1007/978-1-4842-6156-9. url: <http://link.springer.com/10.1007/978-1-4842-6156-9> (visited on 10/21/2022).



Programa Iberoamericano de
Formación en Minería de Datos