



ML2000

Métodos No Supervisados en Python

Descripción

En este curso se definirán los principales conceptos en Machine Learning, con énfasis principalmente en los métodos No Supervisados. Para ello, se verán técnicas relacionadas con exploración de los datos, reducción de la dimensión y clustering. Para esto se hará uso del lenguaje Python y paquetes especializados en Machine Learning.

Objetivos

En este curso el estudiante será capaz de:

1. Entender la necesidad de la utilización de modelos, algoritmos, software especial para un correcto análisis de los datos.
2. Entender las diferencias entre: estadística, análisis de datos, recuperación de la información, ML – “Machine Learning”, minería de datos y Ciencia de Datos.
3. Conocer los principales modelos, técnicas y algoritmos utilizados en los métodos No Supervisados.
4. Efectuar Análisis Exploratorio de Datos en Python.
5. Utilizar Python en métodos no supervisados de análisis de datos, como lo son Análisis en Componentes Principales, Clusterización Jerárquica, K-Medias, entre otros.

Metodología

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

1. Una vídeo 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 No Supervisado como por ejemplo segmentación de carteras de clientes.

Contenido

1. **Conceptos de Machine Learning.**
 - a. Definiciones básicas en Machine Learning y Ciencia de Datos.
2. **Análisis Exploratorio de Datos.**
 - a. Tipos de variables.
 - b. Estadísticas básicas y matriz de correlaciones.
 - c. Tablas de datos y datos atípicos.
 - d. Uso de NumPy.
 - e. Manipulación de Datos con Pandas.
 - f. Visualización de Datos con Matplotlib.
3. **Métodos de condensación de la información.**
 - a. Análisis en Componentes Principales – ACP (PCA, Karhunen-Loeve o K-L Method).
 - b. Método t-SNE.
 - c. Método UMAP.
 - d. Plano principal.
 - e. Círculo de correlaciones.
 - f. Dualidad y sobre-posición de gráficos.
 - g. Análisis Factorial de Correspondencias Múltiples.



Programa Iberoamericano de
Formación en Minería de Datos

4. Clustering Jerárquico Aglomerativo.

- a. ¿Qué es "cluster analysis"?
- b. Clusterización Jerárquica Aglomerativa.
- c. Distancias y matrices de distancias.
- d. Agregaciones.
- e. Jerarquías binarias.
- f. Jerarquías Binarias sobre las Componentes Principales.

5. Método de k-medias (k-means).

- a. Inercia total, inercia inter-clases e inercia intra-clases.
- b. Teorema de Fisher.
- c. Problema combinatorio.
- d. Método de Forgy.
- e. Método de las nubes dinámicas.

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] Giuseppe Bonaccorso. Hands-On Unsupervised Learning with Python: Implement machine learning and deep learning models using Scikit-Learn, TensorFlow, and more. Google-Books-ID: IgyLDwAAQBAJ. Packt Publishing Ltd, Feb. 28, 2019. 375 pp. isbn: 978-1-78934-927-6.
- [2] M. Emre Celebi and Kemal Aydin, eds. Unsupervised Learning Algorithms. Cham: Springer International Publishing, 2016. isbn: 978-3-319-24209-5 978-3-319-24211-8. doi: 10.1007/978-3-319-24211-8. url: <http://link.springer.com/10.1007/978-3-319-24211-8> (visited on 10/21/2022).
- [3] Benjamin Johnston, Aaron Jones, and Christopher Kruger. Applied Unsupervised Learning with Python: Discover hidden patterns and relationships in unstructured data with Python. Google-Books-ID: I66aDwAAQBAJ. Packt Publishing Ltd, May 28, 2019. 483 pp. isbn: 978-1-78995-837-9.
- [4] Ankur A. Patel. Hands-On Unsupervised Learning Using Python: How to Build Applied Machine Learning Solutions from Unlabeled Data. "O'Reilly Media, Inc.", Feb. 21, 2019. 362 pp. isbn: 978-1-4920-3561-9.
- [5] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: MACHINE LEARNING IN PYTHON (), p. 6.
- [6] Zhi-Hua Zhou. Machine Learning. Springer Nature, Aug. 20, 2021. 460 pp. isbn: 9789811519673.



Programa Iberoamericano de
Formación en Minería de Datos