



ME-4000 Métodos NO Supervisados Avanzados en R y Python

Duración

Cuatro semanas

Descripción del Curso

Este curso avanzado cubre técnicas de métodos no supervisados en el análisis de datos, aplicadas a través de R y Python. Los estudiantes explorarán temas como clustering avanzado, reducción de dimensionalidad y análisis factorial de correspondencias, además de técnicas especializadas como el Análisis Factorial de Correspondencias Simples y Múltiples y el Método STATIS para tablas múltiples. Este conocimiento permitirá a los estudiantes aplicar métodos no supervisados en proyectos complejos de ciencia de datos y resolver problemas de agrupamiento y segmentación.

Objetivos

1. Profundizar en técnicas avanzadas de métodos no supervisados en análisis de datos.
2. Aprender a implementar y evaluar técnicas de clustering avanzado y reducción de dimensionalidad.
3. Aplicar el Análisis Factorial de Correspondencias Simples y Múltiples en R y Python.
4. Explorar y aplicar el Método STATIS para el análisis de tablas múltiples.
5. Desarrollar habilidades para aplicar estos métodos en ciencia de datos, análisis de mercado y estudios multidimensionales.

Metodología

El curso combinará teoría y práctica aplicada en R y Python, permitiendo a los estudiantes implementar y analizar métodos no supervisados en problemas reales de ciencia de datos. Las herramientas y recursos incluyen:

1. **Videoconferencias semanales:**
Sesiones en Zoom que serán grabadas para su revisión posterior.
2. **Trabajos prácticos semanales:**
Ejercicios aplicados en R y Python para fortalecer la implementación de métodos no supervisados avanzados.
3. **Foros de discusión:**
Espacio en Moodle para la resolución de dudas y discusión de temas del curso.
4. **Aula virtual en Moodle:**
Acceso a materiales, ejemplos de código y recursos adicionales.

Luego de este curso el estudiante será capaz de:

1. Aplicar técnicas avanzadas de clustering y reducción de dimensionalidad en R y Python.
2. Realizar Análisis Factorial de Correspondencias Simples y Múltiples.
3. Implementar el Método STATIS para el análisis de tablas múltiples.
4. Utilizar métodos no supervisados avanzados para resolver problemas de segmentación, agrupación y análisis exploratorio de datos.
5. Evaluar y comparar métodos no supervisados en el contexto de aplicaciones en ciencia de datos.

Contenido

- **Semana 1:**
Análisis Factorial de Correspondencias Simples y Múltiples
 - a. Introducción al Análisis Factorial de Correspondencias (AFC) para tablas simples.
 - b. Extensión al Análisis Factorial de Correspondencias Múltiples (AFCM) para tablas múltiples.
 - c. Interpretación de los resultados del AFC y AFCM en contextos de ciencia de datos y segmentación de mercado.
 - d. Ejemplo práctico en Python: Implementación de AFC y AFCM.
 - e. Ejemplo práctico en R: Análisis factorial de correspondencias en datos categóricos.

- **Semana 2:**
Método STATIS y STATIS Dual para el Análisis de Tablas Múltiples
 - a. Introducción al Método STATIS y su aplicación en ciencia de datos.
 - b. Análisis de estructuras comunes entre tablas múltiples y su interpretación.
 - c. Aplicaciones de STATIS en estudios multidimensionales y análisis de tendencias.
 - d. Ejemplo práctico en Python y R: Implementación del Método STATIS para tablas múltiples.
- **Semana 3:**
Técnicas Avanzadas de Reducción de Dimensionalidad y Visualización
 - a. Métodos t-SNE, y UMAP.
 - b. Selección de técnicas de reducción de dimensionalidad para visualización y agrupación.
 - c. Ejemplo práctico en Python: Implementación de t-SNE y UMAP para visualización de datos.
 - d. Ejemplo práctico en R: PCA aplicado a datos multidimensionales.
- **Semana 4:**
Modelos no euclidianos de reducción de la dimensión y técnicas de agrupación basadas en densidades
 - a. Análisis de Datos Topológicos.
 - b. Estadística Riemanniana.
 - c. Introducción a métodos de clustering avanzados: DBSCAN, clustering jerárquico, y clustering basado en densidad.
 - d. Evaluación de modelos de clustering y métricas de validación.
 - e. Ejemplo práctico en Python: Implementación de la Estadística Riemanniana.
 - f. Ejemplo práctico en R: Agrupación de datos utilizando clustering basado en densidad.

Evaluación

La evaluación consistirá en 4 tareas prácticas con igual valor, enfocadas en aplicar métodos no supervisados avanzados en análisis de datos.

Bibliografía

- Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. *Journal of Classification*.
- Greenacre, M. (2007). *Correspondence Analysis in Practice*. CRC Press.
- Abdi, H., & Williams, L. J. (2010). *Principal Component Analysis*. Wiley Interdisciplinary Reviews: Computational Statistics.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley.
- Pagès, J. (2015). *Multiple Factor Analysis by Example Using R*. CRC Press.
- Husson, F., Josse, J., & Pagès, J. (2010). Principal Component Methods - Hierarchical Clustering - Partitional Clustering: Why Would We Need to Choose for Visualizing Data?. *Journal of Statistical Software*.
- Gower, J. C., & Dijksterhuis, G. B. (2004). *Procrustes Problems*. Oxford University Press.
- Lé, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*.
- VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media.