



ME-3000

Álgebra Lineal Aplicada en Ciencia de Datos e Inteligencia Artificial

Duración

Cuatro semanas

Descripción del Curso

El álgebra lineal es una herramienta fundamental en la ciencia de datos, la inteligencia artificial y el deep learning. Este curso introduce los conceptos esenciales de álgebra lineal necesarios para comprender y aplicar técnicas avanzadas de análisis de datos, reducción de dimensionalidad y modelos de aprendizaje automático. A través de ejercicios prácticos, los estudiantes aprenderán a utilizar vectores, matrices, y transformaciones lineales en el desarrollo de modelos predictivos y algoritmos de aprendizaje profundo. Los temas incluyen operaciones con vectores y matrices, descomposiciones en valores singulares, y sus aplicaciones en el Análisis en Componentes Principales, Máquinas de Soporte Vectorial y Redes Neuronales.

Objetivos

1. Familiarizar a los estudiantes con los conceptos básicos y avanzados de álgebra lineal aplicados a la ciencia de datos y el aprendizaje automático.
2. Dotar a los estudiantes de las habilidades para realizar operaciones con vectores y matrices, y aplicarlas en la construcción de modelos de aprendizaje supervisado y no supervisado.
3. Explicar la importancia de los valores y vectores propios, y su uso en técnicas de reducción de dimensionalidad y descomposición.
4. Comprender cómo el álgebra lineal sustenta el funcionamiento de algoritmos avanzados en deep learning, como redes neuronales convolucionales y redes de memoria a largo plazo (LSTM).
5. Desarrollar habilidades para aplicar técnicas de álgebra lineal en el análisis de datos reales y en el diseño de modelos de inteligencia artificial.



Programa Iberoamericano de
Formación en Minería de Datos

Metodología

La metodología del curso combina teoría y práctica, asegurando que los estudiantes puedan aplicar directamente los conceptos en problemas reales. Las herramientas y recursos incluyen:

- 1. Videoconferencias semanales:**
Una sesión en Zoom que quedará grabada para que los estudiantes puedan revisarla en cualquier momento.
- 2. Trabajos prácticos semanales:**
Ejercicios y casos de estudio enfocados en la implementación de algoritmos en Python.
- 3. Foros de discusión:**
Espacio en Moodle donde los estudiantes pueden plantear preguntas, discutir temas del curso y colaborar con sus compañeros.
- 4. Aula virtual en Moodle:**
Acceso a materiales de estudio, presentaciones, ejercicios, y recursos adicionales.

Luego de este curso el estudiante será capaz de:

1. Realizar operaciones complejas con vectores y matrices, incluyendo multiplicación, transposición, y cálculo de determinantes e inversas.
2. Aplicar valores y vectores propios en descomposiciones de matrices y reducción de dimensionalidad, como el Análisis en Componentes Principales (PCA).
3. Modelar rectas y planos en el espacio tridimensional y utilizar estas representaciones en visualización de datos y clustering.
4. Desarrollar e implementar algoritmos de Máquinas de Soporte Vectorial (SVM) y técnicas no supervisadas utilizando conceptos de álgebra lineal.
5. Comprender y construir redes neuronales simples aplicando operaciones matriciales, y entender el funcionamiento de capas convolucionales y de retropropagación en redes profundas.

Contenido

- **Semana 1:**
Fundamentos de Álgebra Lineal Aplicada a Ciencia de Datos
 - a. Introducción a vectores y matrices en ciencia de datos.
 - b. Operaciones con vectores (suma, producto escalar, producto cruz) y matrices (suma, multiplicación, determinantes, inversas).
 - c. Aplicaciones en inteligencia artificial: álgebra lineal en transformaciones de datos y normalización.
 - d. Caso práctico: Preprocesamiento de datos utilizando álgebra lineal.

- **Semana 2:**
Determinantes y sus aplicaciones en Análisis de Datos
 - a. Definición e implementación en R y en Python.
 - b. Aplicaciones en: Análisis de colinealidad en regresión, cálculo de verosimilitud en modelos multivariados, detección de anomalías, cálculo de β , independencia en activaciones neuronales, reducción de dimensionalidad, optimización con restricciones, problemas de optimización convexa o bayesiana
 - c. Caso práctico: Implementación de la Regresión Clásica.
- **Semana 3:**
Valores y Vectores Propios y sus aplicaciones en Análisis de Datos
 - a. Introducción a valores y vectores propios y su importancia en ciencia de datos.
 - b. Descomposición en valores propios (Eigen-decomposition) y aplicaciones en reducción de dimensionalidad.
 - c. Introducción a descomposiciones de matrices (e.g., SVD) y su uso en deep learning.
 - d. Caso práctico: Implementación del Análisis en Componentes Principales (PCA) para reducción de dimensionalidad en un conjunto de datos.
- **Semana 4:**
Implementación del Análisis en Componentes Principales
 - a. Modelado de rectas y planos en el espacio.
 - b. Interpretación geométrica de datos en espacios multidimensionales y su aplicación en clustering.
 - c. Algoritmo e implementación del ACP en R y en Python.
 - d. Aplicación práctica en visualización y agrupación de datos.
 - e. Caso práctico: Programación del Algoritmo del Análisis en Componentes Principales.

Evaluación

La evaluación consistirá en 4 tareas con un valor igual. Cada tarea estará enfocada en aplicar los conceptos de álgebra lineal a problemas reales de ciencia de datos e inteligencia artificial.

Bibliografía

- | | |
|---|---|
| Strang, G. (2016). Introduction to Linear Algebra. Wellesley-Cambridge Press. | Nielsen, M. (2015). Neural Networks and Deep Learning. Online book. |
| Shlens, J. (2014). A Tutorial on Principal Component Analysis. arXiv. | Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media. |
| James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer. | Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press. |
| Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. | |
| Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. | |
| Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459. | |
| Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. | |