



ME-4002 Métodos Supervisados Avanzados en R y Python

Duración

Cuatro semanas

Descripción del Curso

Este curso avanzado se centra en el estudio y la implementación de métodos supervisados avanzados en el análisis de datos, utilizando R y Python. Los estudiantes explorarán algoritmos de clasificación y regresión de última generación, optimización de hiperparámetros, técnicas de ensamblado y métodos de validación para mejorar la precisión y eficiencia de los modelos. Al final del curso, los estudiantes serán capaces de implementar y ajustar modelos supervisados avanzados para resolver problemas complejos en ciencia de datos y machine learning.

Objetivos

1. Comprender y aplicar la técnicas de hiperparametrización en algoritmos avanzados de clasificación y regresión.
2. Aprender técnicas de hiperparametrización en ensamblado, como bagging, boosting y stacking, para mejorar la precisión de los modelos.
3. Optimizar los hiperparámetros de modelos supervisados en R y Python.
4. Aplicar hiperparametrización en métodos de validación cruzada y evaluación de rendimiento en modelos de machine learning.
5. Desarrollar habilidades para seleccionar y ajustar modelos supervisados en función de diferentes aplicaciones en ciencia de datos.

Metodología

El curso combina teoría y práctica aplicada en R y Python, permitiendo a los estudiantes implementar y evaluar modelos 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 supervisados avanzados.
3. **Foros de discusión:**
Espacio en Moodle para resolver dudas y discutir temas del curso.
4. **Aula virtual en Moodle:**
Acceso a materiales, ejemplos de código y otros recursos.

Luego de este curso el estudiante será capaz de:

1. Implementar y evaluar algoritmos avanzados de clasificación y regresión en R y Python.
2. Optimizar modelos supervisados mediante técnicas de ajuste de hiperparámetros.
3. Aplicar técnicas de ensamblado, como bagging y boosting, para mejorar el rendimiento de los modelos.
4. Utilizar métodos de validación y métricas de evaluación para medir la eficacia de los modelos.
5. Seleccionar e implementar el modelo supervisado más adecuado según el contexto y los datos disponibles.

Contenido

- **Semana 1:**
Hiperparametrización de Modelos Analíticos en R
 - a. HPT y Optimización de Hiperparámetros (HPO).
 - b. Medidas de desempeño o calidad.
 - c. Grid Search vs. Random Search.
 - d. Optimización basada en modelos sustitutos (SMBO).
 - e. Ejemplo práctico en R: Implementación de Métodos de hiperpametrización en R.

- **Semana 2:**
Hiperparametrización en R usando Validación Cruzada Modelos Supervisados para medir la calidad
 - a. Hiperparametrización en técnicas de validación cruzada y métodos de evaluación.
 - b. Ejemplo práctico en R: Validación cruzada y evaluación de un modelo de clasificación.
 - c. Ejemplo práctico en R: Hiperparametrización y evaluación de rendimiento en modelos de regresión.
- **Semana 3:**
Hiperparametrización mediante Validación cruzada estratificada en Modelos Supervisados en Python
 - a. Métodos de optimización de hiperparámetros: Grid Search, Random Search, optimización bayesiana, entre otros.
 - b. Técnicas de regularización para mejorar la generalización de los modelos.
 - c. Ejemplo práctico en Python: Optimización de hiperparámetros a pie y usando Scikit-learn.
- **Semana 4:**
Técnicas avanzadas de hiperparametrización en Python.
 - a. Hiperparametrización mediante búsqueda gruesa a fina (Coarse-to-Fine Search, CFS).
 - b. Hiperparametrización mediante búsqueda por reducción sucesiva (Successive Halving, SH).
 - c. Hiperparametrización mediante Optimización Bayesiana (BO).
 - d. Relación entre la Optimización Bayesiana y el Teorema de Bayes.
 - e. Ejemplo práctico en Python: Implementación en Python de todos los métodos anteriores.

Evaluación

La evaluación consistirá en 4 tareas prácticas con igual valor, enfocadas en aplicar y optimizar métodos supervisados avanzados en proyectos de ciencia de datos.

Bibliografía

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

VanderPlas, J. (2016). Python Data Science Handbook. O'Reilly Media.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer.

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). Additive Logistic Regression: A Statistical View of Boosting. The Annals of Statistics.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. arXiv preprint arXiv:1603.02754.

Chollet, F. (2018). Deep Learning with Python. Manning Publications.

Brownlee, J. (2016). Master Machine Learning Algorithms. Machine Learning Mastery.



Programa Iberoamericano de
Formación en Minería de Datos