

FUNDAMENTOS DE ANÁLISIS DE DATOS

Nº de Créditos: **6 ECTS**
Segundo Semestre
Primer Curso

EQUIPO DOCENTE

Martínez Muñoz, Gonzalo
Escuela Politécnica Superior
UAM
Coordinador

Arroyo Guardado, David
Escuela Politécnica Superior
UAM

Dorronsoro Íbero, José
Escuela Politécnica Superior
UAM

González Marcos, Ana María
Escuela Politécnica Superior
UAM

Lago Fernández, Luis F.
Escuela Politécnica Superior
UAM

Sánchez Montañés, Manuel
Escuela Politécnica Superior
UAM

OBJETIVOS

- Aplicar técnicas de aprendizaje automático supervisado y no supervisado para el análisis de datos.
- Aplicar técnicas de preprocesado, auditoría y limpieza a conjuntos de datos.
- Identificar las fortalezas y debilidades de las técnicas de aprendizaje automático en el contexto de grandes volúmenes de datos.
- Manejar herramientas y bibliotecas de Aprendizaje Automático en R y Python.

PROGRAMA DETALLADO

1. Introducción al aprendizaje automático
 - 1.1. Tipos de aprendizaje automático, conceptos básicos, tipos de atributos
 - 1.2. Flujo de un proyecto de aprendizaje automático
 - 1.3. Validación de modelos: tasas de error, matriz de confusión, curvas ROC y validación cruzada
 - 1.4. Regresión lineal, regresión logística
 - 1.5. Vecinos próximos en clasificación y regresión
 - 1.6. Sesgo y varianza. Maldición de la dimensionalidad
2. Preprocesado de datos
 - 2.1. Construcción de la base de datos; tratamiento de múltiples fuentes
 - 2.2. Preparación y auditoría de la base de datos
 - Manejo de variables no numéricas: fechas, strings, variables nominales, etc.
 - Manejo de series temporales
 - Variables sintéticas
 - Auditoría de datos; analítica descriptiva.
 - Manejo de missing values y outliers
 - 2.3. Distribución de las variables.
 - Segmentación y agregación de datos
 - Normalización y estandarización de variables
 - Cambio de la distribución de una variable
 - 2.4. Reducción de la dimensionalidad
 - Selección de variables:
 - Métodos de filtrado (filter methods): Información mutua; multiple testing
 - Métodos envolventes (wrapper methods).
 - Detección de falsos predictores
 - Extracción de variables:
 - Análisis de componentes principales
 - Análisis Discriminante de Fisher
 - 2.5. Información no estructurada; casos prácticos.
 - Texto
 - Imágenes
 - Otros
3. Aprendizaje automático
 - 3.1. Clustering
 - K-means
 - Clúster jerárquico
 - Mezclas de Gaussianas
 - 3.2. Redes neuronales
 - Perceptrones de Rosenblatt
 - Perceptrones multicapa

- Estrategias para aprendizaje de perceptrones
- Deep networks
- 3.3. Clasificación y regresión con máquinas de vectores soporte
 - Clasificación de margen máximo
 - Regresión de margen máximo
 - SVM sobre núcleos lineales
- 3.4. Conjuntos de clasificadores y árboles de decisión
 - CART
 - Random Forests
 - Gradient Boosting Regression

BIBLIOGRAFÍA

1. G. James, D. Witten, T. Hastie, R. Tibshirani. An Introduction to Statistical Learning. Springer Texts in Statistics. 2013.
2. T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning. Springer Series in Statistics.
3. C. Rossant. Learning IPython for Interactive Computing and Data Visualization. Packt Publishing. 2013.
4. J. Grus. Data Science from Scratch, O' Reilly, 2015.
5. Pattern Classification (second edition). R. O. Duda, P. E. Hart & D. G. Stork. Wiley-Interscience, 2000.
6. Pattern Recognition (fourth edition). S. Theodoridis and K. Koutroumbas. Academic Press, 2009.

MÉTODOS DOCENTES

- Lección magistral
- Resolución de problemas
- Prácticas de laboratorio
- Estudio de casos

MÉTODOS DE EVALUACIÓN

- Asistencia a clase: **10%**
- Evaluación continua: **40%**
- Examen final: **50%**