

Profesor: Camilo Franco c.franco31@uniandes.edu.co

Horario de clase: Lunes de 15:30 a 18:20, Salón: C_209

DESCRIPCIÓN DEL CURSO

Los grandes volúmenes de datos que se generan hoy en día en nuestra sociedad de la información presentan grandes retos que requieren tanto de métodos clásicos de análisis estadístico como de técnicas avanzadas de aprendizaje computacional e ingeniería de la información. Las técnicas estadísticas ofrecen una gran versatilidad en el manejo de los datos, facilitando su compresión y explotación, extrayendo conocimiento útil para el apoyo a la toma de decisiones. Este curso propone un marco coherente a partir de los métodos clásicos, ilustrando nuevas técnicas de aprendizaje computacional (supervisado y no-supervisado), con el fin de extraer conocimiento relevante a partir de distintas fuentes de información y generar resultados con alto valor para la recomendación y toma de decisiones.

Objetivo general

Capacitar al estudiante con conceptos y técnicas estadísticas y computacionales que permitan extraer conocimiento relevante para la toma de decisiones.

Resultados esperados de aprendizaje (*learning outcomes*)

- Conocimientos fundamentales dentro del área de la Ciencia de Datos, que incluye inferencia estadística, procesamiento de datos, aprendizaje automático, extracción de conocimiento y lenguajes de programación típicos para el área: Python y R.
- Conocimiento aplicado sobre algoritmos para el análisis estadístico, aprendizaje automático y la extracción de conocimiento para el apoyo a la toma de decisiones.
- Apropiación de elementos técnicos del aprendizaje computacional, y aplicación de funciones a nivel de código abierto.
- Capacidad de analizar y relacionarse críticamente con diferentes fuentes de información, conjuntos de datos y procesos de datos, y aplicarlos para estructurar y formular razonamientos basados en datos.
- Familiarización con las teorías, métodos e interpretaciones existentes dentro del área, y trabajo independiente en la aplicación y evaluación de diferentes tecnologías de procesamiento de datos.
- Capacidad de contribuir con nuevas ideas y procesos de innovación mediante la introducción de enfoques basados en datos.

Metodología: Sesiones magistrales apoyadas en talleres computacionales para el aprendizaje activo y colaborativo. Las actividades de clase se ven complementadas por prácticas y un proyecto final donde se resuelven problemas aplicados. Se hace uso intensivo de aplicaciones utilizando software abierto: R y Python.

CONTENIDO

Parte I. Estructura de observaciones multivariadas, inferencia estadística y causalidad

- Análisis de componentes principales
- Análisis de factores
- Ecuaciones estructurales

Parte II. Aprendizaje no supervisado

- K-vecinos más cercanos
- Clustering jerárquico (métodos aglomerativos)
- Clustering no-jerárquico (métodos iterativos)

Parte III. Aprendizaje supervisado con redes neuronales

- Redes neuronales simples y profundas
- Optimización y modelización
- Redes convolucionales y aprendizaje por transferencia

Parte IV. Sistemas de recomendación

- Sistemas basados en contenidos
- Sistemas de filtrado colaborativo
- Métodos basados en conocimiento y sistemas híbridos

BIBLIOGRAFÍA

- Ch. Aggarwal. Recommender Systems. Springer, 2016.
- I. Goodfellow, Y. Bengio, A. Courville. Deep learning. MIT press, 2016.
- S.S. Haykin. Neural networks: a comprehensive foundation. Prentice Hall, 1999.
- J. Leskovec, A. Rajaraman, J. Ullman. Mining of Massive Datasets. Cambridge University Press
- D. Morrison. Multivariate Statistical Methods. McGraw-Hill, 1990.

EVALUACIÓN

Talleres 30%

Prácticas 30%

Examen 20%

Proyecto 20%