



ANÁLISIS BAYESIANO DE DATOS

NOMBRE

Análisis Bayesiano de datos

NIVEL

Posgrado.

OBJETIVOS

La estadística Bayesiana es una técnica flexible y poderosa que ha visto incrementada su adopción en aplicaciones de Ciencia de Datos durante los últimos años. Los principales beneficios de los métodos Bayesianos respecto de métodos clásicos es que estos son más flexibles para ajustar modelos más complejos y realistas. Además sus resultados, al estar expresados en términos de probabilidades, son más simples de interpretar para los no expertos. Sin embargo, estos métodos son computacionalmente más demandantes que los métodos estadísticos clásicos y existen menos opciones de software amigable que la utilicen. Estas limitaciones técnicas junto con la falta generalizada de cursos en Universidades sobre métodos Bayesianos ha dificultado la adopción de métodos de estadística Bayesiana y su uso en Ciencia da Datos.

Durante el curso se introducirán conceptos estadística Bayesiana desde una aproximación conceptual y computacional reduciendo el uso de lenguaje matemático al mínimo necesario. El enfoque del curso será fundamentalmente práctico usando software amigable y requiriendo la participación activa de los estudiantes.

Objetivos:

- Introducir conceptos en estadística Bayesiana
- Enseñar a realizar modelos Bayesianos a fin de responder preguntas científicas
- Introducir las librerías PyMC3 y ArviZ

PROGRAMA ANALITICO

Tema 1: Teorema de Bayes e introducción a la inferencia Bayesiana

Deducción del teorema de Bayes desde nociones básicas de probabilidad. “Partes” del teorema de Bayes. Ejemplos de aplicación simples: el problema del diagnóstico. El problema de la moneda. A prioris informativos, no informativos y débilmente informativos. Críticas y defensa al uso de los a prioris. A prioris conjugados y tratabilidad analítica de la distribución a posteriori. Cómo interpretar el a posteriori e informar el resultado de un análisis Bayesiano. Estimaciones puntuales e intervalos de credibilidad.



Tema 2: Programación probabilística

Programación probabilística. Presentación de las librerías PyMC3 y ArviZ para inferencia Bayesiana. Revisitando el problema de la moneda. Análisis, interpretación y reporte de los resultados. Modelos Gaussianos y t de Student. Comparación de grupos y tamaño del efecto. Modelos jerárquicos. Diagnóstico del muestreo y soluciones a posibles problemas.

Tema 3: Modelos lineales generalizados

Tipos de variables: Métricas, ordinales y categóricas. Combinación lineal de predictores. Combinaciones aditivas y no aditivas. Regresión lineal simple y múltiple. Regresión robusta a datos aberrantes. Funciones de enlace. Regresión logística, regresión de Poisson. Distribuciones cero-infladas.

Tema 4: Selección y promediado de modelos

Medidas de exactitud predictiva. Bondad de ajuste vs. complejidad del modelo. Criterios de información (AIC, WAIC). Validación cruzada y aproximaciones PSIS-LOO-CV. Selección de modelos. Promediado Bayesiano de modelos, pesos tipo Akaike, stacking de distribuciones predictivas a posteriori.

Tema 5: Modelos de mezcla y modelos no-paramétricos

Modelos de mezcla finitos. Model-based clustering. Modelos de mezcla continuos. Modelos no paramétricos, modelos de mezcla infinitos y procesos de Dirichlet. Procesos Gaussianos, kernels.

BIBLIOGRAFÍA

- Richard McElreath. Statistical Rethinking. CRC press. 2015
- Osvaldo Martin. Bayesian Analysis with Python. Second edition. Packt Publishing. 2019
- John Kruschke. Doing Bayesian Data Analysis. Second Edition: Boston: Academic Press; 2014.
- Allen B. Downey. Think Bayes. 1 edition. O'Reilly Media; 2013.
- Devinderjit Sivia, John Skilling. Data Analysis: A Bayesian Tutorial. 2 edition. Oxford University Press; 2006.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3 edition. Boca Raton: Chapman and Hall/CRC; 2013.
- Kevin Murphy. Machine Learning a probabilistic perspective. MIT Press 2012



- Edwin T. Jaynes. Probability Theory: The Logic of Science. Bretthorst GL, editor. Cambridge, UK ; New York, NY: Cambridge University Press; 2003.
- Sumio Watanabe. Mathematical Theory of Bayesian Statistics. Chapman and Hall/CRC. 2018

EXPOSITOR

Dra. Osvaldo Martín. Universidad Nacional de San Luis-CONICET

COORDINADORA

Dra. María Graciela Molina

Departamento de Ciencias de la Computación. FACET-UNT.

CARGA HORARIA Y MODALIDAD DE DICTADO

Carga horaria total: 40 horas

El curso se dictará durante 5 días. Por las mañanas se realizará el dictado de clases teóricas y explicaciones. Durante las tardes se realizarán clases con trabajos prácticos a resolver por los alumnos con consultas al docente, y se trabajará en la preparación de un trabajo final.

FORMACIÓN PREVIA REQUERIDA

Dirigido a alumnos graduados o investigadores de carreras científicas. Conocimientos básicos de lenguajes de programación Python, R o Julia.

SISTEMA DE EVALUACIÓN

Para aprobación se requiere:

- Asistir al menos al 80 % de las clases teórico-prácticas
- Desarrollar un trabajo final.

La nota final calcula como:

$NOTA = 0.4 (\% \text{ de Asistencia}) + 0.6 (\% \text{ trabajo final})$

Aprobación con nota superior a 6 (seis)

LUGAR DE REALIZACIÓN

Laboratorio de Computación Científica

Dpto de Ciencias de la Computación – FACET-UNT

CUPO

Sin límite

FECHAS ESTIMADAS DE REALIZACION

Marzo 2020.