Estadística para la Ciencia de Datos

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Másteres Universitarios

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.

DATOS GENERALES

Breve descripción

Esta asignatura tiene como objetivo proporcionar a los estudiantes conocimientos básicos sobre los métodos estadísticos para la descripción, presentación y análisis de datos. Los objetivos específicos son:

- Conocer los métodos de estadística descriptiva.
- Conocer técnicas de muestreo y Monte Carlo.
- Conocer los fundamentos de inferencia estadística.
- Conocer los contrastes paramétricos y no paramétricos para una y varias muestras.
- Conocer las técnicas de remuestreo (*bootstrap*).
- Conocer los modelos de regresión y estimación de máxima verosimilitud.
- Conocer los métodos de regularización y regresión contraída (*ridge regression*)

Título asignatura

Estadística para la Ciencia de Datos

Código asignatura

102264

Curso académico

2024-25

Planes donde se imparte

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE

Créditos ECTS

6

Carácter de la asignatura

OBLIGATORIA

Duración

Cuatrimestral

Idioma

Castellano e Inglés

CONTENIDOS

Contenidos

1. Estadística descriptiva.
2. Muestreo y Monte Carlo.
3. Fundamentos de inferencia estadística: una y varias variables.
4. Contrastes paramétricos y no paramétricos para una y varias muestras.
5. Técnicas de remuestreo (*bootstrap*).
6. Modelos de regresión. Estimación de máxima verosimilitud.
7. Regularización. Regresión contraída (*ridge regression*).

RESULTADOS DE APRENDIZAJE Y DE FORMACIÓN

Generales

CG1 - Integrarse eficazmente en un grupo de trabajo y trabajar en equipo, compartir la información disponible e integrar su actividad en la actividad del grupo colaborando de forma activa en la consecución de objetivos comunes

CG2 - Capacidad de estudio, síntesis y autonomía suficientes para desarrollar de forma autónoma proyectos básicos de investigación

CG6 - Buscar, obtener, procesar, comunicar información y transformarla en conocimiento

CG7 - Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados

CG8 - Capacidad de actualización de los conocimientos expuestos en el ámbito de la comunidad científica

Transversales

CT1 - Analizar y combinar información utilizando diferentes fuentes

CT3 - Dominio de la gestión del tiempo

CT5 - Capacidad de trabajo autónomo y toma de decisiones

CT6 - Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar

Específicas

DSDA01 - Utilizar el análisis predictivo para analizar grandes volúmenes de datos y descubrir nuevas relaciones

DSDA02 - Utilizar técnicas estadísticas apropiadas sobre los datos disponibles para lograr una visión adecuada de los mismos

DSRM06 - Aplicar el ingenio propio para resolver problemas complejos y desarrollar ideas innovadoras

PLAN DE APRENDIZAJE

Actividades formativas

AF1 - Participación y asistencia a lecciones magistrales y seminarios (30 horas)

AF2 - Realización de prácticas de computación y análisis de datos (30 horas)

AF6 - Tutorías (presenciales o por medio de recursos telemáticos) (7 horas)

AF7 - Elaboración de informes de laboratorio y trabajos (40 horas)

AF8 - Estudio individual de contenidos de la asignatura (40 horas)

A10 - Pruebas de evaluación (3 horas)

Metodologías docentes

Se comenzará por una exposición de los conceptos y métodos básicos, incluyendo ejemplos ilustrativos sencillos de distintas disciplinas, que serán analizados y discutidos en común.

Los estudiantes, organizados en grupos, realizarán un análisis práctico detallado de distintos casos de estudio para problemas de asociación, segmentación, clasificación y predicción aplicando las técnicas adecuadas en cada caso. La parte práctica se realizará utilizando el software estadístico R de uso libre.

Resultados de aprendizaje

- Entender las técnicas para la descripción estadística de un conjunto de datos y saber aplicar críticamente las que mejor se adapten a un caso práctico.
- Entender las técnicas estadísticas de muestreo y remuestreo.
- Entender las técnicas de contraste de hipótesis y saber aplicarlas a problemas típicos.
- Entender el modelo de regresión lineal y saber aplicarlo en casos prácticos.
- Entender el método de máxima verosimilitud para estimación de parámetros.
- Conocer el concepto de regularización.

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

SE1 - Examen (escrito, oral y/o práctico en el aula de computación) (40%)

SE2 - Valoración de informes y trabajos escritos (60%)

PROFESORADO

Profesor responsable

Cruz Rodríguez, Marcos

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

Profesorado

Nieto Reyes, Alicia

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

Casanueva Vicente, Ana

DOCTORA POR LA UNIVERSIDAD DE CANTABRIA.
PROFESORA PERMANENTE LABORAL, ÁREA DE MATEMÁTICA APLICADA.
UNIVERSIDAD DE CANTABRIA.

García Díaz, Daniel

Dr. Ciencia y tecnología.
Contratado.
Instituto de Física de Cantabria (IFCA).

HORARIO

Horario

09/10/2024

15:30 - 16:30

Estadística Descriptiva

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

21/10/2024

16:30 - 17:30

Estadística Descriptiva

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

22/10/2024

15:30 - 16:30

Repaso de distribuciones de probabilidad

Marcos Cruz Rodríguez

Doctor en Ciencias.

Profesor Titular de Universidad.
Universidad de Cantabria.

23/10/2024

16:30 - 17:30

Repaso de distribuciones de probabilidad

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

28/10/2024

15:30 - 16:30

Números aleatorios y teorema central del límite

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

04/11/2024

16:30 - 17:30

Números aleatorios y teorema central del límite

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

05/11/2024

15:30 - 16:30

Simulación: Método del rechazo, método funcion cuantil, método Boxmüller

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

06/11/2024

16:30 - 17:30

Simulación: Método del rechazo, método funcion cuantil, método Boxmüller

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

08/11/2024

17:30 - 19:30

Estimadores estadísticos

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

11/11/2024

17:30 - 19:30

Regresión

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

18/11/2024

17:30 - 19:30

Regresión Lineal

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

22/11/2024

17:30 - 19:30

Clasificación Lineal

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

25/11/2024

17:30 - 19:30

Regresión logística

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

29/11/2024

17:30 - 19:30

Regresión logística

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

02/12/2024

17:30 - 19:30

Regresión logística

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

09/12/2024

17:30 - 19:30

Regresión logística

Alicia Nieto Reyes

Doctora por la Universidad de Cantabria
Profesora Contratada Doctora
Universidad de Cantabria

11/12/2024

15:30 - 17:30

Práctica regresión lineal

Daniel García Díaz

Dr. Ciencia y tecnología.
Contratado.
Instituto de Física de Cantabria (IFCA).

16/12/2024

15:30 - 17:30

Práctica regresión logística

Daniel García Díaz

Dr. Ciencia y tecnología.
Contratado.
Instituto de Física de Cantabria (IFCA).

18/12/2024

15:30 - 17:30

Práctica remuestreo

Daniel García Díaz

Dr. Ciencia y tecnología.
Contratado.
Instituto de Física de Cantabria (IFCA).

07/01/2025

17:30 - 19:30

PCA

Ana Casanueva Vicente

DOCTORA POR LA UNIVERSIDAD DE CANTABRIA.
PROFESORA PERMANENTE LABORAL, ÁREA DE MATEMÁTICA APLICADA.
UNIVERSIDAD DE CANTABRIA.

08/01/2025

15:30 - 17:30

Práctica de PCA

Ana Casanueva Vicente

DOCTORA POR LA UNIVERSIDAD DE CANTABRIA.
PROFESORA PERMANENTE LABORAL, ÁREA DE MATEMÁTICA APLICADA.
UNIVERSIDAD DE CANTABRIA.

13/01/2025

17:30 - 19:30

Selección de variables

Ana Casanueva Vicente

DOCTORA POR LA UNIVERSIDAD DE CANTABRIA.
PROFESORA PERMANENTE LABORAL, ÁREA DE MATEMÁTICA APLICADA.
UNIVERSIDAD DE CANTABRIA.

15/01/2025

17:30 - 19:30

Práctica de selección de variables

Ana Casanueva Vicente

DOCTORA POR LA UNIVERSIDAD DE CANTABRIA.
PROFESORA PERMANENTE LABORAL, ÁREA DE MATEMÁTICA APLICADA.
UNIVERSIDAD DE CANTABRIA.

20/01/2025

15:30 - 17:30

Examen

Marcos Cruz Rodríguez

Doctor en Ciencias.
Profesor Titular de Universidad.
Universidad de Cantabria.

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Bibliografía básica

María Dolores Ugarte, Ana F. Militino, and Alan T. Arnholt: "Probability and Statistics with R", Second Edition. Chapman & Hall 2015.

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani: "An Introduction to Statistical Learning" (with Applications in R). Springer-Verlag 2013.

Bibliografía complementaria

Hastie, Tibshirani and Friedman: "The Elements of Statistical Learning" (2nd edition). Springer-Verlag 2009.

David MacKay: "Information Theory, Inference, and Learning Algorithms" Cambridge University Press, 2003.